
Strathclyde Institute for Pharmaceutical and Biomedical Sciences
Faculty of Science

# Systems Biology Approach for Metabolic Engineering in *Streptomyces*: The Phosphoenolpyruvate-Pyruvate-Oxaloacetate Node

Jana Katharina Hiltner

Thesis presented in fulfilment of the requirement for the degree of
Doctor of Philosophy

2015

# Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.



Signed:

Date:

# Acknowledgments

The work for this thesis would not be what it is without the support and help I have received from a number of people. Thank you all for being part of this.

I would like to thank Paul for his supervision during these four years which allowed me to grow substantially not only as a scientist but also as a person. Thank you for entrusting me with this project, I enjoyed working on it and hope we will be able to continue working on this- I feel, we just got started.

I thank Iain for all the valuable insights and fruitful discussions throughout my PhD, every single time when talking to you I learned something new.

I thank the Scottish University Life Science Alliances (SULSA) BioSkape for the funding of this project, as well as the Microbiology Society and S*f*AM for funding to attend a number of different conferences and meetings.

I thank Paco and Pablo for the collaboration we had going on for most of my PhD and especially Paco for letting me visit his lab in Irapuato in Mexico to explore the evolutionary angle in a lot more detail. Furthermore Nelly and Christian for the help and teaching me how to use command line and to read scripts. I would like to thank the MacRobertson scholarship which I received in order to realise this visit to Mexico and Chicago.

I thank the members of the Hoskisson research group who were there during my PhD: I thank Kirsty for sharing the PhD experience from day one with me and for showing me so many hills and munros. Thank you, Emilio for your friendship, especially for cheering me up so many times over quite a few years now. And for your chemical advice when I got lost in curly arrows or structures! I thank Danny for making me understand some physics and for many insights into your work experience. Thank you, Vartul for your help with some of the bioinformatics and encouraging me to pursue this part of science in more detail. Thank you Sarah, for joining our group during my last year of the PhD, you brought a lot of enthusiasm and happiness. I would like to thank Leena for being my immediate role model in being a mum and a great scientist at the same time. Also thank you for teaching me how to do RNA Sequencing.

I would like to thank Richard for the experimental work we shared for the RNA Seq, as well as Sara for helping me with the qPCR and Govind from JIC for the scripts and insights into the data analysis. And I thank Duska for her advice on the proteome sampling.

I would like to thank Antonella for every Zumba lesson and every laugh we shared. I miss you in the office, my wee little Italian mind-twin!

Also I would like to thank the students I got to supervise during my PhD, I hope you enjoyed it and learned from it, I certainly did- Thank you Joao, Martina and Ashley.

Furthermore the microbiology PIs in our lab Paul, Nick and Arnaud for sharing their insights and offering help as well as everybody else from level 6 of the Hamnett Wing.

I would like to thank Mervyn Bibb, my external examiner and Arnaud, my internal examiner for reading my thesis, their comments and the fruitful discussion during my viva!

I thank my parents Birgit and Uwe for giving me roots and wings to fly and my family from Argentina Gaby, Mario, Juliana, Leandro and Salvador for giving me a second home and opening my eyes to the world.

I thank Jan for his love, belief and companionship over all those years, there is probably nothing that can describe this properly, so simply thank you for being you and for being with me. And thanks to our two cats Luke and Leia who made the writing bearable.

**For Janja -I will always miss you.**

“I have a friend who's an artist and has sometimes taken a view which I don't agree with very well. He'll hold up a flower and say "look how beautiful it is," and I'll agree. Then he says "I as an artist can see how beautiful this is but you as a scientist take this all apart and it becomes a dull thing," and I think that he's kind of nutty. First of all, the beauty that he sees is available to other people and to me too, I believe. Although I may not be quite as refined aesthetically as he is ... I can appreciate the beauty of a flower. At the same time, I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes. The fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting; it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic? All kinds of interesting questions which the science knowledge only adds to the excitement, the mystery and the awe of a flower. It only adds. I don't understand how it subtracts.”

— Richard Feynman

# Contents

# ABSTRACT

The emergence of antibiotic resistance poses a threat to humankind in combination with the significant drop in discovering new antibiotics since the 1980s. Although *Streptomyces* are prolific producers of bioactive metabolites, genome analysis shows more biosynthetic clusters than compounds produced under laboratory conditions. Besides identification of a new potential compound, its biosynthesis needs to be understood to exploit it for industrial production. This thesis focussed on understanding the interaction of primary metabolism, in particular the Phosphoenolpyruvate-Pyruvate-Oxaloacetate (PEP-PYR-OAA) node, with specialised metabolism to identify targets for metabolic engineering. *Streptomyces coelicolor* and its polyketide actinorhodin formed a proof of concept. Many central carbon metabolic enzymes are annotated by isofunctional genes. Such expansions were analysed by creating a database of 614 actinobacterial species: more expansions were observed in genera rich in specialised metabolism. This, in combination with screening mutants for changes in polyketide production, identified phosphofructokinase, pyruvate kinase, pyruvate phosphate dikinase, pyruvate dehydrogenase complex and malic enzyme as potential targets. These are expanded in *Streptomyces* and showed changes in actinorhodin production for at least one of each gene-set. Phosphofructokinase and malic enzymes have been previously studied. Differential gene expression analysis by RNA-Seq for glycolytic and gluconeogeneic conditions showed that most genes of expanded enzymes were differentially-expressed. Analysis of metabolite pools of central carbon metabolism supported this further. The two pyruvate kinases did not show differential expression and were analysed in more detail. A *pyk2* mutant exhibited altered growth on glucose, whereas *pyk1* mutants showed altered actinorhodin production. Biochemical characterisation suggested Pyk2 as the housekeeping enzyme with low $S_{0.5}$ for both substrates, whereas Pyk1 was activated under low energy state conditions.

Overall, the thesis provides a better understanding of the interaction of primary metabolism and expansion in respect to specialised metabolism to be used in strain development for the production of antibacterial compounds.

Part of the Introduction of this thesis has been published as a Book Chapter in Advances in Applied Microbiology. A copy of the review can be found at the end of the appendix to this thesis.

Hiltner, J.K., Hunter, I.S., and Hoskisson, P.A. (2015) Tailoring Specialized Metabolite Production in *Streptomyces*. In *Advances in Applied Microbiology*. Vol 91 Chapter 4. pp. 237–252.

# 1 INTRODUCTION

The widespread use of antibiotics is one of the most successful additions to human healthcare that emerged from the discovery of penicillin in 1928 by Alexander Fleming and its subsequent development to industrial scale production by Howard Florey and co-workers. This discovery of bioactive metabolite production by microorganisms, useful in human health, prompted the so-called golden era of antibiotics starting in the 1950s. Spanning 30 or so years this era of search and discovery resulted in the delivery of many chemical classes of antibiotic, antifungal, anti-helminthic, anti-cancer agents and immunosuppressive drugs to market (Berdy, 2005; Demain and Sanchez, 2009; Aminov, 2010; Davies and Davies, 2010). During this intensive period of research the Actinobacteria came to the fore as prolific producers of bioactive metabolites, particularly antibiotics from the genera *Streptomyces* and *Micromonospora* (Watve *et al.*, 2001; Davies and Davies, 2010). This initiated the widespread public perception that antibiotics were 'wonder-drugs' that signalled the end of life-threatening bacterial infections, yet this was premature. Rapidly after the introduction of these drugs to the clinic, antibiotic resistance was observed (Davies and Davies, 2010). This began a race to discover and bring to the clinic new antibiotics to help combat the emergence of resistance. Yet with the introduction each new drug rapid resistance was observed. This inevitable resistance, coupled to the rising costs of development, tightening of regulatory rules, diminishing discovery rates (and rediscovery of known compounds), and the lack of financial returns due to short durations of treatments resulted in the withdrawal of large pharmaceutical companies from large-scale antibiotic discovery in the 1990's (Projan, 2003). These problems coupled with profligate use of antibiotics in medicine and agriculture lead to the emergence of extensive antibiotic resistance on a global scale. Since 2001 the World Health Organisation is raising concerns that the rise of antibiotic resistance was becoming one of the major threats to human health and that researchers, clinicians, industry and policy makers needed to work together to address this multifactorial problem (WHO, 2001).

Once a promising antibiotic, or for instance any medically interesting compound, has been identified, it needs to be produced in large quantities. However the level of natural synthesis of natural products is not sufficient for human needs and metabolic engineering is one part of the optimisation process to achieve higher antibiotic titers besides improving physical and chemical parameters. The production can be carried out by either the natural or a heterologous host.

Metabolic engineering refers to the modification of metabolic fluxes in order to increase or decrease certain fluxes with the aim to obtain a higher titre of a compound of interest in the producer. These flux modifications are usually achieved by eliminating bottle necks by either introducing a new enzyme or a whole new pathway or by deleting a certain enzyme or a pathway for an unwanted by-product using gene deletion (Bailey, 1991). Furthermore the choice of substrate or additives to the culture medium for the organism can be used to induce certain pathways. The increased use of advanced DNA synthesis and assembly techniques to improve and design new biosynthetic gene clusters is an area of great interest. Moreover this synthetic biology revolution has great potential to transform the more traditional disciplines of metabolic engineering, through creation of novel heterologous hosts with engineered precursor supply or deletion of native biosynthetic clusters that may help overcome metabolic limitations (Gomez-Escribano and Bibb, 2011).

## 1.1 THE GENUS *STREPTOMYCES*

The actinobacterial phylum represents a large lineage of physiological and morphologically diverse Gram-positive bacteria and includes the industrially, agriculturally and medically important genera such as *Bifidobacterium, Corynebacterium, Mycobacterium, Nocardia, Leifsonia, Frankia,* and *Streptomyces* (Ventura *et al.*, 2007). The most speciated family is Streptomycetaceae with more than 600 taxa including the genus of *Streptomyces* which are sporulating bacteria with filamentous growth, characterised by large (> 7 Mbp) high G+C-content and linear genomes (Labeda *et al.*, 2012).

The most recent common ancestor of all *Streptomyces* is believed to have originated about 440 million years ago, just after the colonisation of land by green plants. Thus they played a pivotal role in establishing the soil environment (Embley and Stackebrandt, 1994; Chater and Chandra, 2006).

Their natural habitat is the soil and the rhizosphere and they only have a limited pathogenic importance (Hopwood, 2007). Streptomycetes show a complex life cycle with the development of vegetative mycelia and aerial hyphae as well as spore formation. When a spore germinates, it first forms a substrate mycelium branching into the substrate with few partition walls. At this stage one cell can contain several genomes. Elongation only takes place at the tip of the hyphae. In the next stage, which is induced upon nutrient depletion, aerial hyphae are formed, the substrate mycelium starts programmed cell death and its material is reused for developing aerial hyphae. At this point partition wall formation is more

frequent and branches start to coil. Finally, chains of spores are formed and released at the tips (Figure 1.1; (McGregor, 1954; Chater, 1993; Flärdh and Buttner, 2009). If the cells are deficient in producing aerial hyphae, they are referred to be bald (no 'hair' on the colonies). The *bld* genes are involved in the aerial hyphae formation and another important set of genes in the developmental process are *whi* genes. The name '*whi*' for those genes refers to white as those mutants lack the grey spore pigment and appear as white. The '*whi*' genes are responsible for the segmentation and coiling of the hyphae (Chater, 1993; Hopwood, 2007). *Streptomyces coelicolor* A3(2) is the model organism for this genus. In terms of speciation, this isolate should technically be termed *S. violaceruber*. However the isolate has been continued to be called *S. coelicolor* as a matter of convenience (Hopwood, 1999). Its chromosome is linear and the genome size is of about 8.7 Mb. The genome has 7825 predicted genes of which 965 are predicted to be regulatory proteins and approximately 20 different clusters of known and unknown specialised metabolites (Bentley *et al.*, 2002). The origin of the genus was obscure and believed to be a connective link between bacteria and fungi. However, the absence of a nucleus in *Streptomyces* indicated their bacterial nature (Glauert and Hopwood 1960). Nevertheless many features of them are reminiscent of a eukaryote rather than a prokaryote and are an outstanding group of microorganisms to study. Therefore they have been studied by geneticists and microbiologist since their discovery in the late 1940 (Hopwood, 1999).

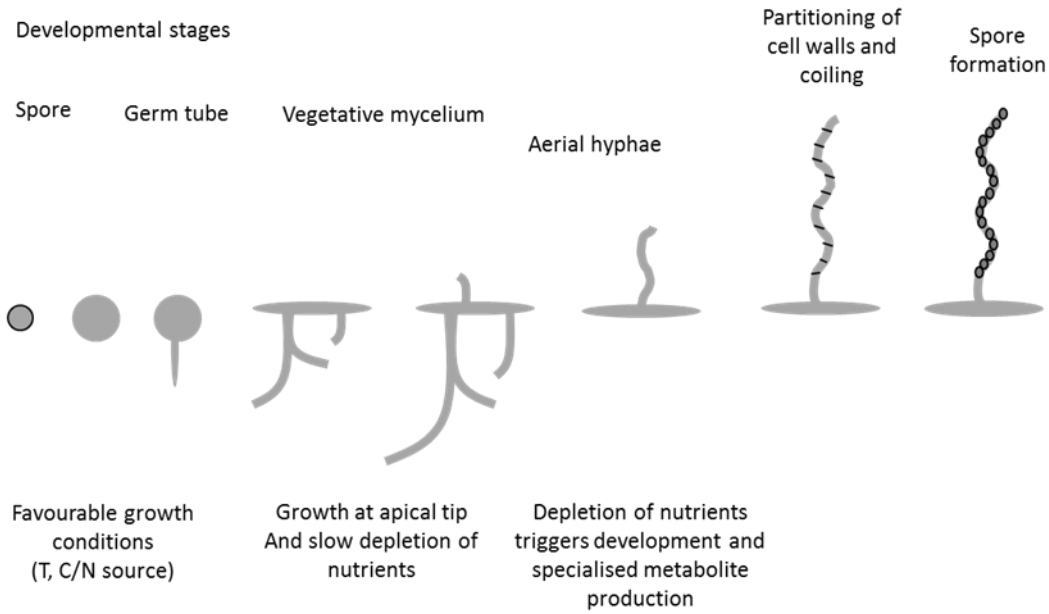



Figure 1.1 Schematic overview of the life cycle of *Streptomyces coelicolor* A3(2)
Favourable conditions will trigger the germination of the spore and the formation of vegetative mycelium. Once the nutrients are depleted, the vegetative mycelium dies and release of its nutrients is utilised to form aerial hyphae that are erected in the air. At this stage multiple genomes are present, once the growth of the aerial hyphae stops, cell walls thicken and the grey spore pigment is produced. Chaplins and rodlins are produced to protect the spore from drought. Septa are formed and one genome per compartment is distributed. Adapted from Kieser *et al.*, 2000; Flärdh and Buttner, 2009
Legend: T = temperature, C = carbon, N = nitrogen

## 1.2 Primary and Specialised Metabolism in Streptomycetes

The distinction of primary and secondary metabolism probably arose through the studies of Albrecht Koessel, who proposed that plants show two different metabolisms: 'primary' and 'secondary', where primary metabolism is common among all organisms and comprises all essential reactions; in contrast to secondary metabolism, that is thought to be specialised, distinct and comprised of species specific pathways (Hartmann, 2008; Firn and Jones, 2009). This view was been widely adopted and adapted through many fields of biology and while the designations imply secondary metabolism is less important than primary metabolism, this view has modified over the years, and it is implicit that secondary metabolism is dependent on supply of precursors from primary metabolism. Firn & Jones (2009) challenged the view of disconnecting primary and secondary metabolism as misleading given the obvious intimate link and the presence of multiple, yet separate metabolic routes may reflect distinct evolutionary selection pressures. This is also in accordance with the development of the term "specialised metabolites" which may be a more useful term in replacing the bias of 'secondary' - as it implies less importance (Davies, 2013; van Keulen and Dyson, 2014; Hiltner *et al.*, 2015).

### 1.2.1 Primary Metabolism

The main pathways in primary metabolism are glycolysis, pentose phosphate pathway (PPP) and the tricarboxylic acid (TCA) cycle including the glyoxylate shunt. Often these are referred to as central carbon metabolism and the reactions involved provide building blocks for RNA, DNA, proteins, fatty acids as well as specialised metabolites (Figure 1.2). The metabolism consists of different processes such as uptake, transport, conversion and storage of different metabolites as well as regulation of all of those processes.

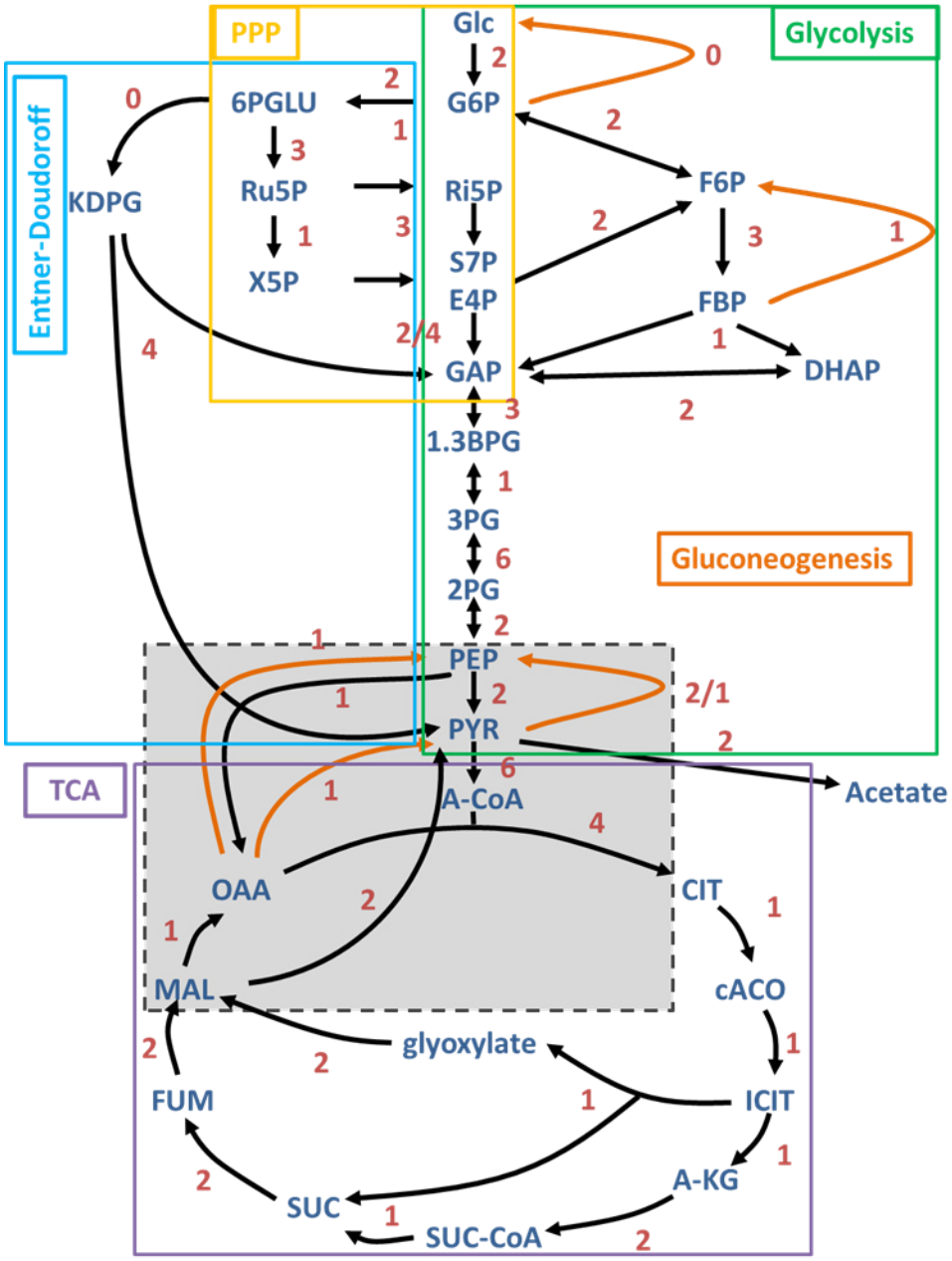



Figure 1.2 Schematic overview with main metabolites of the central carbon metabolism grouped according to pathways: glycolysis (green), pentose phosphate pathway (PPP, yellow), Entner-Doudoroff pathway (blue), tricarboxylic acid cycle (TCA, purple), gluconeogenesis (orange arrows). The grey box indicates the Phosphoenolpyruvate-Pyruvate-Oxaloacetate node. Red numbers indicate number of genes annotated in the genome of *S. coelicolor* A3(2) Legend: Glc = Glucose, G6P = Glucose-6-Phosphate, 6-PGLU = 6-phosphogluconate, Ru5P = Ribulose-5-phosphate, X5P = Xylose-5-Phosphate, KDPG = 2-keto-3-deoxy-6-phosphogluconate, F6P = Fructose-6-Phosphate, FBP = Fructose 1.6-bisphosphate, DHAP = dihydroxyacetone phosphate , Ri5P = Ribose-5-Phosphate, S7P= Seduheptulose-7-Phosphate, E4P = Erythrose-4-Phosphate, GAP = glyceraldehyd-3-phosphate , 1.3BGP = 1.3-bisphosphoglycerate, 3PG = 3-phosphoglycerate, 2PG =2-phosphoglycerate, PEP-= Phosphoenolpyruvate, PYR = Pyruvate, ACoA = Acetyl-CoA, Cit = Citrate, cAco = cisAconitate, ICit = Isocitrate, A-KG-= α-Ketoglutarate, SucCoA = Succinyl-CoA, Suc = Succinate, Fum = Fumarate, Mal = Malate, OAA = Oxaloacetate

**Carbon Uptake and Storage**

The natural habitat of *Streptomyces* offers a carbon rich environment and at the same time is low in phosphate and nitrogen (Hodgson, 2000). The first step for metabolising any substrate is the uptake into the cell. *Streptomyces* are saprophytic and degrade the two most abundant polymers, cellulose and chitin, by secreting necessary enzymes into their surroundings to break the polymers down to transportable compounds (Chater *et al.*, 2010).

A large number of secreted proteins have been predicted in the *Streptomyces* genome, including 60 proteases, 13 chitinases, eight cellulases and endoglucanases, three amylases and two pectate lyases (Bentley *et al.*, 2002). The extracellular enzymes are secreted by various different protein secretion pathways such as Sec (Driessen and Nouwen, 2008), Esx (Akpe San Roman *et al.*, 2010) and Tat (Rose *et al.*, 2002; Dilks *et al.*, 2003; Widdick *et al.*, 2006). Besides degrading enzymes there are also substrate binding proteins such as chitin-binding proteins (Schnellmann *et al.*, 1994; Kolbe *et al.*, 1998) or carbohydrate binding proteins which help the organism to degrade large polymers (Walter *et al.*, 1998).

The nutritional matrix in the natural habitat of *Streptomyces* is reflected in the type of uptake system they possess. In contrast to other bacteria, *Streptomyces* do not possess a PEP-PTS for glucose as known so far, however a phosphotransferase systems (PTS) for fructose inducible by the latter has been identified (Titgemeyer *et al.*, 1995). According to the KEGG database the *S. coelicolor* genome contains transport systems for PEP (PtsI, SCO1391), Maltose (MalX, SCO0434), N-Acetylglucosamine (GlcNAc, NagE, SCO1390, 2905-07), L-Ascorbate (UlaA/UlaC, SCO0136/7) as well as fructose via PEP (FruA/FruB, SCO3196). In *Streptomyces* glucose uptake occurs mainly by GlcP (GlcP1 -SCO5578 and GlcP2- SCO7153) a major facilitator superfamily (MFS) sugar permease. GlcP1 is the major player as its transcription is induced by glucose in contrast to very low expression of GlcP2, which was further confirmed by gene knockout studies (van Wezel *et al.*, 2005).

For instance, catabolite repression is present, referring to the ability to use preferentially one specific carbon source prior to another one. Nevertheless, it shows a different order than in other bacteria. Glucose represses the operons of agarose, arabinose, galactose, fructose, glycerol catabolism as well as maltose transport. Whereas cellobiose represses agarose, arabinose, fructose, galactose and glycerol operons. Moreover an accumulation of organic acids, that is induced by glucose, is observed and is thought to represent decoupled TCA and

glycolysis (Cochrane and Dimmick, 1949; Dekleva and Strohl, 1987; Hodgson, 2000; D'Huys *et al.*, 2011). In addition it has been shown that *S. coelicolor* is able to sense the extracellular concentration of N-Acetylglucosamine (GlcNAc) and that at high concentration the cells are kept in a vegetative growth state. The same study also reveals that DasR, a GntR-like regulator, is responsible for the sensing mechanism (Rigali *et al.*, 2006; Rigali *et al.*, 2008). This links the availability of certain metabolites with primary metabolism, as well as with the developmental biology of the organism. Another important aspect is carbon storage, which can serve as backup nutrient source. *Streptomyces* show a variety of carbon storage molecules such as trehalose, polyhydroxybutyrate, neutral lipids and triaglycerol (Martin *et al.*, 1997). It appears that trehalase activity increases after germination before outgrowth in both rich or defined nutrient medium (McBride and Ensign, 1987). Trehalose phosphate synthase is in low concentrations in spores and increases parallel to the biomass, only when outgrowth commenced. In contrast trehalose phosphate phosphatase is highly active in spores and drops when outgrowth starts. Another common storage molecule, glycogen, is metabolised during spore formation (Bruton *et al.*, 1995; Schneider *et al.*, 2000). Furthermore triacylglycerol is accumulated during exponential growth phase till late stationary phase and can also be used as precursor molecule for antibiotic production (Olukoshi and Packter, 1994).

### 1.2.2 Specialised Metabolism in *Streptomyces*

The plethora of specialised metabolites produced by this group of organisms represents a high degree of chemical diversity including classes such as polyketides, terpenes, lactams, aminoglycosides and non-ribosomal peptides with many species able to produce multiple examples of the same class of specialised metabolite (Berdy, 2005). An idea of this diversity can be found in the StreptomeDB database http://www.pharmaceutical-bioinformatics.de/streptomedb/), where the structures of more than 2400 compounds from more than 1900 species are held (Lucas *et al.*, 2013). These estimates are and will continue to increase exponentially along with the expansion of whole genome sequencing projects and specialised metabolite predictions software such as antiSMASH (Medema *et al.*, 2011) or EvoMining (Cruz-Morales *et al.*, 2015). In addition, the sequences of about 7758 actinobacteria genomes are available in the Genomes Online Database (GOLD; Reddy *et al.*, 2015) and represent a vast resource for identification of biotechnologically useful genes and gene clusters. One particularly revealing actinobacterial genome feature that the next-generation sequencing revolution has opened the eyes to is the vast array of antibiotic (and

other specialised metabolite) biosynthetic gene clusters present in the genomes of actinobacterial strains. Significantly many of these biosynthetic clusters are cryptic or poorly expressed in their natural hosts and offer a great potential for the discovery of novel, clinically useful compounds. Moreover these gene clusters represent a significant resource of genes for synthetic biology to create novel metabolites through synthetic biology, semi-synthesis methods, where existing compounds may be biosynthesised as a chemical backbone, which are then further modified through synthetic chemistry and great advances have been made in this area in recent years (Wu, 2000). However semi-synthetic derivatives of natural products are often limited through availability of starting material. Semi-synthetic derivatives of erythromycin such as azithromycin require increasing amounts of the starter molecule yet the natural producing strains have been difficult to engineer to high production levels when compared to related species (Wu *et al.*, 2011). Given that specialised metabolites, such as antibiotics, are derived from primary metabolic starter units, a deeper understanding of how these precursors are synthesised and channelled in to the biosynthesis of these specialised metabolites offers great potential in metabolic engineering and manipulation through synthetic biology approaches for the development of novel compounds. Many specialised metabolites have acetyl-CoA, malonyl-CoA, propionyl-CoA, methylmalonyl-CoA (Olano *et al.*, 2008) or amino acids as direct precursors (Stirrett *et al.*, 2009) and competition for the same precursors between central metabolism and specialised metabolism may represent key conflicts in precursor supply such as in fatty acid metabolism and polyketide synthesis (Rodríguez *et al.*, 2012). Given these issues, a better understanding of precursor supply may enable increased production of the poorly understood cryptic/silent biosynthetic clusters (Figure 1.3). Therefore there are several challenges facing researchers in this field, such as how to awaken or enhance production of the array of cryptic gene clusters emerging from whole genome sequencing projects and how we can understand better the links between primary and specialised metabolism so that we can increase metabolite flow to maximise industrial yield of new and existing medically useful compounds.

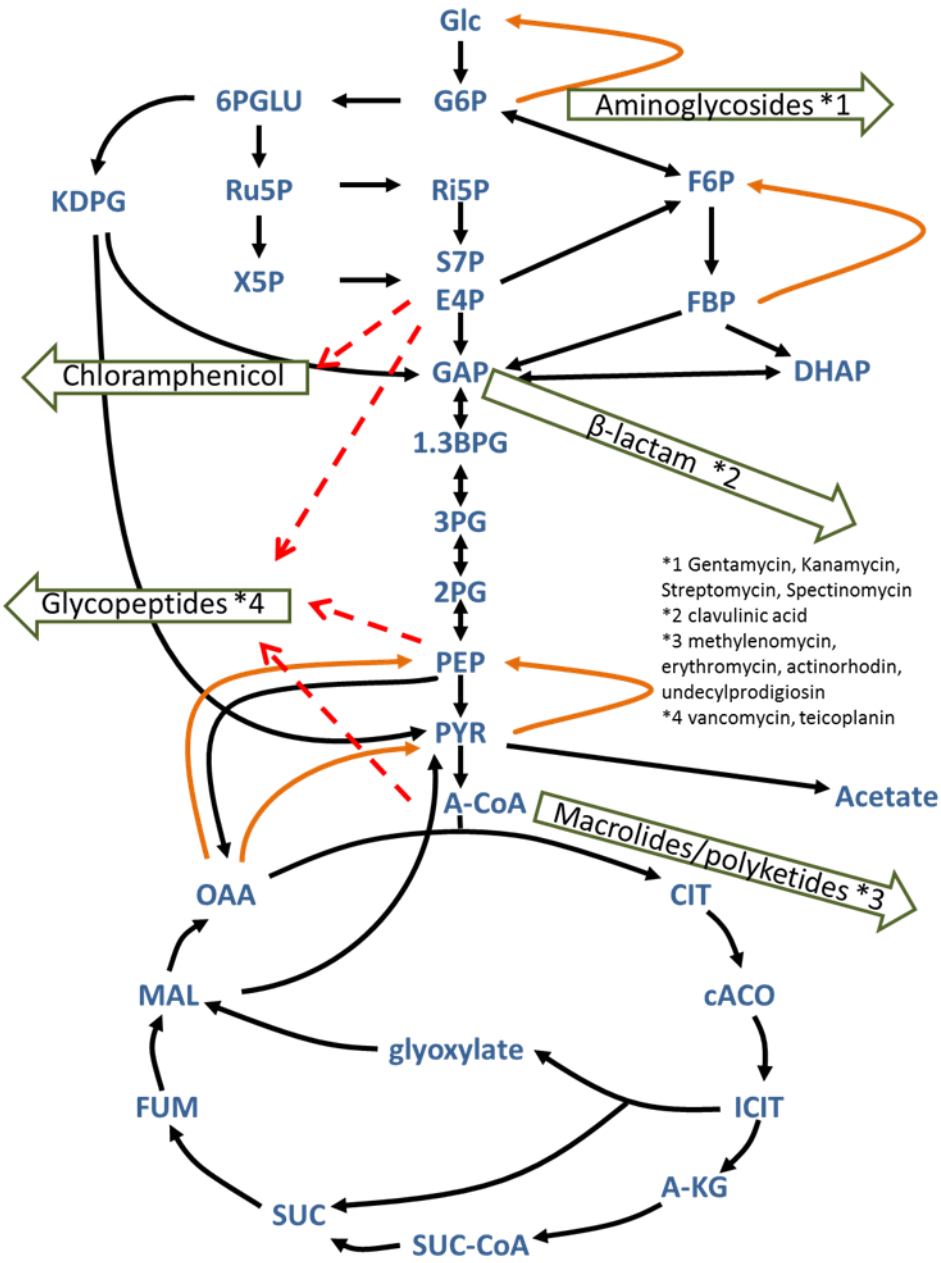



Figure 1.3 Overview of precursor withdrawal from different points of central carbon metabolism for different types of specialised metabolites, many of which are made by *Streptomyces*. Legend: Glc = Glucose, G6P = Glucose-6-Phosphate, 6-PGLU = 6-phosphogluconate, Ru5P = Ribulose-5-phosphate, X5P = Xylose-5-Phosphate, KDPG = 2-keto-3-deoxy-6-phosphogluconate, F6P = Fructose-6-Phosphate, FBP = Fructose 1.6-bisphosphate, DHAP = dihydroxyacetone phosphate , Ri5P = Ribose-5-Phosphate, S7P= Seduheptulose-7-Phosphate, E4P = Erythrose-4-Phosphate, GAP = glyceraldehyd-3-phosphate , 1.3BGP = 1.3-bisphosphoglycerate, 3PG = 3-phosphoglycerate, 2PG =2-phosphoglycerate, PEP-= Phosphoenolpyruvate, PYR = Pyruvate, ACoA = Acetyl-CoA, Cit = Citrate, cAco = cisAconitate, ICit = Isocitrate, AKG-= α-Ketoglutarate, SucCoA = Succinyl-CoA, Suc = Succinate, Fum = Fumarate, Mal = Malate, OAA = Oxaloacetate; orange arrows indicate gluconeogenesis and red dotted arrows withdrawal of precursor

### 1.2.3 Regulation of Development and Specialised Metabolism

The complex life cycle as well as the production of numerous specialised metabolites requires a tight coordination of the different pathways by regulation. Furthermore the primary metabolism generating precursor supplies for specialised metabolites needs to be regulated as well. An external change such as depletion of a nutrient and availability of a different substrate or other environmental factors can also result in changed routes of the carbon flux.

According to the literature major regulatory points of central carbon metabolism are phosphofructokinase and pyruvate kinase in the glycolysis and pentose phosphate pathways and citrate synthase and pyruvate dehydrogenase in the TCA cycle (Alves *et al.*, 1997; Alves *et al.*, 2001; Butler *et al.*, 2002).

External signals such as growth cessation generally leads to an accumulation of intermediates which can either be utilised for specialised metabolite production or result in overflow metabolism like the secretion of organic acids (Hodgson, 2000).

In *Streptomyces* it has been observed that if grown under carbon excess and especially glucose, they compensate the constant transport of carbon into the cell by an overflow metabolism, since they are usually limited in nitrogen and cannot metabolise it all at the same time. They secrete those overflow metabolites, such as succinic acid or pyruvate (Cochrane and Dimmick, 1949; D'Huys *et al.*, 2011) into the medium, which leads to a temporary acidification of the medium, but later on they uptake those metabolites again (Dekleva and Strohl, 1987; Viollier *et al.*, 2001). *Streptomyces* are also able to grow on C2 compounds such as acetate or aliphatic compounds. These substrates imply anaplerotic reactions to ensure the availability of TCA intermediates as building blocks for the catabolism. Furthermore isocitrate lyase is absent when grown on C2, so an alternative route to glyoxylate is necessary (Bramwell *et al.*, 1993).

Studies of the production of methylenomycin (MMY) showed a switch from glycolysis towards the PPP when MMY was produced, which supports the idea that NADPH is required in higher amounts for the production of polyketides (Obanye *et al.*, 1996). The large genome size and high number of homologues as well as the fungus like life cycle show the ability of *Streptomyces* to adapt to their habitat with complex carbon sources and low concentrations of phosphate and nitrogen. So far it is not yet known why they also have

such a complex specialised metabolism as compared to other organisms occurring in the same habitat (Bentley *et al.*, 2002).

The model organism *S. coelicolor* produces four antibiotics under laboratory conditions, which are actinorhodin ACT (Wright and Hopwood, 1976), undecylprodigiosin (Feitelson *et al.*, 1985), calcium-dependent antibiotic CDA (Hopwood and Wright, 1983) and MMY (Hobbs *et al.*, 1992). However there are more gene clusters present with unknown products which might be of interest (Bentley *et al.*, 2002; Hesketh *et al.*, 2002; Challis and Hopwood, 2003; Challis, 2008). Generally production of antibiotics is growth phase dependent in liquid culture and goes along on solid media with the differentiation of the cells at the formation of aerial hyphae (Bibb, 2005). It is influenced by the growth rate, some signal molecules such as γ-butyrolactone or ppGpp, produced by RelA under nitrogen limiting conditions and by an imbalance of metabolism or physiological stress (Bibb, 2005). Additionally the carbon storage plays an important role in the precursor supply for actinorhodin production (Ryu *et al.*, 2006). Furthermore they are often subject to nitrogen repression (Doull and Vining, 1990).

A general study of the metabolic switches at transcriptional level showed that the strongest transcriptional change coincides with phosphate depletion as the classic metabolic switch (Nieselt *et al.*, 2010). It is marked by strong up-regulation of the phosphate regulon at 36 h and a biosynthetic gene cluster for a secondary phosphate free polymer of the cell wall. The gene cluster of ribosomal protein genes and the protein biosynthetic genes are initially highly expressed and gradually decline upon transient and stationary phases. The nitrogen metabolism gene cluster shows high expression at 20 h after the start of production and shows another transient peak around 28 h. This is a different pattern than expected as no nitrogen depletion was present, because the medium was rich in glutamate. The type I polyketide synthase gene cluster (*cpk*) showed an increase of expression at 22 h and declined at 25 h. Butyrolactone genes had a peak in their expression at 21 h till 22 h and declined after 24 h. Regulatory genes for *cpk* genes *scoT* (SCO6287) a putative enzyme for cleavage of polyketide core carbon chains shows an activity increase after 23 h. At 33 h a peak of the gene cluster including nitrate reductase cluster and developmental genes as chaplins, *bld* and *whi* genes was observed. The transcriptional activity increased at 38 h for undecylprodigiosin and at 43 h for actinorhodin (Nieselt *et al.*, 2010). Nitrogen metabolism is under control of *phoP* (Rodríguez *et al.*, 2001). The depletion of nitrate in the medium coincided with aerial hyphae formation. In the second growth phase accumulation of DNA, reduction of RNA

levels, cell protein accretion, accumulation of glycogen and lipids, increase of α-Ketoglutarate and a decrease of pH took place (Karandikar *et al.*, 1997). Accumulation of glycogen occurs in older substrate mycelium from which aerial hyphae emerge as well as in immature spore chains (Braña *et al.*, 1986; Bruton *et al.*, 1995; Martin *et al.*, 1997). Trehalose is present at all growth stages and seems to be important during sporulation where it accumulates (Braña *et al.*, 1986; Schneider *et al.*, 2000). There seems to be a link to interconvert glycogen and trehalose to meet the needs during the development (Schneider *et al.*, 2000). In presence of excess carbon glycogen can serve as energy reservoir (Hoskisson *et al.*, 2004).

In order to gain an insight of the metabolism it is necessary to understand in detail how it functions and it is regulated. An overview of some of the genes involved in the regulation of the specialised metabolism and development is given in Table 1-1. In *Streptomyces* it is not possible to separate development and specialised metabolites as many regulatory mechanisms affect both at the same time. Not only intracellular regulation takes place, but also the environment such as the carbon source influences the antibiotic production and the development. Cell wall breakdown stimulates antibiotic production, probably by the release of GlcNAc which acts as a signalling molecule in *S.coelicolor* (Rigali *et al.*, 2006; Chi *et al.*, 2011; Świątek *et al.*, 2012). Even though this is critically viewed by other publications who detected that the presence of *Myxococcus xanthus* which prey on *Streptomyces* hyphae also increases actinorhodin production and stimulates the development (D'Alia *et al.*, 2011). In another study it has been shown that heat killed *Bacillus subtilis* cells added to the culture increases the production of undecylprodigiosin (Luti and Mavituna, 2011). Furthermore it has been shown that antibiotic production is greatly influenced by the presence of other *Streptomyces* species, pointing towards a social role of antibiotic production which increases production to suppress the competitor and inhibit the production of potentially threatening compounds. Asocial production, referring to the killing of a competitor, is less often observed (Abrudan *et al.*, 2015; Traxler and Kolter, 2015). This competition and constant struggle actually has a positive effect on the biodiversity and stability of the habitat (Bergstrom and Kerr, 2015; Kelsic *et al.*, 2015).

Table 1-1 Overview of genes involved in secondary metabolite and developmental regulation, SCO numbers were obtained from StrepDB (http://strepdb.streptomyces.org.uk). Legend: P = phosphate N= nitrogen, PTS= phospotransferase system

| Gene | SCO | Regulatory function | Reference |
|---|---|---|---|
| *absB* | 5572 | ribonuclease III autoregulator | Adamidis and Champness, 1992 |
| *absC* | 5405 | MarR like regulator, includes regulation of SCO7412 | Yang *et al.*, 2010 |
| *abrA1-2* | 1744-45 | negative regulation of specialised metabolism (sensor kinase, response regulator) | Yepes *et al.*, 2011; Rico *et al.*, 2014 |
| *abrC1-3* | 4596-98 | histidine kinases and one response regulator on ACT and CDA | Yepes *et al.*, 2011 |
| *absR1-2* | 6992-3 | regulatory protein StrepDB | http://strepdb.streptomyces.org.uk |
| *accE/B* | 5535-6 | part of ACCase involved in malonyl-CoA synthesis (antibiotic precursor) | Rodríguez *et al.*, 2001 |
| *afsR* | 4426 | stimulator for Act, DNA binding protein | Horinouchi *et al.*, 1989; Santos-Beneit *et al.*, 2011 |
| *afsR2* | 4425 | regulator of specialised metabolism, σ like factor | Kim *et al.*, 2001 |
| *afsK* | 4423 | serine/threonine protein kinase | Jin *et al.*, 2011 |
| *bldA* | RNA | tRNA for Leu | Bibb, 1996; White and Bibb, 1997 |
| *bldB* | 5723 | polymerase s-factor for aerial hyphae formation | Champness, 1988 |
| *bldC/D* | 4091/189 | DNA-binding protein | den Hengst *et al.*, 2010 |
| *bldG* | 3549 | RNA polymerases-factor | Bignell *et al.*, 2000 |
| *bldH* | 2792 | global transcriptional regulator (araC family) | Higo *et al.*, 2011 |
| *bldM* | 4768 | two component regulator | Bibb *et al.*, 2000 |
| *bldN* | 3323 | anti s-factor antagonist | Bibb *et al.*, 2000 |
| *crp* | 3571 | cAMP receptor protein, global transcription regulator | Gao *et al.*, 2012 |
| *dasR* | 5231 | GlcNAc uptake, transcriptional repressor in absence of GlcNAc | Rigali *et al.*, 2006 |
| *devA* | 4190 | GntR like transcriptional regulator | Hoskisson *et al.*, 2006 |
| *glnR/RII* | 2213/4159 | N metabolism, transcription activator | Fink *et al.*, 2002 |
| *phoR/P* | 4229-30 | two component regulator, *afsS* 3 phosphodiesterases and a phytase,indirectly also *glnA/R* | Sola-Landa *et al.*, 2008; D'Alia *et al.*, 2011 |
| *pstS* | 4142 | interaction P and N regulation, PTS | D'Alia *et al.*, 2011 |

| | | | |
|---|---|---|---|
| *ohkA* | 1596 | global regulator secondary metabolism and morphology | Lu *et al.*, 2011 |
| *relA* | 1513 | ppGpp synthase, global regulator | Chakraburtty and Bibb, 1997; Kang *et al.*, 1998; Gatewood and Jones, 2010; D'Alia *et al.*, 2011; Yu *et al.*, 2014 |
| *scbA* | | γ-butyrolactone synthase | D'Alia *et al.*, 2011; Wang *et al.*, 2011 |
| *scbR* | | γ -butyrolactone receptor for yellow CPK | Gottelt *et al.*, 2010 |
| *tamR* | 3133 | MarR regulator, increase of flux through TCA cycle | Huang and Grove, 2013; Huang *et al.*, 2015 |
| *whiA* | 1950 | hypothetical protein | Aínsa *et al.*, 2000 |
| *whiB* | 3034 | sporulation regulatory protein | Fowler-Goldsworthy *et al.*, 2011 |
| *whiD* | 4767 | putative regulatory protein | Fowler-Goldsworthy *et al.*, 2011 |
| *whiE* | 5315 | polyketide cyclase | Ryding *et al.*, 1999 |
| | 5316 | acyl carrier protein | Ryding *et al.*, 1999 |
| | 5317 | polyketide beta-ketoacyl synthase beta | Ryding *et al.*, 1999 |
| | 5318 | polyketide beta-ketoacyl synthase a | Ryding *et al.*, 1999 |
| | 5321 | polyketide hydroxylase | Ryding *et al.*, 1999 |
| *whiG* | 5621 | RNA polymerase sigma factor | den Hengst *et al.*, 2010 |
| *whiH* | 5819 | sporulation transcription factor, maturation of spores | Shen *et al.*, 2008 |
| *whiI* | 6029 | two component regulator | Ryding *et al.*, 1999 |
| *whiJ* | 4543 | hypothetical protein | Aínsa *et al.*, 2000 |
| *zwf1/2* | 1937 | isoenzymes of G6P dehydrogenase of PP | Butler *et al.*, 2002 |
| | 0877 | LAL regulator of ACT production | Guerra *et al.*, 2012 |
| | 7173 | LAL regulator of ACT production | Guerra *et al.*, 2012 |

### 1.2.4 POLYKETIDE BIOSYNTHESIS

Polyketides are a heterogenous group of natural substances with a common biosynthesis by decarboxylative condensation of propionyl-CoA, methylmalonyl-CoA, malonyl-CoA or acetyl-CoA. However due to tailoring enzymes, which modify the polymer during and after the synthesis, a remarkable diversity of the final structure of polyketides is observed (Shen and Hutchinson, 1993, Figure 1.4). This group of compounds has been of interest for a long time due to their diversity and understanding how they are synthesized has been one of the many puzzles to be solved (Staunton and Weissman, 2001). One of the first theories was established by James Collie in 1893 by chance when studying the structure of dehydroacetic acid where he discovered unexpectedly orcinol, an aromatic product which he later explained of having been formed as polyketone intermediate as well as being generated in living cells (Collie and Myers, 1893). Unfortunately he was ahead of his time and his discoveries were buried over decades. Only in the 1950s, when Birch became interested in polyketides, research in the area was taken up again. He discovered that the precursors were acetate units that were linked by repeated condensation. He also established isotopically labelled precursor feeding to producing organisms (Birch *et al.*, 1955).

Polyketides are formed by polyketides synthases, a family of large enzymes called polyketide synthases (PKS), some that are multi-domain enzymes. The genes responsible for the biosynthesis are organised in clusters containing the biosynthetic, resistance as well as pathway specific regulatory genes (Martín and Gil, 1984).

There are different classes of PKS described which are type I, type II and type III. The type I are the best understood group and are represented by either modularly or iteratively organised enzymes. In the modular type each module is comprised of at least an acyl-carrier protein (ACP), an acyltransferase (AT) and a ketosynthase (KS). To release the final chain a thioesterase (TE) is needed too. There can be additional enzymes present, the tailoring enzymes, which modify the basic chain in each of the modules. Polyketides are formed by the condensation of acyl-CoA precursors in a series analogous to those occurring during fatty acid biosynthesis (Birch, 1968; O'Hagan, 1992; Jenke-Kodama and Dittmann, 2005; Ridley *et al.*, 2008). The best understood compound of this type is Erythromycin synthesized by *Saccharopolyspora erythraea* (Figure 1.5; Staunton, 1998). An example for an iteratively produced type I polyketide is lovastatin in filamentous fungi (Guo and Wang, 2014).

*S .coelicolor* A3(2) produces coelimycin P1, a yellow polyketide, that only recently had its structure elucidated together with a putative biosynthesis (Gomez-Escribano *et al.*, 2012). It is made by a type I PKS encoded by the genes of *cpkABC*. It was first described by knockout studies of the respective regulatory gene *scbR2* leading to an increased production of a yellow pigment with a maximum absorption at 460 nm in the culture (Gottelt *et al.*, 2010). However Gomez- Escribano and co-workers (2012) identified the yellow pigment not to be the actual product of the pathway but an intermediate containing 1,5-oxalthiocane, the actual metabolite is bis-epoxide which is the compound showing antibacterial properties.

Type II polyketide synthases are aggregates of monofunctional proteins such as in the case of actinorhodin, a blue pigmented antibiotic made by *Streptomyces coelicolor* A3(2) (Malpartida and Hopwood, 1986). All genes required for the biosynthesis are present in the same cluster in the genome. The compound backbone is made by condensation of one acetyl-CoA and seven malonyl-CoA by a minimal polyketide synthase of the type II. This basic backbone is cyclised into the three ring intermediate 4-dihydro-9-hydroxy-1-methyl-10-oxo-3-H-naptho-[2,3-c]-pyran-3-(S)-acetic acid ((s)-DNPA) and further modified to dihydrokalafungin (=DHK) by tailoring enzymes, two molecules of DHK are finally dimerised to form actinorhodin. Some genes in the cluster have yet to have a function assigned (Craney *et al.*, 2012). The pathway-specific regulator (*actII*-4) interacts with a number of global regulators connecting global stress response ppGpp and nutrient depletion with the onset of the production (Uguru *et al.*, 2005; Craney *et al.*, 2012). A schematic overview of the gene organisation of the actinorhodin cluster is shown in Figure 1.4 B.

The type III polyketide syntheses are characterised by a lack of ACP domain. An example for a type III product is germicidin from *Streptomyces* which acts as a self-germination inhibitor (Aoki *et al.*, 2011).

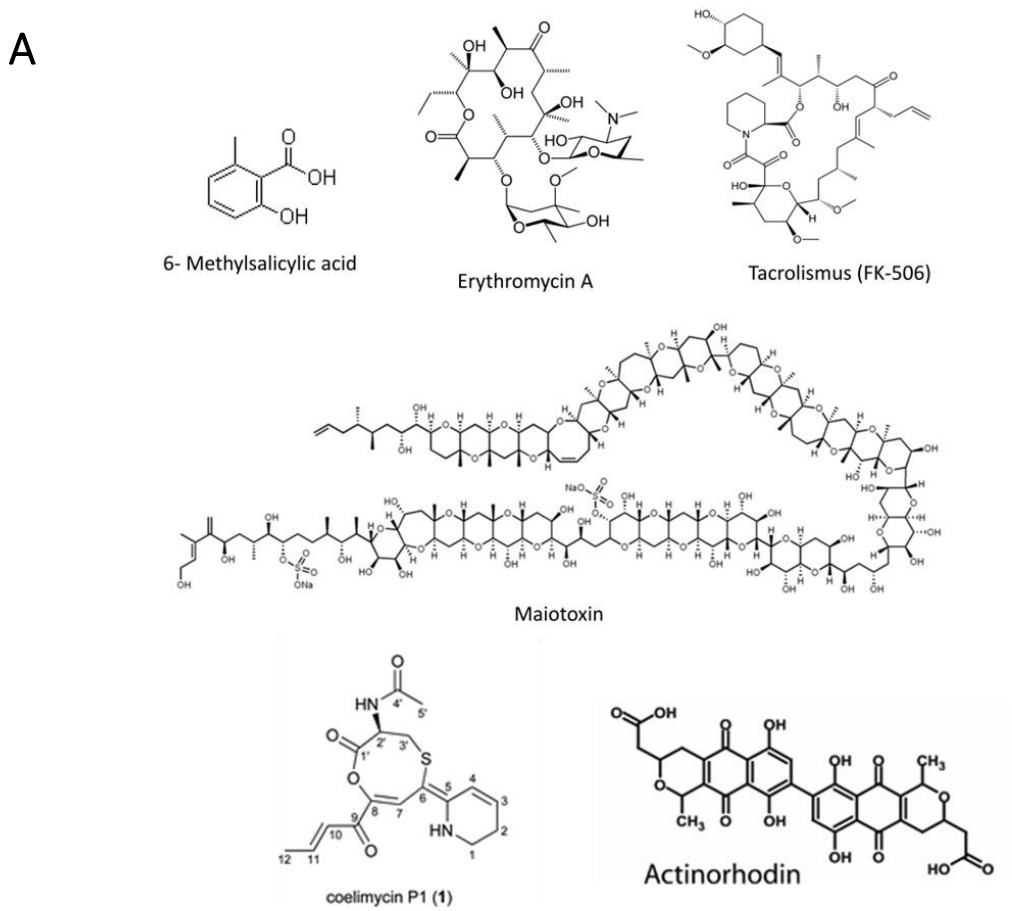



Figure 1.4 A Examples of different polyketides to illustrate the diversity of the group, they all however share the mechanism of their synthesis: 6 - Methylsalicylic acid is an intermediate in the biosynthesis of patulin in *Pencillium* (Forrester and Gaucher, 1972; Spencer and Jordan, 1992)- Erythromycin A is macrolide antibiotic used to treat a number of bacterial infections, the molecule is synthesized by *Saccharopolspora erythraea* (Dhillon *et al.*, 1989; Weissman and Leadlay, 2005) - Tarcolimus (FK-506) is an immunosuppressor produced by *Streptomyces tsukubaensis* (Hatanaka *et al.*, 1989; Turło *et al.*, 2012) -Malotoxin, a toxin produced by the dinoflagellate *Gambierdiscus toxicus* (Yokoyama *et al.*, 1988)- coelimycin, a yellow pigmented polyketide from *S. coelicolor* (Gomez-Escribano *et al.*, 2012) – actinorhodin, a blue pigmented polyketide from *S. coelicolor* (Craney *et al.*, 2012) B Genetic organisation of the biosynthetic cluster for actinorhodin in the genome modified after (Okamoto *et al.*, 2009) numbers indicate biosynthetic order,Legend for arrows: unfilled = function unknown, blue= tailoring enzymes, red= putative resistance genes, green – regulatory genes, yellow = minimal type II PKS

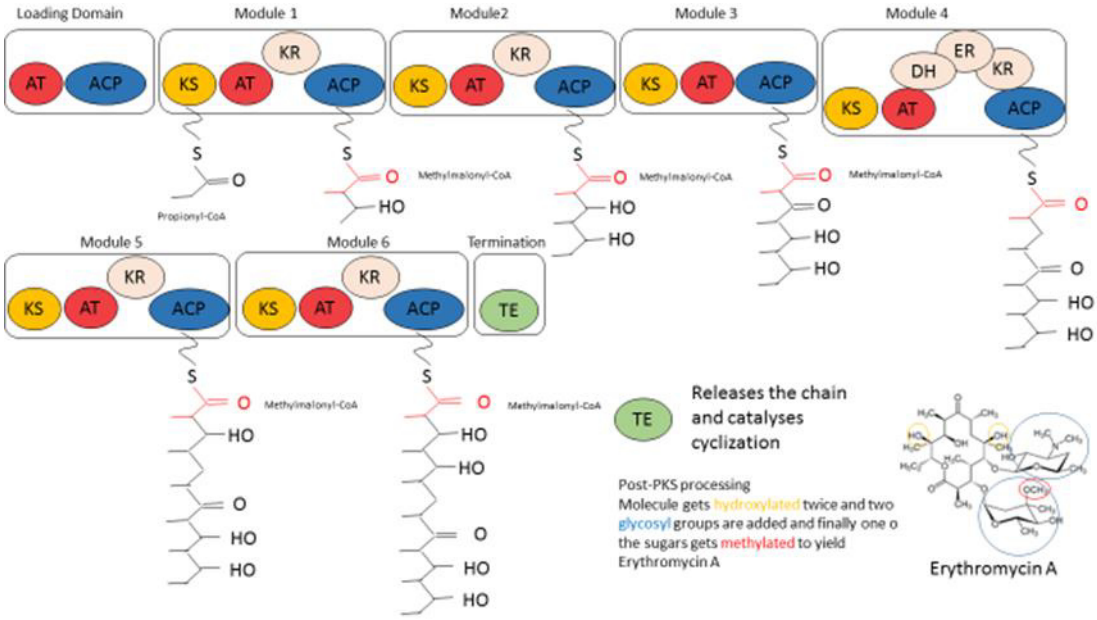



Figure 1.5 Simple illustration of polyketide synthesis of the highly modular type I: Erythromycin A biosynthesis. Modified after Weissman and Leadlay, 2005

A propionyl-CoA is the starter unit is bound at their carboxy group to the SH group of the acyl-carrier protein (ACP) by the acyltransferase (AT). This is then transferred to the ketosynthase (KS) of module 1. The following unit (methylmalonyl-CoA) is bound to the next modules ACP domain by the AT and reacts with the previous bound starter group by a Claisen condensation releasing $CO_2$, releasing the KS and an ACP-bound elongated polyketide chain. This process is repeated, there can be additional enzyme present such as ketoreductases (KR), dehydratases (DH) or enoylreductases (ER) that can modify the chain at each elongation step. This is where the huge diversity is derived from. To terminate the molecule another enzyme is required, a thiolesterase (TE) which hydrolyses the polyketides form the last ACP-domain and in this case also catalyses a cyclisation. After the initial chain is finished there can be further modification such as hydroxylation, glycosylation or methylation as in the case of Erythromycin. The principle process is highly similar to fatty acid synthesis and polyketide synthesis is believed of having evolved from it (Jenke-Kodama *et al.*, 2005) Legend: AT= acyltransferase, ACP=acyl carrier protein, KS= ketosynthase, KR= keto reductase, DH =dehydratase, ER = enoylreductase, TE= thiolesterase

## 1.3 Evolution of Primary and Specialised Metabolism

An interesting feature of streptomycete genomes is that there are often multiple genes annotated for the same biochemical function in central carbon metabolism (Bentley *et al.*, 2002; Hiltner *et al.*, 2015, Figure 1.2). Understanding this relationship between function and evolution is key to elucidating metabolic plasticity and exploiting this in biotechnology.

Surprisingly many primary metabolic enzymes are non-essential for survival due to genetic redundancy through the presence of isoenzymes or alternative reactions, that allow the cell to adapt to changing environmental conditions providing an adaptive robustness to metabolism (Kim and Copley, 2007). The use of the term 'redundancy' may be misleading as it suggests non-essentiality, yet the presence of multiple routes to a metabolic intermediate may be significant under a given set of environmental conditions, therefore 'contingency', 'metabolic flexibility' or 'enzyme expansion' may be more suitable terms to reflect this phenomenon (Treangen and Rocha, 2011; Noda-García and Barona-Gómez, 2013). Such terms would also account for so called 'moonlighting' enzyme functions and catalytic promiscuity where the main catalytic function is supplemented by the catalysis of additional reactions (Copley, 2003; Copley, 2012; Copley, 2014; Copley, 2015).

In metabolic models these homologous functions are often all combined into one flux pathway, yet this does not reflect the nuances of regulation and allostery for each gene product. However, it is necessary in order to understand the role of these gene expansions to appreciate functionality at the biochemical level. A hypothesis to explain these redundant functions would be that they might allow cellular flexibility and adaptation in dynamic environments.

Interestingly, given that streptomycetes exhibit so many of these gene expansions, it may also in part reflect their ability to produce such a vast variety of specialised metabolites, as these pathways are usually connected to core metabolism and use common intermediates. There are well-studied examples of such gene family expansions in streptomycete developmental gene and DNA-binding protein families (Clark and Hoskisson, 2011; Girard *et al.*, 2013) however little attention has been paid to central metabolism. An example of gene expansion and evolution relating to specialised metabolite biosynthesis is that of polyketides where it appears the multi-modular and iterative polyketide synthases (PKS), which perform the stepwise condensation activated carboxylic acid subunits to provide carbon backbones

for polyketides have evolved from fatty acid synthases (FAS; (Jenke-Kodama *et al.*, 2005). This is based on the presence of highly conserved modules such as the keto-acyl synthase (KS) domains and acyl carrier proteins (ACP), which through duplications, deletion and horizontal gene transfer (HGT) have lead to the diverse chemistry of polyketides that is observed today (Jenke-Kodama *et al.*, 2005; Ridley *et al.*, 2008).

The nature of these gene expansions also requires consideration as it is generally assumed that these expansions arise through gene duplications, however little consideration has been paid to the role of HGT in metabolic gene expansion (Noda-García *et al.*, 2013; Noda-García and Barona-Gómez, 2013). This reflects the idea that gene duplication is an important source of biological innovation, where orthologous genes exhibit conserved functionality, and paralogs tend towards diverged function. However integrating the role of HGT in this process requires sophisticated tools to identify HGT events in metabolic genes, coupled with thorough studies of enzyme function to tease out the detailed mechanisms. Recently, Noda-Garcia showed that horizontally acquired metabolic genes could drive evolution of existing metabolic function through altered substrate specificity, and this may reflect a common, but under-appreciated mechanism for enzyme expansion in prokaryotic genomes (Noda-García and Barona-Gómez, 2013).

This is a developing area of interest that requires integration of a range of techniques in genomics, molecular genetics, biochemistry, X-ray crystallography, molecular dynamics simulations, and evolutionary modelling, but offers great potential for deep understanding of evolution of enzyme function and how this has contributed to metabolic plasticity. Ultimately studies such as these can be valuable for informing metabolic engineering for biotechnology.

## 1.4 THE PEP-PYR-OAA NODE OF *STREPTOMYCES COELICOLOR* A3 (2)

This metabolic node represents a major flux distribution point for carbon skeletons in the cell with the key metabolites being phosphoenolpyruvate (PEP), pyruvate (PYR) and oxaloacetate (OAA). Pyruvate can be metabolised further into either OAA or Acetyl-CoA, the former being a precursor for amino acids such as aspartate, lysine, methionine, threonine and isoleucine and the latter for fatty acids and polyketides. The architecture of this node in a range of species can be highly variable and likely reflects the ecology of individual organisms (Figure 1.6). Many key precursors for specialised metabolites such as polyketides are derived from

here or pathways are limited by flux through this node, which is surprisingly diverse among bacteria. Sauer and Eikmans (2005) examined this node in *Escherichia coli*, *Corynebacterium glutamicum* and *Bacillus subtilis* in their role as major workhorses for the industrial production of bulk chemicals such as amino acids, organic acids or proteins. These authors concluded that this node of metabolism is a key target for metabolic engineering in bacteria (Sauer and Eikmanns, 2005). A summary of the reaction of each enzyme of the metabolic node in *S. coelicolor* with their gene ID, their UniProt ID is given in Table 1-2 and a PFAM domain description for each enzyme is given in Figure 1.7.

In comparison to other bacteria several genes for the same reaction are present in *S. coelicolor*. Out of the ten reactions, there are seven that have at least two genes or more, only PEP carboxylase, PEP carboxykinase and Pyr carboxylase have a single gene in the *S. coelicolor* genome.

Phosphoenolpyruvate carboxylase (PEPCx; SCO3127, EC 4.1.1.31) condenses Phosphoenolpyruvate (PEP) and $CO_2$ to form oxaloacetate (OAA) and orthophosphate. The enzyme can be stimulated by acetyl-CoA, nucleoside triphosphates and succinate. Aspartate and citrate can act as inhibitors and its activity is increased during actinorhodin production (Vorísek *et al.*, 1970). It was therefore tested by the same authors if it could also convert malonyl-CoA, a precursor of actinorhodin, however, no evidence for this could be detected in cell free extracts (Birch, 1968; O'Hagan, 1992; Bramwell *et al.*, 1993).

PEP carboxykinase (PEPCk; SCO4979, EC 4.1.1.32) is a gluconeogenic enzyme transforming OAA using the phosphate from a GTP molecule to yield PEP and $CO_2$. The enzyme requires $Mn^{2+}$ or $Mg^{2+}$ for optimal activity. It is monomeric, 68 kDa in size in *C. glutamicum* and shows almost 58 % sequence identity at the amino acid level (Aich *et al.*, 2003). The enzyme is essential for growth on lactate and acetate in *C. glutamicum* (Riedel *et al.*, 2001).

Pyruvate carboxylase (PCx; SCO0546, EC 6.4.1.1) catabolises the reaction of pyruvate and hydrogencarbonate to form OAA while consuming energy in form of ATP and was first described in 1963 (Utter and Keech, 1963). PCx plays an important role in the replenishment of OAA levels for the TCA (Sauer and Eikmanns, 2005). The enzyme is biotinylated and uses the biotin to transfer the carboxyl group (Owen *et al.*, 2002). L-aspartate and α- ketoglutarate inhibit the activity in a non-competitive fashion and molecules like AMP, ADP and pyruvate

competitively inhibit PCx (Dunn *et al.*, 2001). Acetyl-CoA however shows a positive effect on the enzyme activity (Modak and Kelly, 1995).

Malic enzyme (ME; SCO2951 and SCO5261, EC 1.1.1.38) is also an OAA decarboxylase which catalyses irreversible decarboxylation of OAA to pyruvate and $CO_2$. An additional activity of decarboxylating malate to pyruvate via reduction of NAD to NADH is described (Klaffl and Eikmanns, 2010). It has already been investigated that the two genes present in the *S. coelicolor* genome have different cofactors, SCO2951 is $NAD^+$ dependent, whereas SCO5261 is $NADP^+$ dependent (Rodríguez *et al.*, 2012).

Pyruvate dehydrogenase or pyruvate oxidase (SCO6155 and SCO7412, EC 1.2.5.1) catalyses the reaction of pyruvate and ubiquinone to acetate with $CO_2$ and ubiquinol. It is located on the inner surface of the cytoplasmic membrane and is coupled to the respiratory chain via ubiquinone (Cunningham and Hager, 1975; Koland *et al.*, 1984). In *E. coli* the enzyme requires thiamine diphosphate and FAD (O'Brien *et al.*, 1976).

Pyruvate kinase, PpdK and PDHC are explained in more detail in separate sections below. They all are connected to the metabolite pyruvate, one of the key metabolites of the node and an important precursor for polyketide synthesis.

Table 1-2 Overview of the enzymes involved in the PEP-PYR-OAA node with name, EC number, Gene ID, UniProt ID and the reaction they catalyse, the reactions of the PDHC are coupled with each other Source: KEGG database (http://www.genome.jp), Legend: PEP Phosphoenolpyruvate, OAA Oxaloacetate, MAL Malate, PYR Pyruvate, ACE Acetate, HeThdP 2-(α-Hydroxyethyl)thiamine diphosphate, ThdP Thiamin diphosphate

| **Enzyme (EC number)** | **Gene ID** | **UniProt ID** | **Reaction** |
|---|---|---|---|
| **PEP carboxylase (EC 4.1.1.31)** | SCO3127 | Q9RNU9 | $PO_4^{3-}$ + OAA ←→ $H_2O$ + PEP + $CO_2$ |
| **PEP carboxykinase (EC 4.1.1.32)** | SCO4979 | Q93JL5 | GTP + OAA ←→ GDP + PEP + $CO_2$ |
| **Malic enzyme (EC 1.1.1.38)** | SCO2951<br>SCO5261 | Q9L1U4<br>Q9F3K4 | (S)-MAL + NAD + ←→ PYR + $CO_2$+ NADH + $H^+$<br>OAA ←→ PYR + $CO_2$ |
| **Pyr carboxylase (EC 6.4.1.1)** | SCO0546 | Q9RK64 | ATP + PYR + $HCO_3$- ←→ ADP + $PO_4^{3-}$ + OAA |
| **Pyr phosphate dikinase (EC 2.7.9.1)** | SCO0208<br>SCO2494 | Q9RI40<br>Q9L2C5 | ATP + PYR + P ←→ AMP + PEP + diP |
| **Pyruvate kinase (EC 2.7.1.40)** | SCO2014<br>SCO5423 | Q9L299<br>Q9S2I9 | ATP + PYR ←→ ADP + PEP |
| **$PDHC_{E1}$ (EC 1.2.4.1)** | SCO2183<br>SCO2371<br>SCO7124 | Q9S2Q3<br>Q9KY19<br>Q9FC62 | PYR + ThdP ←→ HeThdP +$CO_2$ |
| **$PDHC_{E1\alpha/\beta}$ (EC 1.2.4.1)** | SCO1269/70<br>SCO3816/17<br>SCO3830/31 | Q9K3H1/ Q9K3H0<br>Q9XA61/ Q9XA60<br>Q9XA48/ Q8CJV5 | HeThdP + Lipoamide-E ←→ S-Acetyldihydrolipoamide-E + ThdP |
| **$PDHC_{E2}$ (EC 2.3.1.12)** | SCO3815<br>SCO3829<br>SCO7123 | Q9XA62<br>Q9XA49<br>Q9FC63 | Acetyl-CoA + Lipoamide-E <=> CoA + S-Acetyldihydrolipoamide-E |
| **$PDHC_{E3}$ (EC 1.8.1.4)** | SCO0884<br>SCO2180<br>SCO4919 | Q9RD26<br>Q9S2Q6<br>Q9EWF0 | PYR + CoA + NAD+ ←→ Acetyl-CoA + $CO_2$+ NADH + $H^+$ |
| **Pyruvate dehydrogenase (EC 1.2.5.1)** | SCO6155<br>SCO7412 | Q9ZBT3<br>Q9L147 | PYR + Ubiquinone + $H_2O$ ←→ ACE + Ubiquinol + $CO_2$ |

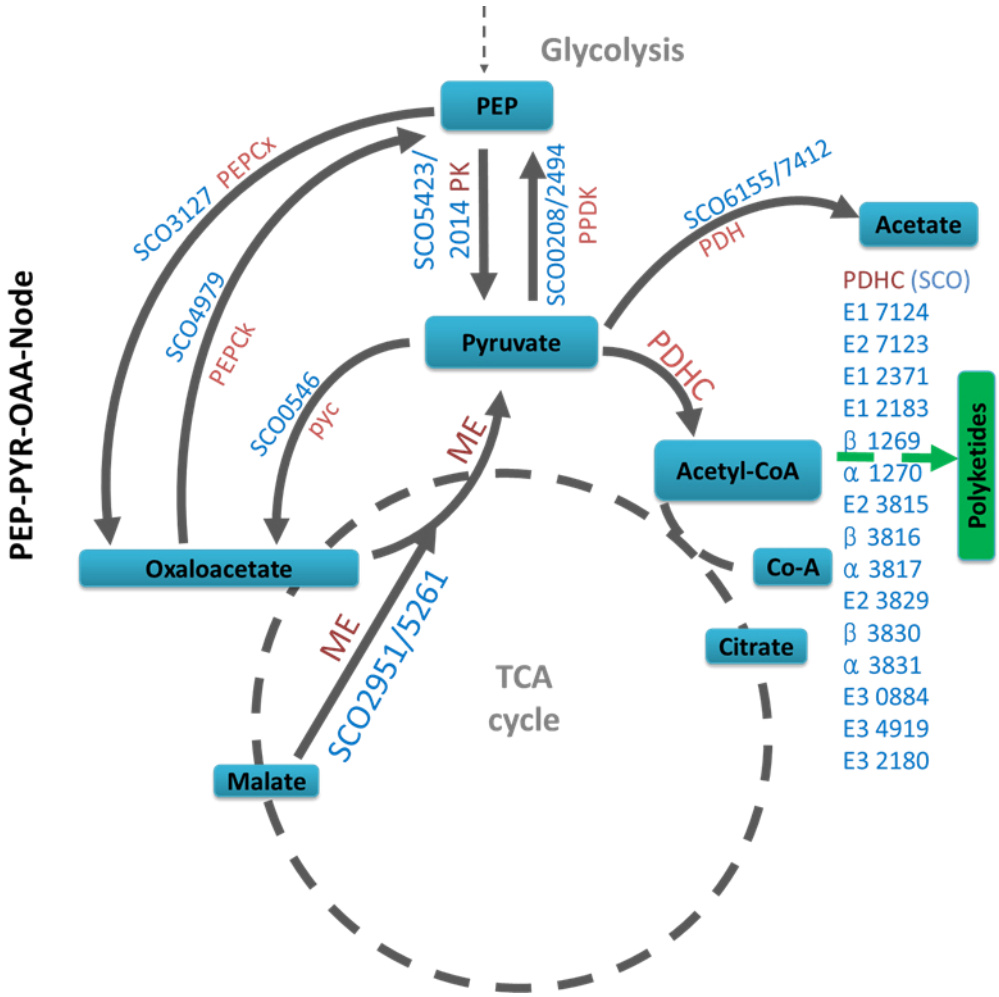



Figure 1.6 General overview of the reactions that form the Phosphoenolpyruvate-Pyruvate-Oxaloacetate Node with gene identifier of *Streptomyces coelicolor* genome. Legend: SCO3127, PEPCx EC 4.1.1.31, SCO4979 PEPCK EC 4.1.1.32, SCO2951/5261 ME EC 1.1.1.38, PCx SCO0546 EC 6.4.1.1, SCO0208/2494 PpdK EC 2.7.9.1, SCO2014/5423 PYK EC 2.7.1.40, SCO2183/2371/7124/1269/1270/3816/3817/3830/3831 $PDHC_{E1}$ EC 1.2.4.1, SCO3815/3829/7123 $PDHC_{E2}$ 2.3.1.12, SCO0884/2180/4919 $PDHC_{E3}$ EC 1.8.1.4, SCO6155/7412 Pyruvate dehydrogenase EC 1.2.5.1 , also see Table 1-2

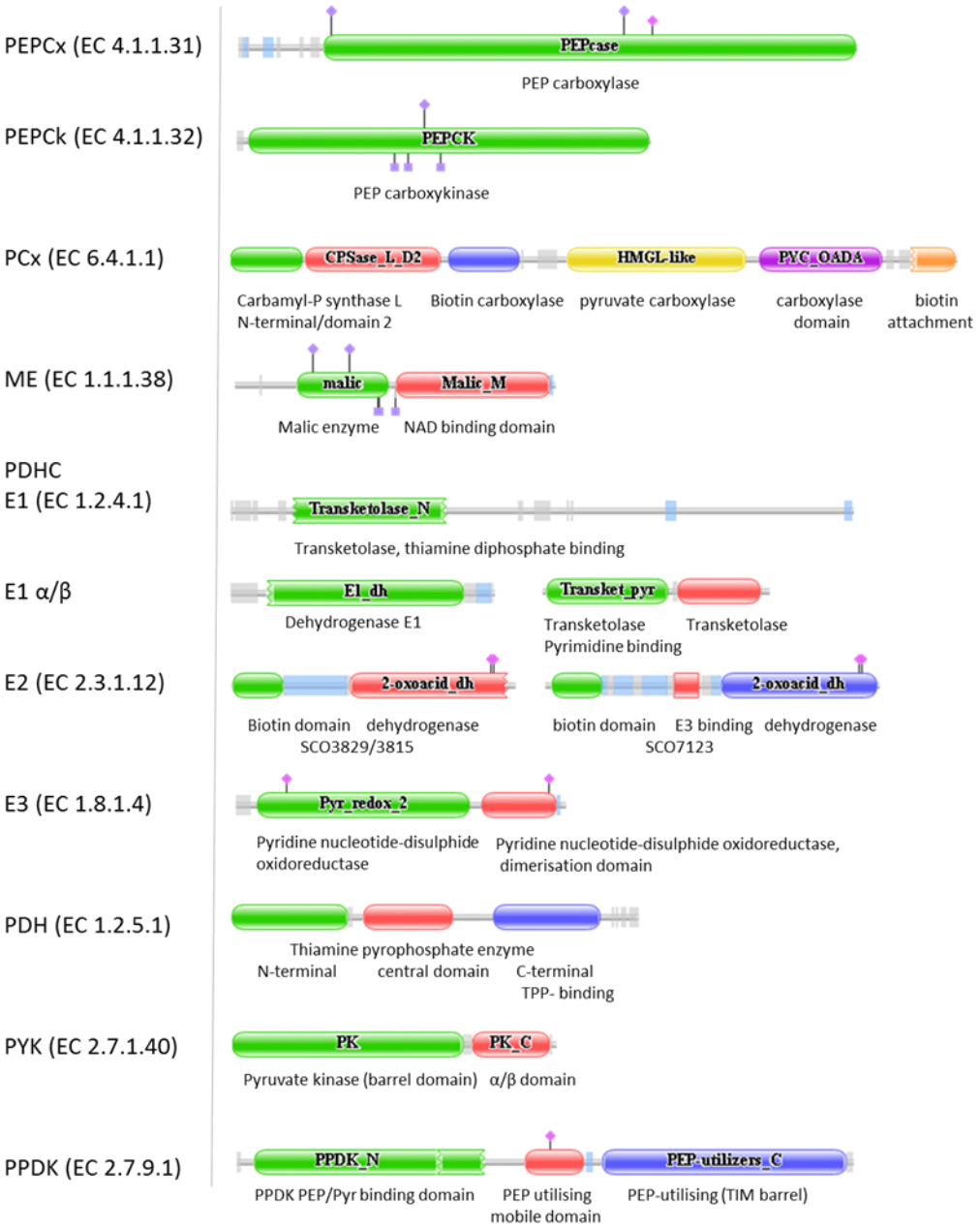



Figure 1.7 PFAM protein domain overview of the enzymes of the PEP-PYR-OAA-Node. Legend: PEPCx = PEP carboxylase, PEPCk = PEP carboxykinase, PCx = Pyruvate carboxylase, ME = malic enzyme, PDHC = pyruvate dehydrogenase complex, PDH = pyruvate dehydrogenase, PYK = pyruvate kinase, PpdK = pyruvate phosphate dikinase

### 1.4.1 PYRUVATE KINASE

Pyruvate kinase, one of the key enzymes in the glycolysis catalyses the irreversible transfer of a phosphate from phosphoenolpyruvate to ADP to form pyruvate and ATP (Kayne and Price, 1973). *S. coelicolor* has two pyruvate kinases annotated in its genome, which encode Pyk1 (SCO2014) and Pyk2 (SCO5423). A divalent cation is required by pyruvate kinase to exhibit activity. Under physiological conditions $Mg^{2+}$ is optimal. However, $Co^{+2}$, $Mn^{+2}$, and $Ni^{+2}$ can also work (Nowak and Suelter, 1981; Jurica *et al.*, 1998).

Additionally the enzyme requires $K^+$, which is involved in the formation of the active form of the enzyme. Without $K^+$ the reaction is ordered with PEP being its first substrate and only being able to bind ADP once PEP has bound. In the presence of $K^+$ both substrates can bind independently to the enzyme (Oria-Hernández *et al.*, 2005). However there are enzymes that do not require $K^+$ for activity. The dependency has been analysed phylogenetically and the dependency and linked to the presence of either glutamine or lysine at position 117, where glutamine is indicative for $K^+$ dependency and lysine for independency (Oria-Hernández *et al.*, 2006). However the *Streptomyces* Pyk do not show either amino acid at the position, but glycine in Pyk1 and alanine in Pyk2.

The enzyme usually shows allosteric properties for PEP (Mattevi *et al.*, 1995) and common effectors are F1,6BP in mammalian cells and *E. coli*, F2,6PB in protozoans (van Schaftingen *et al.*, 1985; Morgan *et al.*, 2014), some for R5P (Valentini *et al.*, 2000) or AMP for PykA of *E. coli*, as well as for Pyk in *S. aureus, G. stearothermophilus* and *C. glutamicum* (Jetten *et al.*, 1994; Mattevi *et al.*, 1995; Suzuki *et al.*, 2008; Zoraghi *et al.*, 2011).

A range of structures has been determined, most common is a homotetramer (Kayne and Price, 1973; Schramm *et al.*, 2000), but also homodimer (Pawluk *et al.*, 1986), monomer (Knowles *et al.*, 1989) or heterohexamer (Plaxton *et al.*, 1990; Plaxton, 1991; Negm *et al.*, 1995).

The protein consists of three or four domains. The N-terminal is not present in bacteria, however domain A, B and C are present in all. The active site is located between the A and B domain and the effector binds to the C domain. *G. stearothermophilus* and *S. aureus* additionally show a C' domain (Morgan *et al.*, 2014). Both Pyk1 and Pyk2 from *S. coelicolor* lack this domain, so do PykA and PykF from *E. coli*. In the final structure adjacent C domains form the small interface and adjacent A domains the large interface and the B domain forms

a mobile 'lid' at one end of the α/β barrelled A domain, like a door to the active site (Morgan *et al.*, 2014).

It has been reported recently that the tetrameric structure of Pyk in *Trypanosoma* and *Leishmania* is stabilized by 8˚ rigid body motion of A and C domain in each subunit as well as salt bridges across two of the four subunits leading to enhanced tetramer stability and increased specificity constant ($K_{cat}/S_{0.5}$). $K_{cat}$ is the substrate turnover number meaning the number of substrate molecules converted per active site per unit of time. $S_{0.5}$ is defined as the substrate concentration at which half of the maximum velocity is reached. Maximum velocity is referring to all active sites are occupied (Cornish Bowen, 2012). However the structure is different in human and bacterial pyruvate kinases, which have different allosteric regulators (Morgan *et al.*, 2014).

Furthermore this enzyme plays a role in controlling the carbon flux of glycolysis (Allert *et al.*, 1991; Valentini *et al.*, 2000). It is ubiquitous throughout eukaryotes and prokaryotes, but it has been reported that it can be 'replaced' by the enzyme pyruvate phosphate dikinase (PpdK, Saavedra-Lira *et al.*, 1998). Many bacteria have two isoenzymes. However a single gene is not uncommon in bacteria too. Plants have a plastid and a cytoplasmic pyruvate kinase and in mammalian cells four tissue specific pyruvate kinases (L (liver), R (erythrocyte), M1 and M2 (muscle isoenzymes) can be found (Muñoz and Ponce, 2003).

The best bacterial characterisation of pyruvate kinase has been carried out for *E. coli*. Two genes are annotated in the genome for pyruvate kinase and the resulting enzymes differ in their allosteric regulation: type I (PykF) is activated by Fructose-1.6-bisphosphate and type II (PykA) by AMP (Valentini *et al.*, 2000). This allows an effective switching of glycolysis and gluconeogenesis. When considering this enzyme as a metabolic engineering target, it has to be considered that when grown on non-fermentable sugars, the cell requires gluconeogenesis and PEP is generated from pyruvate via oxaloacetate. It might therefore be necessary to block pyruvate kinase in order to prevent futile cycling of PEP and pyruvate (Muñoz and Ponce, 2003). In *E. coli* it shows allosteric binding properties for PEP and requires monovalent cations for its catalytic activity (Muñoz and Ponce, 2003). Furthermore it can occur in two different states, the inactive T-state and active R-state. It has been shown that the inactivation of pyruvate kinase can increase the flux into a particular metabolite (Ponce *et al.*, 1995; Ponce *et al.*, 1998; Emmerling *et al.*, 1999).

### 1.4.2 PYRUVATE PHOSPHATE DIKINASE

Pyruvate phosphate dikinase (PpdK) is a phosphotransferase (EC 2.7.9.1) and a key enzyme in gluconeogenesis. It was first described in the leaves of maize, sugar cane and sorghum, where it plays a role in C4 catabolism and gluconeogenesis (Hatch and Slack, 1969). It was independently discovered in *Entamoeba histolytica*, an anaerobic unicellular parasitic protozoan, which lacks Pyk and the suggestion was that this enzyme replaces the function of Pyk (Reeves, 1968). The *Streptomyces* genome has two genes annotated with this function SCO0208 (*ppdK*1) and SCO2494 (*ppdK*2).

PpdK converts pyruvate, ATP and inorganic phosphate into AMP, PEP and diphosphate. The reaction is reversible in contrast to the pyruvate kinase reaction and occurs in a three step process. PpdK first binds to ATP and converts it to AMP resulting in a diphosphorylated PpdK. The diphosphorylated enzyme then binds to inorganic phosphate to produce diphosphate and leaving a mono-phoshorylated PpdK which then binds to pyruvate and transfers its phosphate onto it to form PEP and regenerating the PpdK (Evans and Wood, 1968). It requires divalent cations such as $Co^{2+}$, $Mn^{2+}$, $Ni^{2+}$ or $Mg^{2+}$ for its activity (Reeves, 1971).

The enzyme is most abundant in bacteria and plants and is absent in vertebrates with a few exceptions (Feng *et al.*, 2008). A gene search on NCBI Gene resulted in a total of 1092 hits, of which 292 were in plants, but also 63 in animals, 16 in fungi, 359 hits in bacteria, 267 in archae, 8 in viruses and 87 in other organisms such as *Entamoeba*, kinetoplastids, oomycetes and trichomonads for example (NCBI Gene Database, http://www.ncbi.nlm.nih.gov/gene , revised on 30.07.2015).

In the literature the focus has mainly been as a potential drug target to treat parasitic diseases (Wu *et al.*, 2013). It is the preferred catabolic reaction in *Trichomonas tenax*, an archae and is modulated by ADP and central carbon metabolism (Tjaden *et al.*, 2006). Glycolytic flux in *Phytophthora cinnamomi* is mainly run by PpdK in the sporulation hyphae (Marshall *et al.*, 2001) and it has been reported to play a role in early development of rice seeds (Chastain *et al.*, 2006). It has also been linked to photosynthesis and stress in plant roots. Its activity can be regulated by ABA, mannitol and PEG (Moons *et al.*, 1998; Feng *et al.*, 2008). Additionally to the two genes annotated as PpdK there is another enzyme PEP synthetase or pyruvate water dikinase which forms PEP from pyruvate, but by a different mechanism. In the genome of *S. coelicolor* one gene is annotated with this function: SCO5896.

### 1.4.3 PYRUVATE DEHYDROGENASE COMPLEX

The pyruvate dehydrogenase complex is composed of three different enzymes, simply named $PDHC_{E1}$ (pyruvate dehydrogenase, 1.2.4.1), $PDHC_{E2}$ (dihydrolipoamide acetyltransferase, 2.3.1.12) and $PDHC_{E3}$ (dihydrolipoamide dehydrogenase 1.8.1.4) catalysing the conversion of pyruvate into acetyl-CoA in a multistep process (de Kok *et al.*, 1998; Neveling *et al.*, 1998; Ren *et al.*, 2011). $PDHC_{E1}$ catalyses the irreversible oxidative decarboxylation of 2-oxoacid followed by the aceylation of lipoyl prosthetic group attached to the $PDHC_{E2}$. $PDHC_{E2}$ then catalyses the transfer of the acyl group from the lipoyl group to CoA resulting in a dihydrolipoyl group. $PDHC_{E3}$ reoxidises the latter and generates $NADH+H^+$ and NAD (Figure 1.8). The $PDHC_{E3}$ is shared with several enzyme complexes, such as the oxoglutarate dehydrogenase, acetoin cleaving system and the α-keto acid dehydrogenase complexes (Schreiner *et al.*, 2005). Analysis of the general gene organisation in different organisms shows that Gram-negative bacteria have three different polypeptides involved forming an octahedral structure containing 12 $PDHC_{E1}$ (homodimer), 24 $PDHC_{E2}$ and 6 $PDHC_{E3}$ per complex. Whereas for Gram-positive bacteria the general structure is of 4 different polypeptides forming an icosahedral symmetric complex. It is comprised of 30 E1 components (heterotetramer 2 $PDHC_{E1\alpha}$ and 2 $PDHC_{E1\beta}$), 60 $PDHC_{E2}$ and 6 $PDHC_{E3}$ (Neveling *et al.*, 1998). In Gram-negative bacteria the $PDHC_{E1}$ and $PDHC_{E2}$ are most often found in the same operon, whereas $PDHC_{E3}$ which is shared with complexes can be in a different location. *Streptomyces* $PDHC_{E1}$ is generally located on its own with no other genes of the complex. Other Gram-positive bacteria show a cluster of $PDHC_{E1\alpha}$ and $PDHC_{E1\beta}$ subunit often together with $PDHC_{E2}$and $PDHC_{E3}$ as well. Clustering of $PDHC_{E1\alpha}$, $PDHC_{E1\beta}$ and $PDHC_{E2}$ is also observed in *Streptomyces*, but not $PDHC_{E3}$, even though it has to be said that generally one $PDHC_{E3}$ homologues was in close proximity to the cluster. The bacterium *Zymomonas mobilis* shows a different organisations from those described above, where there is a cluster of $PDHC_{E1}$ subunits and a second cluster of the $PDHC_{E2}$ and $PDHC_{E3}$ gene. Generally the gene organisation in eukaryotes is as in Gram-positive bacteria. *Streptomyces* shows similarities to both Gram-positive and Gram-negative forms but there are differences: $PDHC_{E1}$ is usually on its own (as in Gram-negatives) and $PDHC_{E3}$ is not clustered with $PDHC_{E2}$ in any case, as is normal in Gram-positives (Figure 1.8). Other important aspects of the complex are, that pyruvate is a positive effector and acetyl-CoA is a negative effector of activity, while NADH causes feedback inhibition, with the $PDHC_{E1}$ component is described as the rate limiting step of the reaction (de Kok *et al.*, 1998).

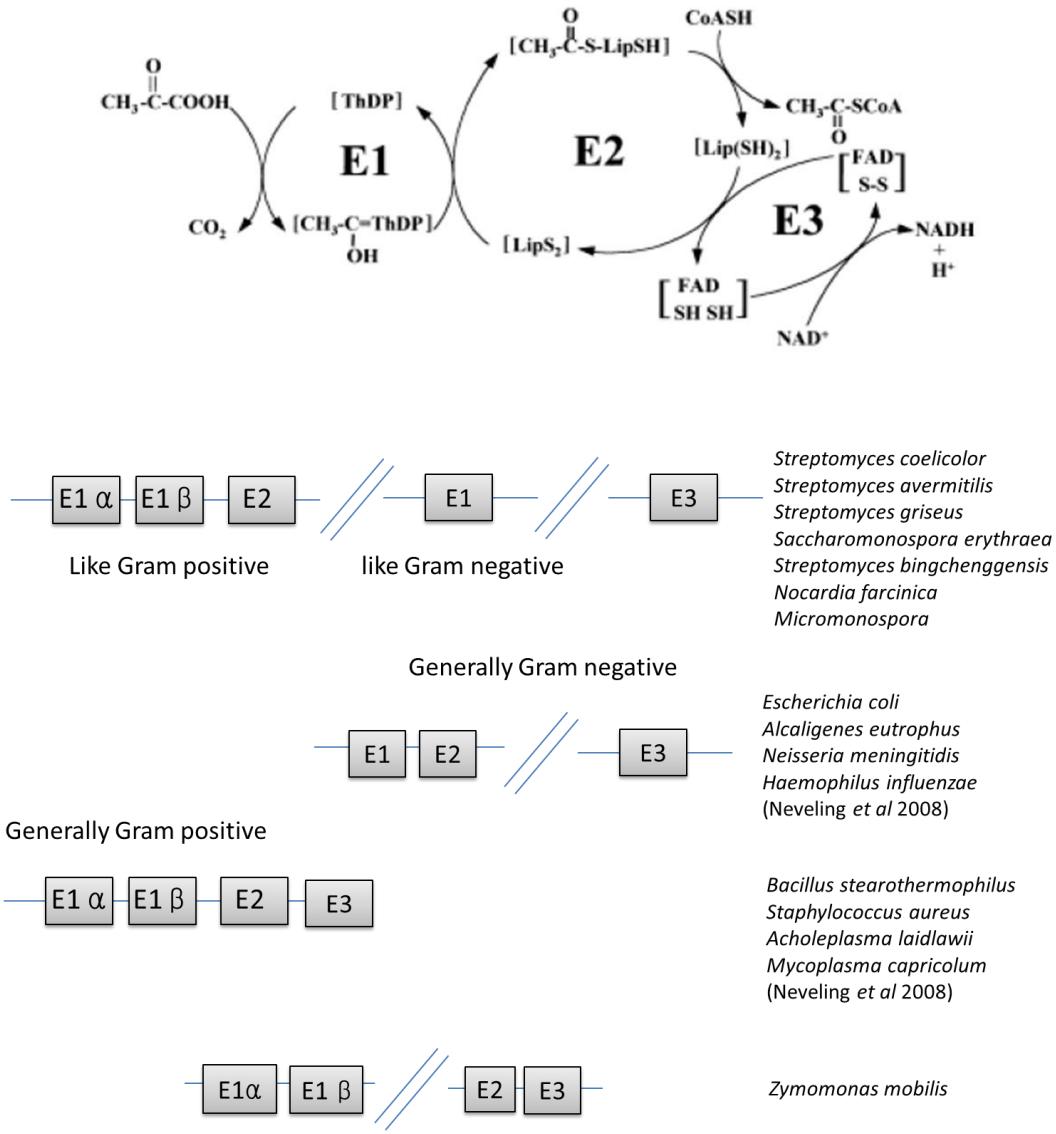



Figure 1.8 Top: Schematic overview of PDHC reaction catalysed by $PDHC_{E1}$, $PDHC_{E2}$ and $PDHC_{E3}$ from (de Kok *et al.*, 1998). E1 catalyses the irreversible oxidative decarboxylation of 2-oxoacid followed by the aceylation of lipoyl prosthetic group attached to the $PDHC_{E2}$. $PDHC_{E2}$ then catalyses the transfer of the acyl group from lipoyl group to CoA resulting in a dihydrolipoyl group. $PDHC_{E3}$ reoxidises the latter and generates $NADH+H^+$ and NAD. Bottom: Schematic overview of gene organisation of genes annotated for the enzyme complex

### 1.4.4 THE PEP-PYR-OAA-NODE AS METABOLIC ENGINEERING TARGET

Manipulating primary metabolic pathways in the PEP-PYR-OAA node can lead to higher product yields in a range of bacterial species. In *Corynebacterium glutamicum*, a well-studied amino acid producing organism, it is well established that this node is crucial for amino acid production. Much can be learned from studying the PEP-PYR-OAA node in this organism in terms of how the basal node functions. PEP carboxylase (PEPC) is not essential for lysine production and has no effect on the growth rate when deleted. However a mutant lacking both PEPC and pyruvate kinase (Pyk) show decreased growth rates as well as reduced lysine production rates (Koffas and Stephanopoulos, 2005). Moreover, studies that show that inducing low expression levels of *aceE* ($PDHC_{E1}$) through promoter exchange and deletion of *pqo* (pyruvate:quinone oxidoreductase) and *ppc* (PEP carboxylase) leads to an increase in L-valine production (Buchholz *et al.*, 2013). It was also shown that L-lysine production could be increased by reduced PDHC levels indicating that this enzyme activity levels has an influence on the carbon flux and can increase pyruvate derived molecules (Blombach *et al.*, 2009; Buchholz *et al.*, 2013; Eikmanns and Blombach, 2014).

The PEP carboxykinase (PEPCk) gene in *C. glutamicum* also influences the production of glutamate and lysine. Inactivation leads to an increase in production whereas overexpression of PEPCk decreases production (Riedel *et al.*, 2001). This indicates that blocking certain routes within metabolism can increase flux in diverse pathways. Disrupting these processes can also be mediated through mutating the transcriptional regulators. For example, deletion of *pckR* (Cg0196), a negative repressor of PEPCk during growth on glucose, results in cellular PEPCk activity even in the presence of glucose (Hyeon *et al.*, 2012).

Overexpression of pyruvate carboxylase (*pyc*) in *C. glutamicum* results in increased PCx activity and glutamate production in an optimised lysine-producing strain. However inactivation of *pyc* resulted in no PCx activity and lower glutamate production level (Peters-Wendisch *et al.*, 1998). A similar effect was observed for threonine production and its precursor homoserine, pointing towards the importance of PCx in amino acid metabolism and industrial production (Peters-Wendisch *et al.*, 2001). The deletion of pyruvate kinase (*pyk*) in a lysine over-producing strain showed similar growth rates but had higher rates of overflow metabolites such as dihydroxyacetone and glycerol as well as a shift from pyruvate carboxylase to phosphoenolpyruvate carboxylase flux on glucose which enabled a metabolic

bypass via malic enzyme to account for deletion of *pyk* (Becker *et al.*, 2008). This highlights the level of metabolic flexibility that this node provides in central carbon metabolism.

The Gram-negative bacterium *E. coli* has two genes encoding for pyruvate kinase; *pykA* and *pykE* (Muñoz and Ponce, 2003), and disruption of a single pyruvate kinase (*pykA*) in a phenylalanine overproducing strain showed that in combination with inactivation of a PTS sugar transporter lead to redistribution of cellular carbon flux when glucose is the substrate. PTS and PykA (which is allosterically regulated by AMP) inactivation resulted in decreased PCx, PEPCk and TCA cycle activity whereas PTS and PykF inactivation (which is allosterically regulated by Fructose 1,6 biphosphate) resulted in increased OAA formation from PEP and flux through the TCA cycle. Interestingly, both strains showed increased phenylalanine production compared to the parental strain (Meza *et al.*, 2012). The presence of two pyruvate kinases in *E. coli* compared to *C. glutamicum* and the results reported by (Meza *et al.*, 2012) indicates that dissecting these aspects of enzyme expansion in metabolism are difficult to tease apart. In the streptomycetes, this is especially true when up to six copies of some enzymes within the PEP-PYR-OAA node are present.

The PEP-PYR-OAA node is poorly understood in streptomycetes, but some primary metabolic enzymes that have undergone gene expansion in *Streptomyces* have been studied in more detail. One such example is the presence of three copies of the glycolytic enzyme phosphofructokinase (SCO1214 - *pfkA*3; SCO2119 - *pfkA*1; and SCO5426 - *PfkA*2) catalysing the addition of a second phosphate to fructose-6-phosphate at the C1 position. Deletion of *pfkA*2 leads to an increase of undecylprodigiosine and actinorhodin production on defined medium with $NH^{4+}$ as nitrogen source and glucose as carbon source, whereas deletions of *pfkA*1 and *pfkA*3 do not show the same phenotype. The *pfkA*2 mutant also resulted in higher intracellular pools of glucose-6-phosphate and fructose-6-phosphate and radiolabelling experiments indicate increased flux through the pentose phosphate pathway (PPP) and concomitant increased levels of NADPH showing that despite having three genes encoding the same function they have different physiological roles (Borodina *et al.*, 2008). These data also correlate with higher production of methylenomycin and increased PPP flux during slow growth (Obanye *et al.*, 1996).

In *S. clavuligerus* two glyceraldehyde-3-phosphate dehydrogenase (GAPDH) genes are present in the genome, *gap1* and *gap2*. Disruption of *gap1* leads to an increase in clavulanic acid production, but not when *gap2* is deleted. Since clavulanic acid biosynthesis starts with

the condensation of L-arginine and glyceraldehyde-3-phosphate (G3P), this downstream block of *gap1* appears to redirect flux towards clavulanic acid biosynthesis rather than glycolysis. Furthermore it demonstrates the different physiological roles played by the different isoforms GAPDH (Li and Townsend, 2006).

In the model streptomycete *S. coelicolor,* disruption of either of the genes encoding the two isoforms of glucose-6-phosphate dehydrogenase *zwf1* (SCO6661), *zwf2* (SCO1937) leads to changes in the production of the polyketide actinorhodin. Again, two genes encoding the same function seem to play different roles, it appears that *zwf2* plays a more important role than *zwf1* for directing the flux towards actinorhodin production (Butler *et al.*, 2002; Ryu *et al.*, 2006). Similarly it has also been shown that disruption of *zwf* in *S .lividans*, which also has two copies yields a higher actinorhodin and undecylprodigiosin production (Butler *et al.*, 2002). Further studies indicate that the flux of carbon was dependent on growth rate and the carbon source in chemostat culture of *S. lividans* (Rossa *et al.*, 2002). When gluconate was utilized as carbon source a higher flux through PPP was observed than on glucose and a decline in secondary metabolite production when growth rate was increased (Rossa *et al.*, 2002).

Within the PEP-PYR-OAA node of *Streptomyces* only one gene has been studied in detail, the major anaplerotic enzyme, phosphoenolpyruvate carboxylase (PEPCx) from *S. coelicolor* (Bramwell *et al.*, 1993). Activity of this enzyme was shown to increase during actinorhodin biosynthesis and overexpression of PEPCx in *S. lividans* reduced the growth rate of the strain, delaying actinorhodin biosynthesis. These data suggest the importance of this node to specialised metabolite production, especially for polyketides and highlights the potential for further investigation.

Recent global metabolomics studies in *Streptomyces* have also revealed key findings about the metabolic flexibility at the PEP-PYR-OAA node at onset of specialised metabolite production. Studies conducted during phosphate and L-glutamate depletion indicate that most changes in the global metabolite pool (metabolome) were in amino acid and organic acid levels (Wentzel, Sletta, *et al.*, 2012). During phosphate depletion the amino acid pools of histidine, phenylalanine, tyrosine, alanine, valine, leucine, glycine, proline, isoleucine and lysine were increased in addition to the intracellular succinate and ornithine pools. Glutamate and aspartate pools were both down as was pyruvate, citrate, α- ketoglutarate, fumarate and malate. Under glutamate depletion the pools of histidine, phenylalanine,

tyrosine, alanine, valine, leucine, glycine, proline and lysine decreased initially before recovering at around 40hrs, this was also observed for citrate and succinate. Pyruvate, α-ketoglutarate, fumarate, malate and ornithine were all reduced in addition to the amino acids glutamate, glutamine and aspartate (Wentzel, Sletta, *et al.*, 2012). These data confirm the central role for pyruvate in balancing central carbon metabolism during growth and reinforces our understanding of key branch points in specialised metabolite production. An additional study by this group to develop cultivation media for studying metabolic shifts, tested a wide range of carbon sources (arabinose, alanine, aspartate, glucose/glutamate, glucose, glutamate, proline, Tween 20, Tween 40, Tween 60, Tween 80 and xylose) and examined expression of the pyruvate dehydrogenase complex genes - only one of the genes showed a decreased expression (SCO2183). On Tween and alanine, SCO2183, SCO2181, SCO2180 and SCO4919 showed increased expression, yet under all other conditions no expression differences were observed, which may suggest that the pyruvate dehydrogenase complex genes may act as a metabolic bottleneck during growth on glucose (Alam *et al.*, 2010; Wentzel, Bruheim, *et al.*, 2012).

Furthermore studies of carbon preferences for glucose and glutamate during culture on both substrates as a mixed carbon source, using $^{13}$C-glucose to label metabolites, indicate that during rapid growth glycolysis and PPP are enriched for $^{13}$C compounds, yet up on growth cessation $^{13}$C-labelled PPP intermediates decreased. The TCA cycle is generally low in $^{13}$C-labelled intermediates indicating glutamate is a preferred carbon source, being catabolised via α-ketoglutarate following deamination and release of ammonium ions. Interestingly α-ketoglutarate can be decarboxylated in the TCA cycle to form malate and further decarboxylated to pyruvate where PDHC can convert to acetyl-coA. Acetyl-CoA is an important precursor for fatty acids and polyketides. Clearly glutamate is the preferred carbon source providing the main cellular carbon and glucose playing an ancillary role (Wentzel, Bruheim, *et al.*, 2012). Interestingly, it is known that *Streptomyces* secrete pyruvate and α-ketoglutarate during growth under certain conditions, prior to specialised metabolite production (Dekleva and Strohl, 1987; Madden *et al.*, 1996; Hodgson, 2000), which may reflect the inefficiency of acetyl-CoA formation from pyruvate under some physiological conditions. PDHC expansion in streptomycetes may be an evolutionary solution to this phenomenon. These natural examples may provide a framework for metabolic engineering strategies.

## 1.5 Scope of the PhD Project

Detailed insight into the evolution, regulation and biochemistry of primary metabolism and how this feeds into specialised metabolism will help to better understand complex biological systems and will allow targeting of key points in metabolism for metabolic engineering. Global 'omics' studies and modelling can help with general phenomena, however a full understanding of gene function in a classical reductionist manner is the only way to gain true biological insight. These approaches can help synthetic biology approaches to strain and pathway construction and increased production of medically useful metabolites. Key to this is to understand the interaction of the different metabolic pathways as well as understanding how the metabolic flexibility has arisen.

An excellent metabolic engineering target such as the PEP-PYR-OAA node is particularly well suited for engineering metabolite production and will make it more efficient and easier to establish industrial production processes. The model organism *S. coelicolor A3*(2) producing the polyketide actinorhodin has been used as a proof of concept for the lab work.

## 1.6 Specific Aims

The aims of this thesis were the following:

- Identification of most promising targets for metabolic engineering within the chosen metabolic node for increasing precursor supply into polyketide synthesis
- Characterisation of promising targets in terms of their evolution, role in growth and production of polyketides
- Understanding the metabolic expansion in central carbon metabolism and in particular the PEP-PYR-OAA node
- Understanding primary metabolism influence specialised metabolism and how it can be modified to increase production titres

# 2 MATERIAL AND METHODS

All chemicals and reagents were ordered from the following suppliers: Sigma Aldrich, Fisher, VWR, Promega, New England Biolabs, BioLine, Life Technologies, Bio Rad or GE Healthcare. The supplier will be specified in parenthesis.

## 2.1 CREATION OF AN ACTINOBACTERIAL GENOME DATABASE

This part of the thesis was carried out during a visit to the laboratory of Dr Francisco Barona Gomez at the LangeBio in Irapuato, Mexico. The work was conducted in collaboration with Nelly Selem, Christian Martinez and Dr Pablo Cruz Morales.

The actinobacterial genomes in the database were selected from NCBI (http://www.ncbi.nlm.nih.gov/genbank/wgs) with a minimum coverage of 25x and less than 30 contigs per MB. To ensure a wide range of phylogeny in the end a selection of 614 species from 83 genera were included by using the taxonomy browser of NCBI (http://www.ncbi.nlm.nih.gov/taxonomy) searching within Actinobacteria and displaying the location of a genome sequence with sufficient quality.

Each genome sequence was downloaded and re-annotated on Rapid Annotation using Subsystem Technology (RAST, Aziz *et al.*, 2008; Overbeek *et al.*, 2014; Brettin *et al.*, 2015). A subsystem is defined "set of functional roles that together implement a specific biological process or structural complex" (Overbeek *et al.*, 2005). It can be understood as a generalisation of pathways that goes beyond metabolism, so a subsystem could be glycolysis but also a transport system or a cellular process. The subsystems are curated manually and from this technology, which has been developed in order to obtain consistent and reproducible annotation among different genomes, FIGfams are automatically derived. FIGfams are sets of proteins that are isofunctional homologs, meaning that they correspond to each other by having the same function and were derived from a common ancestor (Meyer *et al.*, 2009).

The annotation files of all the genomes were downloaded and the number of occurrence of each functional role (several of which will form a subsystem) in each genome was counted and used to generate a table with all functional roles and all genomes. In order to obtain a manageable amount of data to be analysed and to prevent biases by overrepresented genera, the mean number of occurrences of each functional role was calculated per genus

and those were highlighted that had an equal or higher value than the mean plus its standard deviation. The results were stored in a separate table organised by pathways of interest, which were glycolysis, gluconeogenesis, pentose phosphate pathway, TCA cycle and amino acid metabolism as those are precursor routes (Chapter 3.1).

The highlighted functional roles in this table were treated as gene expansion events in a given genus and were analysed in more detail for *Streptomyces* (Chapter 3.1). In addition to the expansions detected in the pathways mentioned above, all members of the PEP-PYR-OAA node were also analysed in detail to examine their phylogenetic history.

This more detailed analysis was carried out as follows, all protein sequences were extracted from the actinobacterial genome database to form a database that can be used to run a blastp analysis (Altschul *et al.*, 1990). This protein database was then queried with protein sequences of interest from the node and the expansion hits. The amino acid sequences for the query sequences were obtained in fasta format from StrepDB (http://strepdb.streptomyces.org.uk). The obtained blastp sequences were then aligned with MUSCLE (Edgar, 2004a; Edgar, 2004b). The sequence alignment was then analysed using Jalview (Waterhouse *et al.*, 2009) to ensure that at least 25 % coverage with the query were achieved. If necessary too short or badly aligned sequences (less than 25% coverage with query) were deleted and the alignment was re-run. If a badly aligned sequence was deleted, also all other hits from this genome were excluded in order to prevent obtaining false information.

The final alignment was then utilised in Quicktree (version 1.1, Neighbor Joining Tree,Howe *et al.*, 2002) or MrBayes (V3.2,Ronquist *et al.*, 2012) for the construction of a phylogenetic tree. The trees were visualised in FigTree (V1.4.2, http://tree.bio.ed.ac.uk/software/figtree/) and colour coded according to their taxonomic family. A schematic overview of the workflow can be found in Figure 2.1.

The respective scripts used for each step of the analyses and the settings for the different softwares can be found in the appendix (10.1).

## 2.2 Synonymous and Non-synonymous Changes Analysis in pyruvate kinases

Pyruvate kinase gene sequences were obtained from NCBI gene for ten different *Streptomyces* species and an alignment of the twenty gene sequences was carried out using ClustalW algorithm in MEGA V6.06 (Tamura *et al.*, 2013). The synonymous (dS) and nonsynonymous (dN) changes were determined using MEGA V6.06 Software Suite (Tamura *et al.*, 2013) using the Distance Model Function with Nei Gojobori with Jukes Cantor algorithm with a 1000 times bootstrap (Nei and Gojobori, 1986). Additionally a phylogenetic tree of the twenty pyruvate kinases was constructed using Maximum Likelihood algorithm using MEGA V6.06 (Tamura *et al.*, 2013).

## 2.3 Creation of a 3D Structural Model for Pyk1 and Pyk2 (Biasini *et al.*, 2014)

Both Pyk1 and Pyk2 were modelled using the webtool SwissModel from ExPASy at the Biozentrum in Basel, Switzerland (http://swissmodel.expasy.org/). The crystal structure of *E. coli* PykF was utilised (PDB Entry: 4YNG) as template for modelling. Default settings were used. Briefly, the submitted amino acid sequence was blasted against a template library (SMTL version 2015-07-30, PDB release 2015-07-24) available on the server and the *E. coli* PykF was the best fitting match. The target and the template were then structurally aligned using Promod-II (Guex and Peitsch, 1997). Conserved coordinates between the target and the template were copied from the template to the model, followed by remodelling of insertions and deletions using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field.

The model built is only a hypothesis and has not been verified and needs to be analysed with great care. The obtained model was then analysed using PyMol (Schroedinger LLC, version 1.3), which was also used to create images of the relevant sites.

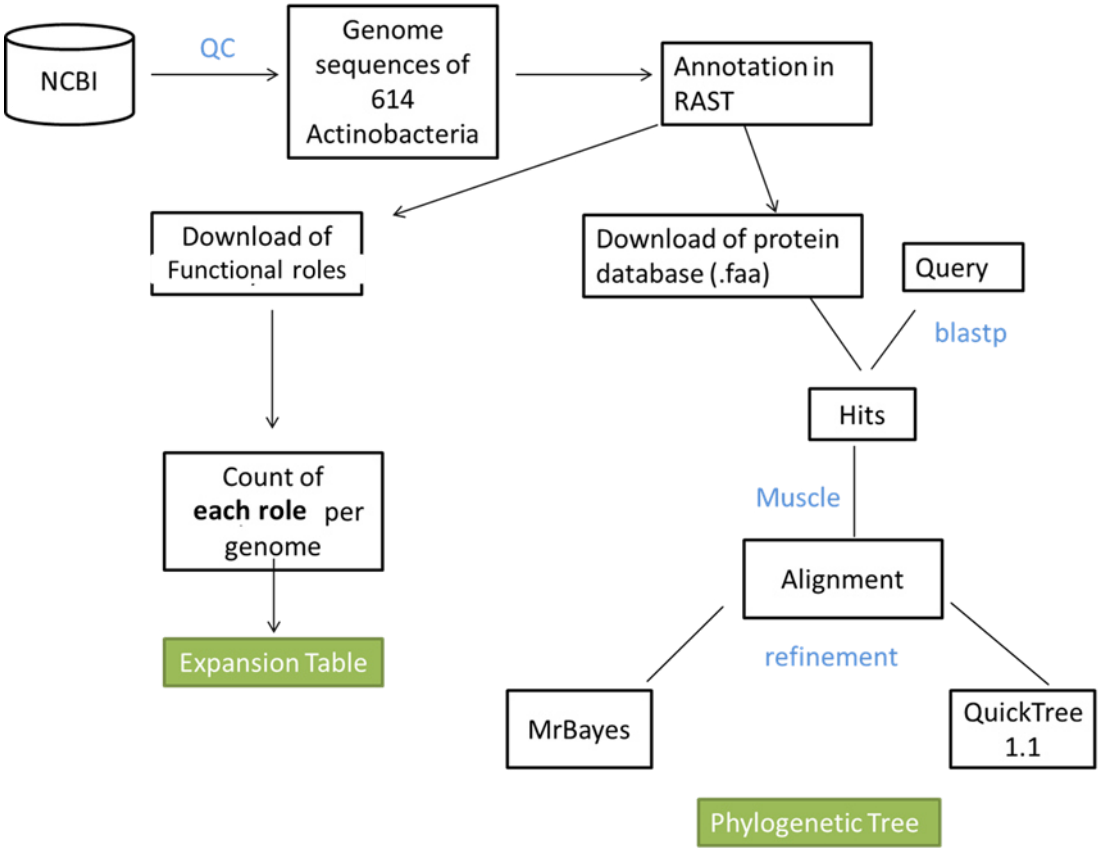



Figure 2.1 Schematic workflow for the creation of an actinobacterial genome database and subsequent mining of data for expansion and phylogenetic analysis. Genomes were selected from NCBI Taxonomy browser and uploaded for annotation to RAST (Aziz *et al.*, 2008). The annotated genomes were used for two different analyses. Firstly, the functional roles were downloaded and for each functional role the number of occurences per genomes were counted in order to obtain an expansion table by comparing the mean of each genus to the overall mean of all genera. Secondly all the genomes were used to extract all protein sequences in fasta format which could then be questioned with a query sequence using blastp (Altschul *et al.*, 1990). The hits were aligned in MUSCLE (Edgar, 2004a; Edgar, 2004b) and after refinement the alignment was used to construct phylogenetic trees in MrBayes (Ronquist *et al.*, 2012) and Quicktree (Howe *et al.*, 2002).

## 2.4 MEDIA

All media were autoclaved for sterilisation (121˚C for 15 min) unless a substance was heat sensitive such as the trace element solutions or antibiotics, which were filter sterilised (0.22 µm). The recipes for each medium can be found in Table 2-1. Depending on the bacterial strain an antibiotic was added to the medium (Table 2-2). Supplemented minimal medium (SMM) and minimal agar medium have been supplemented with different carbon sources, the stock solutions for this can be found in Table 2-3.

## 2.5 MICROBIAL STRAINS, PLASMIDS, COSMIDS AND OLIGOS

In the following the microbial strains with their respective genotype are shown (Table 2-4). *E. coli* glycerol stocks were made from mixing 500 µl of the respective strain from an overnight culture with an equal volume of sterile 50 % glycerol (v/v) and were stored at -80˚C. A *S. coelicolor* strain grown on MS medium for seven days until confluent was used to prepare glycerol spore stocks. Aliquots (7 ml) of 20% glycerol were poured onto the plate and a sterile cotton bud was used to retrieve the spores from the surface of the plate. The spore suspension was then aliquoted into cryovials and vials were stored at −20˚C.

All plasmids constructed and utilised in this thesis are listed in Table 2-5. The cosmids (Table 2-6) were ordered from the transposon insertion library from the University of Swansea (Redenbach *et al.*, 1996; Bishop *et al.*, 2004). Successful incorporation can be selected on apramycin resistance and screened for loss of kanamycin resistance of the strains. A schematic of the transposons utilised is shown in Figure 2.2. All oligos designed and utilised in this thesis can be found in Table 2-7.

Table 2-1 Culture media utilised to grow *Streptomyces* and *E.coli* strains

| **Medium (Reference)** | **Recipe** |
|---|---|
| Minimal Medium (Hopwood, 1967) | 0.5 g or 1 g L-asparagine or $(NH_4)_2SO_4$<br>0.5 g $K_2HPO_4$<br>0.2 g $MgSO_4*7H_2O$<br>0.01 g $FeSO_4$ *7 $H_2O$<br>10 g Agar<br>Add 1000 ml $dH_2O$, pH to 7.0 with 5M NaOH, refer to Table 2-3 for 1% carbon source added after autoclaving |
| Minimal Medium K97 (Karandikar *et al.*, 1997) | 1.7 g $NaNO_3$<br>1 g $K_2HPO_4$<br>0.5 g KCl<br>0.5 g $MgSO_4*7$ $H_2O$<br>0.01 g $FeSO_4*7$ $H_2O$<br>2.5 ml/l Trace element solution *<br>15 g Agar<br>Add 1000 ml $dH_2O$, pH to 7.0 with 5M NaOH, 50 g/l glucose after autoclaving |
| Supplemented Minimal Medium (Strauch *et al.*, 1991) | 5 % PEG 6000 (6.1% stock solution)<br>2.5 mM $MgSO_4*7H_2O$ (24 g/l)<br>0.25 mM TES buffer (pH 7.2),<br>1 mM $NaH_2PO_4$ (50 mM)<br>1 mM $K_2HPO_4$ (50 mM)<br>1% carbon source (Table 2-3)<br>1 ml/l trace element solution ** (Filter sterilised)<br>0.2 % Casein (20%)<br>0.5% (20%) Glycine<br>Add 1000 ml $dH_2O$ |
| Hobbs Minimal Medium (Hobbs *et al.*, 1989) | 5 g NaCl<br>5 g $Na_2SO_4$<br>4.5 g $NaNO_3$<br>1.2 g Tris(hydroxymethyl)aminomethane (TRIS)<br>2 g $MgSO_4*$ 7 $H_2O$<br>0.01 g $ZnSO_4*$ 7 $H_2O$<br>Add 1000 ml $dH_2O$, pH was adjusted to 7.2.<br>Added after autoclaving:<br>2g $K_2HPO4$ (20 g/l stock)<br>1 ml/l trace element solution *<br>either glucose or Tween40 were used as carbon source (0.33 mol carbon)<br>20 µl of Sigma Antifoam A |
| Mannitol Soya Flour Medium (Hobbs *et al.*, 1989) | 16 g Mannitol<br>16 g Soya Flour (from Holland and Barreth Health Shop, UK)<br>16 g Agar<br>Add 1000 ml tap water |
| R2 (Okanishi *et al.*, 1974) | 103 g Sucrose<br>0.25 g $K_2SO_4$<br>5 g $MgCl_2$ *6 $H_2O$<br>10 g glucose |

| | |
|---|---|
| | 0.1 g casein hydrolysate<br>2 ml/l trace element solution *<br>5.73 g triethylsilane (TES)<br>22 g agar<br>Add 1000 ml $dH_2O$. Added after autoclaving (per 100 ml)<br>1 ml 0.5% $KH_2PO_4$<br>0.4 ml 3.68% $CaCl_2$*2 $H_2O$<br>1.5 ml 20 % L-Proline<br>0.7 ml 1 M NaOH |
| R5 (Thompson *et al.*, 1980) | 103 g sucrose<br>0.25 g $K_2SO_4$<br>10.12 g $MgCl_2$*6 $H_2O$<br>10 g glucose<br>0.1 g casein hydrolysate<br>2 ml/l trace element solution<br>5 g yeast extract<br>5.73 g triethylsilane (TES)<br>27.5 g agar<br>Add 1000 ml $dH_2O$, additional components added after autoclaving (per 80 ml)<br>1 ml 0.5% $KH_2PO_4$<br>0.4 ml 5 M $CaCl_2$*2 $H_2O$<br>1.5 ml 20% L-Proline<br>0.7 ml of 1 M NaOH |
| 2xYT (Kieser *et al.*, 2000) | 16 g tryptone<br>10 g yeast extract<br>5 g NaCl<br>Add 1000 ml $dH_2O$ |
| YEME (Kieser *et al.*, 2000) | 5 g peptone<br>3 g yeast extract<br>3 g malt extract<br>10 g glucose<br>340 g sucrose or 10 g agar<br>Add 1000 ml $dH_2O$<br>pH was adjusted to 7.2 with 5 M NaOH<br>after autoclaving 2 µl/l of a 2.5 M $MgCl_2$*6 $H_2O$ were added |
| Nutrient Broth (Oxoid) | 1g Lab-Lemco<br>2g Yeast extract<br>5g Peptone<br>5g NaCl<br>10g Agar<br>Add 1000 ml $dH_2O$ |
| M9 Medium (Sambrook and Russell, 2000) | 5 x salt solution (pH of 7.0):<br>33.9 g $Na_2HPO_4$<br>15 g $KH_2PO_4$,<br>5 g $NH_4Cl$<br>2.5 g NaCl<br>200 ml 5 x salt solution (as specified above)<br>2 ml 1M $MgSO_4$<br>100 µl of 1 M $CaCl_2$<br>1 % glucose or 0.4 % acetate<br>Add 1000 ml $dH_2O$ |

| | |
|---|---|
| Lysogeny Broth (Bertani, 1951) | 10 g tryptone<br>5 g yeast extract<br>5 g NaCl<br>10 g Agar for solid medium<br>Add 1000 ml $dH_2O$ |
| Autoinduction Medium AIM (Studier, 2005) | 10 g tryptone<br>5 g yeast extract<br>3.3 g $(NH_4)_2SO_4$<br>6.8 g $KH_2PO_4$<br>7.1 g $Na_2HPO_4$<br>0.5 g Glucose<br>2.0 g α-Lactose<br>0.15 g/l $MgSO_4$<br>Add 1000 ml $dH_2O$ |

*trace element solution was composed of: 8.78 g/l $FeCl_3$, 2.04 g/l $ZnSO_4$, 1.02 g/l $MnCl_2$*4 $H_2O$, 0.43 g/l $CuSO_4$*2 $H_2O$, 0.42 g/l NaI, 0.31 g/l $H_3BO_3$, 0.24 g/l $CaCl_2$*6 $H_2O$ and 0.24 g/l $Na_2MoO_4$*2 $H_2O$ ** trace element solution of (Hobbs *et al.*, 1989) consists of (g/100 ml; Filter sterilised): 0.1 $ZnSO_4$*7 $H_2O$, 0.1 $FeSo_4$ *7 $H_2O$, 0.1 $MnCl_2$ * 4 $H_2O$, 0.1 $CaCl_2$ * 6 $H_2O$ and 0.1 NaCl, can be stored at 4◦ C for 2–4 weeks. It is diluted 1:10 before added to the medium

Table 2-2 Antibiotics used for selection pressure in culture media

| **Antibiotic** | **Stock [mg/ml]** | **Final concentration [µg /ml]** |
|---|---|---|
| Apramycin | 100 in $dH_2O$ | 50 |
| Hygromycin | 50 in $dH_2O$ | 100 |
| Kanamycin | 50 in $dH_2O$ | 50 |
| Chloramphenicol | 25 in 100% EtOH | 25 |
| Carbenicillin | 100 in $dH_2O$ | 100 |
| Thiostrepton | 50 in 100% DMSO | 50 |
| Nalidixic acid | 25 in $dH_2O$ | 25 |

Table 2-3 Carbon sources added to minimal medium for screening

| **Carbon source** | **Stock [%]/pH** | **ml to 1l medium (1 %)** |
|---|---|---|
| Glucose | 50 | 20 |
| Mannitol | 25 | 40 |
| Tween80 | 20 | 50 |
| Pyruvate | 20/ pH 7.2 | 50 |
| Na-Acetate | 20/ pH 7.2 | 50 |
| Na-Citrate | 20/ pH 7.2 | 50 |
| Malic acid | 20/ pH 7.2 | 50 |
| PEP | 20 | 50 |
| α-KG | 20 | 50 |
| GlcNac | 20 | 50 |

Table 2-4. Microbial strains used in this thesis with their genotype

| Strain | Genotype | Reference |
|---|---|---|
| *E. coli* BL21 | *dcm, ompT, hsdS (rB, mB), gal* | Studier and Moffatt, 1986 |
| *E. coli* Rosetta | *ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (CamR)* | Novagen® |
| *E. coli* C41 | *F⁻ ompT gal dcm hsdS$_B$($r_B^-$ $m_B^-$)(DE3)* | Novagen® |
| *E. coli* C43 | *F⁻ ompT gal dcm hsdS$_B$($r_B^-$ $m_B^-$)(DE3)* | Novagen® |
| *E. coli* Origami | *ΔtrxB/gor Leu⁻* | Novagen® |
| *E. coli* OrigamiB | *ΔtrxB/gor ΔlacZY* | Novagen® |
| *E. coli* Tuner | *ΔlacZY ΔIon ΔompT* | Novagen® |
| *E. coli* Tuner p*LacI* | *ΔlacZY ΔIon ΔompT pLacI* | Novagen® |
| *E. coli* BW25113 pIJ790 | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | Datsenko and Wanner, 2000 |
| *E. coli* BW25113 | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* BW25113 Δ*pykA* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | gift from Ian Henderson |
| *E. coli* BW25113 Δ*pykF* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | gift from Ian Henderson |
| *E. coli* BW25113 Δ*pykA*Δ*pykF* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* BW25113 Δ*pykA* + *pyk1* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* BW25113 Δ*pykA* + *pyk2* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* BW25113 Δ*pykF* + *pyk1* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* BW25113 Δ*pykF* + *pyk2* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* BW25113 Δ*pykA*Δ*pykF* + *pyk1* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* BW25113 Δ*pykA*Δ*pykF* + *pyk2* | *lacI⁺ rrnB T14 ΔlacZ WJI6 hsdR514 ΔaraBAD AH33 ΔrhaBAD LD78* | this study |
| *E. coli* DH5α | *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17* | Grant *et al.*, 1990 |
| *E. coli* ET12567/pUZ8002 | *dam-13::Tn9 dcm-6 hsdM hsdR recF143, zij201::Tn10, galK2, galT22, ara14, lacYI, xylS, leuB6, thi-1, tonA31, rpsL136, hisG4, tsx78, mtli, glnV44, F-* | MacNeil *et al.*, 1992 |
| *S. coelicolor* M145 A3(2) | SCP1- SCP2- | Kieser *et al.*, 2000 |
| *S. coelicolor* sJH_01 | *pyc*::Tn5062 (SCO0546) | this study |
| *S. coelicolor* sJH_02 | *pyk1*::Tn5062 (SCO2014) | Dr L. Fernández-Martínez |
| *S. coelicolor* sJH_03 | *meNAD*::Tn5062 (SCO2951) | This study |
| *S. coelicolor* sJH_04 | *ppc*::Tn5062 (SCO3127) | L. Fernández-Martínez |
| *S. coelicolor* sJH_05 | *ppck*::Tn5062 (SCO4979) | L. Fernández-Martínez |
| *S. coelicolor* sJH_06 | *pyk2*::Tn5062 (SCO5423) | this study |

| | | |
|---|---|---|
| *S. coelicolor* sJH_07 | *poxB1*::Tn5062 (SCO6155) | this study |
| *S. coelicolor* sJH_08 | *aceE3*::Tn5062 (SCO7124) | L. Fernández-Martínez |
| *S. coelicolor* sJH_09 | *poxB*::Tn5062 (SCO7412) | L. Fernández-Martínez |
| *S. coelicolor* Δ*pyk1* | *pyk1* replaced with Apr (pIJ773) | P. Cruz Morales |
| *S. coelicolor* sJH10 | *Δpyk1 +* pIJ6902 (empty vector control) | this study |
| *S. coelicolor* sJH11 | *Δpyk1 + pyk1* (complemented pIJ6902_*pyk1)* | this study |
| *S. coelicolor* sJH12 | *Δpyk1 + pyk2* (complemented pIJ6902_*pyk2)* | this study |
| *S. coelicolor* Δ*pyk2* | *pyk2* replaced with Apr (pIJ773) | P. Cruz Morales |
| *S. coelicolor* sJH13 | *Δpyk2 +* pIJ6902 (empty vector control) | this study |
| *S. coelicolor* sJH14 | *Δpyk2 + pyk1* (complemented pIJ6902_*pyk1)* | this study |
| *S. coelicolor* sJH15 | *Δpyk2 + pyk2* (complemented pIJ6902_*pyk2)* | this study |
| *S. coelicolor* sJH16 | *pyk1*::Tn5062 + pIJ6902 *(*empty vector control) | this study |
| *S. coelicolor* sJH17 | *pyk1*::Tn5062 *+ pyk1* (complemented pIJ6902_*pyk1)* | this study |
| *S. coelicolor* sJH18 | *pyk1*::Tn5062 *+ pyk2* (complemented pIJ6902_*pyk2)* | this study |
| *S. coelicolor* sJH19 | *pyk2*::Tn5062 + pIJ6902 (empty vector control) | this study |
| *S. coelicolor* sJH20 | *pyk2*::Tn5062 *+ pyk1* (complemented pIJ6902_*pyk1)* | this study |
| *S. coelicolor* sJH21 | *pyk2*::Tn5065 *+ pyk2* (complemented pIJ6902_*pyk2*) | this study |
| *S. coelicolor* sJH22 | *aceE3*::Tn5062 + pIJ6902 *(*empty vector control) | this study |
| *S. coelicolor* sJH23 | *aceE3*::Tn5062 *+ aceE3* (complemented pIJ6902_*aceE3*) | this study |
| *S. coelicolor* sJH24 | *Δpyk1 pyk2*::Tn5066 | this study |
| *S. coelicolor* sJH25 | *pyk2*::Tn5066 | this study |

Table 2-5 Plasmids that were used in the work for this thesis

| **Plasmids** | **Antibiotic marker** | **Reference** |
|---|---|---|
| pET100_TOPO | Carb | Invitrogen |
| pET15b | Carb | Novagen |
| pGEM Teasy | Carb | Promega |
| pIJ6902 | Apr | Huang *et al.*, 2005 |
| pET28a | Kan | Novagen |
| pEX-K4-SCO2014s | Kan | synthetic gene (MWG Eurofins) |
| pJH01 (TOPO pET100_SCO2014) | Carb | this study |
| pJH02 (TOPO pET100_SCO5423) | Carb | this study |
| pJH03 (pGEM_SCO2014) | Carb | this study |
| pJH04 (pGEM_SCO5423) | Carb | this study |
| pJH05 (pGEM_SCO7124) | Carb | this study |
| pJH06 (pIJ6902_SCO2014) | Apr, Thio | this study |
| pJH07 (pIJ6902_SCO7124) | Apr, Thio | this study |
| pJH08 (pIJ6902_SCO5423) | Apr, Thio | this study |
| pJH09 (pGEM_SCO2014cterm) | Carb | this study |
| pJH10 (pET28a_SCO2014cterm) | Kan | this study |
| pJH11 (pET15b_SCO2014s) | Carb | this study |

Table 2-6 List of cosmids used for the strain construction in *S. coelicolor* with the respective gene of interest

| Gene of interest | Cosmid |
|---|---|
| SCO0546 | SCF11.26c |
| SCO2014 | 1D05 |
| SCO2951 | SCE59.10c |
| SCO3127 | SCE66.06c |
| SCO4979 | 2SCK36.02 |
| SCO5423 | SC8F4.27c |
| SCO6155 | SC1A9.19 |
| SCO7124 | SC4B10.25c |
| SCO7412 | SC6D11.08 |

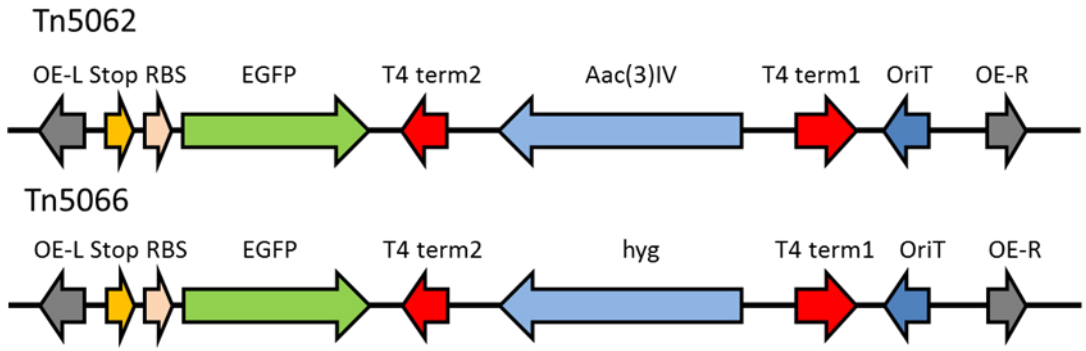



Figure 2.2 Transposon insertions Tn5062 (Bishop *et al.*, 2004) and Tn5066 (Fernández-Martínez *et al.*, 2011). Transposon Tn5062 (3442 bp): OE-L/R=inverted repeats, Stop=translational stop codon, RBS= streptomyces consensus ribosomal binding site, EGFP= GFP coding sequence, T4 term1/2= terminators, Aac(3)IV= apramycin resistance, OriT= RK2 origin of transfer.
Transposon Tn5066 (4173 bp, Accession number: HQ213785, Fernández-Martínez *et al.*, 2011): OE-L/R=inverted repeats, Stop=translational stop codon, RBS= streptomyces consensus ribosomal binding site, EGFP= GFP coding sequence, T4 term1/2= terminators, hyg= hygromycin resistance, OriT= RK2 origin of transfer

Table 2-7 Overview of all primer utilised in this thesis, specifying for which gene, direction, sequence,melting temperature, amplicon size and use in which experiment

| Gene | Direc -tion | Sequence | Tm (°C) | Amplicon (bp) | Use |
|---|---|---|---|---|---|
| Tn5062 | Rv | GAACTTCAGGGTCAGCTTGC | 53.8 | variable | Strain verification |
| | Fw | CCTTCCAGACGAACGAAGAG | 53.8 | | |
| SCO2014 (*pyk1*) | Fw | GGTGACCCGTACACACACAG | 59.9 | 1703 | Insert for complemen tation |
| | Rv | CATCTCCGTAGCCCGTAAAG | 59.7 | | |
| SCO5423 (*pyk2*) | Fw | GGGCTGAACGCCGTAC | 55.6 | 1921 | |
| | Rv | CATGTACACCGAGAGTATTTCG | 57.7 | | |
| SCO7124 (*aceE3*) | Fw | CATCACACCTTCACCCAGTG | 60 | 3092 | |
| | Rv | ATCTCGGTGTCCACCTTGTC | 60 | | |
| SCO2014 (*pyk1*) | Fw | CACCATGCGCCGAGCAAAGATC | 69.7 | 1437 | Insert for overexpress ion |
| | Rv | TCACTTGGGAATGTCGTCCACCCC | 70.7 | | |
| SCO5423 (*pyk2*) | Fw | CACCATGCGCCGTTCGAAAATC | 69.2 | 1431 | |
| | Rv | TCAGCCGCGCCGTGTCTC | 68 | | |
| SCO2014 C-term (*pyk1*) | Fw | CCATGGGTATGCGCCGTGCA | 57.9 | 1447 | |
| | Rv | GC GGC CGC CTT GGG AAT GTC | 60 | | |
| SCO2014 (*pyk1*) | Fw | CGACTCGTACGACCAGATCA | 60 | 489 | RT-PCR |
| | Rv | AGGTCGTCCTCGTCCTTCTT | 60 | | |
| SCO5423 (*pyk2*) | Fw | CGGTGACGAGTTCACCATC | 60.1 | 438 | |
| | Rv | ATCACGACGTCCTCCATGTT | 60.4 | | |
| SCO5820 (*hrdB*) | Fw | GAGGCGACCGAGGAGCCGAA | 65.5 | 324 | |
| | Rv | GCGGAGGTTGGCCTCCAGCA | 65.5 | | |
| SCO2014 (*pyk1*) | Fw | CGACTCGTACGACCAGATCA | 59.8 | 103 | qPCR |
| | Rv | GTGGTAGCGCTCTTCGTGTT | 60.4 | | |
| SCO5423 (*pyk2*) | Fw | ACCAGGTGCTCATCAACGAC | 61.1 | 172 | |
| | Rv | AGGTCCTCGACGTCCTTCTC | 60.7 | | |
| SCO5820 (*hrdB*) | Fw | TGCTCTTCCTGGACCTCATC | 60.3 | 100 | |
| | Rv | GTGGCGTACGTGGAGAACTT | 60.1 | | |
| SCO2494 (*ppdK2*) | Fw | AAGATCCCCAGGTAGCGAAC | 59.2 | 177 | |
| | Rv | TGTCGAGGTAGGTCTTGCAG | 59.1 | | |
| SCO0208 (*ppdK1*) | Fw | AGTTCTTCTCCGCCTACCTC | 58.5 | 184 | |
| | Rv | TGGAAGAAGTGCACCGAGTC | 59.9 | | |
| *pykF*_mut apr | Up | TCTCCCATCCTTCTCAACTTAAAGACTAAGACTGTCATG ATTCCGGGGATCCGTCGACC | 87.7 | | *E. coli* mutant construction and verification |
| | Dn | GCGCTTCGATATACAAATTAATTCACAAAAGCAATATTA TGTAGGCTGGAGCTGCTTC | 81.7 | | |
| Kan_Int | Fw | CCTGCAAAGTAAACTGGATG | 49.7 | | |
| | Rv | CATGCTCTTCGTGCAGATCA | 51.8 | | |
| *pykA* | Fw | TTGAAGCGGGTCAAAGAAGC | 51.8 | | |
| | Rv | CGAAATCCAGCCGTATAAGC | 51.8 | | |
| *pykF* | Fw | TTTCAGCGTATAATGCGCGC | 51.8 | | |
| | Rv | AAAGAAGCATCGAACGCTGG | 51.8 | | |
| *pykF*_flank | Fw | GGCACCACCACTTTCGTAAT | 51.8 | | |
| | Rv | ATCAGGGCGCTTCGATATAC | 51.8 | | |
| Apr_Int | Rv | TCGGTCAGCTTCTCAACCTT | 51.8 | | |
| | Fw | ACCAACTTGCCATCCTGAAG | 51.8 | | |

## 2.6 GROWTH AND PHYSIOLOGY

### 2.6.1 GROWTH CURVES OF *STREPTOMYCES*

Spores were germinated in 20 – 50 ml of 2x YT medium in a 250 ml Erlenmeyer flask containing a metal spring by inoculation with 500 µl of a dense spore suspension of the respective strain. For small scale cultivation, the germination was carried out in 20 ml Universals with 10 ml 2x YT. Cultures were incubated for up to 8h at 30°C and 250 rpm until emerging germ tubes were visible by microscopy. The culture was centrifuged at 1,000 x g for 5 min and washed twice with 0.25 M TES buffer (57.3 g/l, pH 7.2). The supernatant was decanted and 5 ml of fresh respective medium was used for resuspension by vortexing for 45 s and to disperse any possible aggregates.

The actual growth were performed at 200-400 ml scale in 2 l flasks containing a metal spring at 30°C and 250 rpm. Inoculation was either directly from spore stocks or from germinated spores. Samples were taken at adequate time intervals ranging from 5 -15 ml sample volume depending on the amount of biomass present. The first sample was taken after 17-24 h followed by a sample each 2-3 h in order to capture the exponential phase for up to 72 h.

Growth curves were also performed at small scale for *Streptomyces* in 96 well plates using a Microplate Reader (Bio-Tek Multi-Detection Microplate Reader Synergy HT) as shaking incubator measuring the $OD_{450}$ every 15 min over 48h. The working volume was of 200 µl and 1.5 x $10^5$ cells were used to inoculate each well from a germinated culture grown in 2x YT.

### 2.6.2 SMALL SCALE SCREENING FOR CHANGES IN ACTINORHODIN PRODUCTION ON LIQUID MEDIUM

Initial screening of transposon mutant strains of the metabolic node were carried out in 24 well plates. Each well of the plate was filled with three 2 mm diameter glass beads to ensure sufficient oxygen supply and to prevent large aggregate formation. The working volume was 1.5 ml of medium per well. The wells on the outer site of the plate were not inoculated with cells to prevent evaporation. Inoculation was performed directly from spore stock, washing the cells previously with medium to remove any glycerol. The inoculation volume per well was of 15 µl (1 %). Plates were incubated at 30°C and 170 rpm for 6 days.

### 2.6.3 Sample analysis- Cell Dry Weight (CDW) Determination

Samples taken from the growth curves were analysed for biomass content using cell dry weight. The sample was vacuum filtered on pre-weighed glass fibre paper (1 µm retention, 5.5 cm, Fisher) and washed three times with 10 ml of $dH_2O$. The filter with the biomass was then dried at 900 W for 10 min in a microwave oven (Prestige SE28S). The weight of the filter paper prior to adding the sample was subtracted after drying the biomass on the filter, to obtain the biomass. The value was divided by the sample volume to obtain mg/ml.

### 2.6.4 Sample analysis- Actinorhodin Quantification (Kang *et al.*, 1998)

In addition to the determination of biomass, samples taken from growth curves were analysed for the concentration of actinorhodin. Equal volumes of culture supernatant and 1M NaOH were mixed by vortexing. The samples were then centrifuged at 1000 g for 5 min and the absorbance at 633 nm was measured. The molar extinction coefficient for actinorhodin is 15,135 $m^{-1}$ $cm^{-1}$ and by using the Lambert-Beer-Law the concentration was calculated (Hobbs *et al.*, 1990; Kang *et al.*, 1998). The specific production yield was calculated by normalising it to µM per mg of CDW.

### 2.6.5 Sample analysis- Quantification yCPK (Gottelt *et al.*, 2010)

Supernatants of *Streptomyces* cultures were measured at 460 nm (Gottelt *et al.*, 2010) to determine production of yellow pigment (yCPK). The yield was calculated as $AU_{460}$ per mg of CDW.

### 2.6.6 Interspecies Complementation of pyruvate kinase mutant

*E. coli* pyruvate kinase mutants were derived from *E. coli* BW25113, both single mutants and the double mutant were complemented individually with the *Streptomyces* pyruvate kinases *pyk1* and *pyk2* (Table 2-4) using pET100_TOPO constructs (pJH01 and pJH02, Table 2-5).

*E. coli* growth curves were carried out in 250 ml flasks with a working volume of 50 ml of either LB or M9 medium with 1% glucose or 0.4% acetate as carbon source. The flasks were inoculated with from an overnight culture (1% v/v) of the strain of interest. Respective antibiotics were added as well as 1 mM IPTG to induce the expression. Samples were taken at regular intervals over 10h and an endpoint measurement after 24h. Growth was followed at $OD_{600}$ with dilutions when the $OD_{600}$ was higher than 0.4. The flasks were incubated at 37°C and 250 rpm. The specific growth rate was determined as a fitness indicator for each strain.

Growth of microbial cells can be described as increase in number of cells in a given time interval, expressed as growth rate, so number of cells per unit of time.

N= number of cells, t= time, μ=growth rate

$$\frac{dN}{dt} = \mu N$$

This equation will result in the following after integration- so for a certain time interval between t1 and t2 and the respective number of cells at t1 and t2 at a specific rate μ.

$$\ln\left(\frac{N2}{N1}\right) = \mu(t2 - t1)$$

During log phase cells will double in a certain time interval and if you plot the *ln* of the cell number over time you will get a linear correlation, which resulting slope equals the specific growth rate.

### 2.6.7 Spore viability of *Streptomyces* strains

A fresh spore preparation of the respective strain was diluted to 5 x $10^8$ cfu/ml and incubated for 10 min at 50˚C. The sample was then serially diluted by factor 10 six times and 10 µl were plated of each dilution in triplicates on nutrient agar with the respective antibiotic and were incubated for three days and at least three dilutions were counted and used to calculate the cfu/ml.

## 2.7 Molecular Biology

### 2.7.1 Preparation of Electrocompetent Cells and Electroporation in *E. coli* (Shigekawa and Dower, 1988; Sambrook and Russell, 2000)

A 5 ml aliquot of LB was inoculated with a single colony from a plate of the respective *E. coli* strain and grown overnight at 37˚C at 250 rpm. The overnight culture was used to inoculate 50 ml LB medium in a 250 ml flask and was grown at 37˚C until an $OD_{600}$ of 0.6 was reached. The cells were harvested by centrifugation for 5 min at 4000 rpm and were resuspended in 20 ml ice cold 10% glycerol (v/v). The washing was repeated and the cells were resuspended in 10% glycerol (v/v) and were left on ice for 30 min. The cells were aliquoted in to 50 µl volumes and either used directly or flash frozen in liquid nitrogen and stored at−80°C. To electrotransform, a 50 µl aliquot of electrocompetent cells was mixed with 1 µl of cosmid

DNA (approximately 0.01 µg DNA) and incubated on ice for 20 min. The mixture was transferred to an electroporation cuvette and was electroporated using a Biorad electroporator set at 1.8 kV and one pulse (pre-set program EC1). Following electroporation, 1 ml of LB medium was added to the cells and these were incubated at 37°C in a shaking incubator for 60 to 180 min. The cells were then spread on LB agar with the appropriate antibiotics and incubated overnight at 30°C or 37°C. Colonies were picked to inoculate 5 ml LB medium containing the respective antibiotics.

### 2.7.2 Preparation of Chemical Competent Cells and Transformation in *E. coli* (Sambrook and Russell, 2000)

A 5 ml aliquot of LB was inoculated with a single colony from a plate of the respective *E. coli* strain and grown overnight at 37°C at 250 rpm. 150 µl overnight culture was used to inoculate 50 ml LB medium in a 250 ml flask and were grown at 37°C until an $OD_{600}$ of 0.3-0.4 was reached. The cells were harvested by centrifugation for 10 min at 4000 rpm and were resuspended carefully using a pipette in 12.5 ml ice cold 100 mM $MgCl_2$. Cells were centrifuged for 10 min at 4000 rpm. Resuspension was carried out carefully in 3 ml of ice cold 100 mM $CaCl_2$, another 22 ml were added and the suspension was kept on ice for at least 20 min. The cell suspension was centrifuged at 4000 rpm for 10 min and was resuspended in 1 ml sterile 100 mM $CaCl_2$ in 20% glycerol. 100 µl were aliquoted and frozen at −80°C. A 50 µl aliquot of competent cells were mixed with plasmid DNA (1 − 5 µl) and were incubated on ice for 30 min. The mixture was heat shocked for 30 s at 42°C and placed immediately on ice for 2 min, then 1 ml of LB medium was added to the cells and were incubated at 37°C in a shaking incubator for 60 to 180 min. The cells were then spread on LB agar containing the appropriate antibiotics and incubated overnight at 37°C. Colonies were picked to inoculate 5 ml LB medium with the antibiotics.

### 2.7.3 *Streptomyces* mutant construction and verification

*Streptomyces* mutants were constructed using transposon insertion, the respective cosmid with the transposon was obtained from Dr Fernández-Martínez (John Innes Centre; see Table 2-6). The respective cosmid was first verified by restriction digest before it was transformed by electroporation into *E. coli* ET12567/puZ8002 (2.7.1) and then transferred into *Streptomyces* by conjugation (2.7.4). The strains were verified by PCR using one primer inside the insertion cassette and one either up or downstream of the gene, using primers that were designed for complementation (Table 2-7).

### 2.7.4 Intergenic Conjugation of Plasmids and Cosmids from *E. coli* to *Streptomyces* (Mazodier *et al.*, 1989; Kieser *et al.*, 2000)

An overnight culture of the *E. coli* strain ET12567/puZ8002 in LB with 50 µg/ml kanamycin and 25 µg/ml chloramphenicol as well as the respective antibiotic depending on the marker was transformed with the respective plasmid or cosmid and was grown to an $OD_{600}$ of 0.4 to 0.6 and cells were harvested by centrifugation for 5 min at 4000 rpm. The cells were washed twice with LB medium without any antibiotics and meanwhile a spore suspension of *S. coeliolor* was centrifuged and diluted to a spore number of 2 x $10^8$ spores per ml using 2xYT. The mixture was heated to 50°C for 10 min for germination. 500 µl of each organism were mixed and plated onto MS agar and incubated for 14-18 h. After that the plates were overlaid with nalidixic acid (25 µg/ml) and the selective antibiotic. The plates were incubated for two days at 30°C and the *Streptomyces* colonies were transferred onto nutrient agar (one with the respective antibiotic of the cassette containing plate and one with the same plus kanamycin containing plate) to screen for double crossover strains. Primary exconjugants were transferred onto nutrient agar with the respective antibiotics and were incubated for two days. The mutants are obtained by homologous recombination of the flanking genomic regions. A single crossover is when the entire cosmid is integrated, resulting in one disrupted copy of the gene plus the original gene copy, whereas a double crossover indicate an allelic replacement of the gene of interest and will lead to a replacement of the gene by two homologous recombination events in the flanking region. Thus screening for the loss of the cosmid backbone marker (here kanamycin) can be assumed to be double crossovers. The colonies were streaked on MS agar for the preparation of spore stocks for further characterisation.

### 2.7.5 Isolation of Plasmid and Cosmid DNA from *E.coli* Cultures using the Alkaline Lysis Method (Birnboim and Doly, 1979; Ish-Horowicz and Burke, 1981)

An aliquot of 1.5 ml of an overnight culture of the respective strain was harvested by centrifugation at 13000 rpm for 2 min and the pellet was resuspended in 100 µl of ice cold solution 1 (50 mM glucose, 10 mM EDTA, 25 mM TRIS-HCl, pH 8.0), then 200 µl of solution 2 (0.2 M NaOH, 1 % SDS) was added and mixed by inverting the tube. Finally 150 µl of solution 3 (3 M K-acetate, 11.5 % glacial acid) was added. The contents were mixed by vortexing shortly and were stored on ice for 5 min. The mixture was centrifuged at 13,000 rpm for 5

min and the supernatant was transferred to a fresh tube, two volumes of ethanol were added, mixed by inversion and allowed to stand at room temperature for 2 min. The tube was centrifuged for 5 min at 13,000 rpm and the supernatant was removed, the pellet was rinsed with 1 ml 70% ethanol and allowed to air dry at room temperature for 10 min. The pellets were resuspended in 50 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

### 2.7.6 Construction of Complementation Vectors

The general strategy for the construction of the complementation vectors consisted of the amplification of the target genes by PCR including the flanking region upstream to have the natural promoter of the gene present and then cloning it into pGEM-T Easy vector (Promega). Then the insert was cut out of the vector at the *EcoR*I site and was ligated into the pIJ6902 plasmid (Huang *et al.*, 2005, selective markers are apramycin and thiostrepton), which can be used to transform the transposon insertion mutants with the respective construct. Amplification of the inserts was carried out using MyTaq polymerase as specified in the suppliers manual (BioLine) in either 25 or 50 µl total volume. The PCR was carried out with an initial denaturation step at 95˚C for ten minutes, followed by 35 cycles of denaturation, annealing and extension for 60 s at 95˚C, 45 s at 50 -58 ˚C (see gene specific primer in Table 2-7 for each gene) and 60 s at 72 ˚C respectively followed by a final extension at 72 ˚C for 10 min. Additionally 2% DMSO were added to the reaction due the high GC content of the template.

### 2.7.7 Construction of Overexpression Vectors for Target Enzymes

The cloning for the overexpression of the target enzymes was carried out using the Champion™ pET100 Directional TOPO Expression Kit (Invitrogen). The primers were designed as described in the manual by the supplier. The inserts for the vector were obtained by PCR using High Phusion Polymerase (New England Biolabs). The PCR was carried out according to the supplier's instruction. GC-buffer and 2 % DMSO were used because of the high GC content of the DNA, in a reaction volume of 50 µl. The PCR consisted of an initial denaturation step at 98˚C for five minutes, followed by 30 cycles of denaturation, annealing and extension for 10s at 98˚C, 30s at 50 -58 ˚C (see gene specific primer in Table 2-7 for each gene) and 45s at 72 ˚C respectively followed by a final extension at 72 ˚C for 10 min.

### 2.7.8 LIGATION

Inserts were obtained either by PCR or by restriction digest from a plasmid and were purified with a PCR/Gel Purification Kit (Bioline), the respective vector was digested with the respective restriction enzyme and dephosphorylated with 1 µl thermosensitive alkaline phosphatase (TSAP, Promega) for 15 min at 37˚C and purified from the enzyme using a PCR cleanup Kit (Bioline). A re-ligation control with no insert was set up in parallel with different insert to vector ratios of 2:1, 3:1, 5:1 and 10:1 for some constructs. T4 DNA ligase (Promega) was used in the reaction, the reaction was set up on ice as specified by the manufacturer and left at 4˚C overnight before transformation into DH5α to check for positive clones.

### 2.7.9 *E.COLI* MUTANT CONSTRUCTION AND VERIFICATION

In order to obtain a pyruvate kinase double mutant in E. coli BW25113, the strain BW25113 Δ*pykA* (Baba *et al.*, 2006) was transformed with pIJ790 to introduce the λ red genes (Gust *et al.*, 2002). The plasmid is temperature sensitive and has to be grown at 30˚C and the genes are induced by 10 mM arabinose. This strain was then electroporated with the PCR product using pIJ773 (Gust *et al.*, 2003). The PCR was carried out using Q5 polymerase (New England Biolabs) with pIJ773 as template; conditions were chosen as described by the manufacturer. Additionally 5% DMSO as suggested by the manual were added to the reaction. The reaction was carried out with an initial denaturation step at 98˚C for two minutes, followed by 10 cycles of denaturation, annealing and extension for 45s at 98˚C, 45s at 50 ˚C and 45s at 72 ˚C respectively. Then 15 cycles of denaturation, annealing and extension for 45s at 94˚C, 45s at 55 ˚C and 90s at 72 ˚C respectively followed by a final extension at 72 ˚C for 10 min. Positive clones were selected for kanamycin (existing mutation) and apramycin resistance (pIJ773). The strain was then verified by PCR using MyTaq polymerase as described in section 2.7.6.

## 2.8 BIOCHEMICAL ANALYSIS

### 2.8.1 PROTEIN OVEREXPRESSION

Protein overexpression was carried out in *E. coli* Rosetta or Origami B on LB with 1 % glucose with a range of IPTG concentration for induction or on AIM medium under a range of temperatures. For specific protocols see results section 5.3.

### 2.8.2 PROTEIN PURIFICATION

Cells were disrupted by sonication for 2.25 sec per $OD_{600}$, so if the resuspended pellet had an $OD_{600}$ of 40, it was sonicated for 90 s in 30 s intervals with 30 s of pause in between to ensure the sample was not overheating (Chapter 5.4).

The cell extract was centrifuged for 30 min at 15,000 g and 4˚C. The supernatant was used to load a column of HisTrap TM FF crude (GE Healthcare) in Buffer A (100 mM $KH_2PO_4$ pH7.2, 10% glycerol (v/v), 100 mM NaCl, 20 mM imidazole) on Äkta purifier (GE Healthcare). Elution was carried out by a gradient with increasing imidazole concentration (Buffer B: 100 mM $KH_2PO_4$ pH7.2, 10% glycerol (v/v), 100 mM NaCl, 1 M imidazole). The eluted protein was collected in 1 ml fractions and the highest concentrations were pooled. In the case of Pyk1 it was necessary to concentrate the purified protein, this was performed with Merck MilliPore concentration columns with a cutoff of 10,000 MW.

### 2.8.3 GEL FILTRATION

In order to obtain an estimation of the size of the two pyruvate kinases, a gel filtration was carried out using a GE Healthcare HiLoad 16/60 Superdex200 prep grade. The standard was BioRad 151-1901 covering a range of 670 -1.35 kDa. The samples were concentrated to a volume of 2 ml. The buffer used was composed of 100 mM $K_2HPO_4$ (pH 7.2), 100 mM NaCl and 10 % Glycerol and was degassed. The column was equilibrated overnight and the standards were run first before loading the samples.

### 2.8.4 CIRCULAR DICHROMISM (CD)

Circular Dichromism measurements were determined at 20˚C using a Chirascan Plus (Applied Photophysics) with a 0.1 mm quartz cuvette. Fractions of the gel filtration were pooled and concentrated using Merck MilliPore concentration columns (cutoff 10,000 MW). Each sample was measured in duplicate in 180 nm to 280 nm range. Duplicate measurements were averaged and baseline (air) and buffer (100 mM $K_2HPO_4$, 100mM NaCl, 10% Glycerol) traces were subtracted from the peaks. All data analysis was completed using Global3 software (Applied Photophysics).

### 2.8.5 Protein concentration determination by Bradford Assay (Bradford, 1976)

The Bradford Assay Reagent (BioRad) was diluted 1:5 to a working solution and was then mixed with the sample or standard. The standard for a calibration curve consisted of a stock solution of 1 mg/ml BSA which was then diluted to cover a range of 2 µg/ml to 50 µg/ml. The assay volume was of 1 ml. A fresh standard for calibration was prepared each time. Usually 5 µl of sample or standard (diluted appropriately) were mixed with 995 µl of Bradford Reagent by vortexing and incubated for 5 min at RT. The absorbance was measured at 595 nm. The calibration curve was used to determine the protein content of the unknown samples.

### 2.8.6 SDS-PAGE (Laemmli, 1970)

SDS-PAGE has been performed as described by GE Healthcare using the Amersham ECL Gel horizontal electrophoresis system. Briefly, samples were prepared by mixing the respective protein sample 1:2 with 2 x SDS sample buffer (63 mM Trizma Base, 2% SDS, 10% glycerol, 5% β-mercaptoethanol and 0.001% bromophenol blue) and incubated for 5 min at 75˚C. For the fractions collected during the purification, the cell extract and cell pellet 10 µl and 1 ml in case of the samples directly from the flask. The latter samples were centrifuged for a minute, 900 µl of the supernatant was discarded and the remaining 100 µl were mixed with 100 µl of SDS sample buffer and were treated as mentioned above. SDS-PAGE was carried out on 4-12 % gradient gels (GE Healthcare) and each gel was run at 160 V for 60 min in SDS Running buffer (10 x stock: 30 g/l Trizma Base 144 g glycine, 10 g/l SDS).

## 2.9 Pyruvate kinase activity Assay (Bergemeyer *et al.*, 1974)

Purified pyruvate kinase samples were used for the determination of the kinetic characteristics of each pyruvate kinase. The assay is a coupled assay with the following reactions:

$$Phosphoenolpyruvate + ADP \rightarrow Pyruvate + ATP$$

$$Pyruvate + \beta - NADH \rightarrow L - Lactate + \beta - NAD$$

The first reaction is carried out by pyruvate kinase in presence of $Mg^{2+}$ and $K^{+}$ and the second reaction by L-lactic dehydrogenase (LDH). The conditions were of 30°C, pH 7.2, light path of

1 cm and following the reaction at $A_{340}$ for 5 min. The reaction setup can be found in Table 2-8.

Table 2-8 Reaction setup for pyruvate kinase assay

| Component (stock) | Assay concentration | Test (µl) | Blank (µl) |
|---|---|---|---|
| $dH_2O$ | - | Ad 1476 | Ad 1476 |
| 100 mM $K_2HPO_4$ pH 7.2 | 39 mM | 400 | 450 |
| 17 mM PEP | 0 -7 mM | variable | - |
| 1.3 mM β-NADH | 0.11 mM | 125 | 125 |
| 100 mM $MgSO_4$ | 6.8 mM | 100 | 100 |
| 44 mM ADP | 0 – 5 mM | 50 | - |
| 5000 U/ml LDH | 10 U | 1 | 1 |
| pyruvate kinase | | 50 | 50 |
| Total volume [µl] | | 1476 | 1476 |

The monitored kinetics, for each condition carried out in triplicates, were utilised to calculate the specific activity of the enzyme (U/mg). First the activity per volume was calculated as following, where 1 unit equals 1 µM of phosphoenolpyruvate to pyruvate per minute at pH 7.2 at 30°C, V= reaction volume in ml, DF = dilution factor, E = molar extinction coefficient for β-NADH and $V_e$ the volume of enzyme added to the reaction in ml. U/ml can easily be converted into U/mg protein by dividing the value with the protein concentration of the purified sample determined with Bradford's reagent.

$$\frac{U}{mL} = \frac{(\Delta A_{340}\,(Sample) - \Delta A_{340}(\text{Blank})) * V * DF}{\varepsilon\, NADH\, Ve}$$

$$\frac{U}{mg} = \frac{U/ml}{mg/ml}$$

The specific activities were then used to calculate the kinetic characteristics.

Controls of the reactions were all components without either of the two substrates or without one substrate or LDH or pyruvate kinase. The blank sample used for the calculation the no substrate sample was used.

## 2.10 RNA ISOLATION

*S. coelicolor* biomass of 15 ml sample from liquid culture was taken (RNA-Seq) or biomass scraped of a cellophane disc from an agar medium (semi quantitative RT-PCR) was used as starting material. Sterile cellophane discs (previously cut and boiled 10 min in $dH_2O$ in the

microwave, then placed between pre-wetted Whatman No.1 paper and autoclaved) were inoculated with $3x10^6$ spores, which had been germinated and washed in $dH_2O$. The plates were incubated at 30° C. The samples were then spun down for 5 min at 4˚C and 4000 rpm, the supernatant discarded and 2 ml of RNAprotect (Qiagen) were added and left at room temperature for 5 min. RNAprotect (Qiagen) was immediately added to the biomass from agar plates. The samples in RNAprotect (Qiagen) were centrifuged (5 min 4˚C, 4000 rpm) and the supernatant was discarded. The biomass was then resuspended in 1ml 1x TE buffer containing 15 mg/ml lysozyme. Tubes were vortexed for 10 s and incubated at room temperature for 60 min whilst shaking. 1 ml RLT buffer (Qiagen RNA Isolation Kit) + 10 µl β-mercaptoethanol was added and vortexed, then one volume of chloroform/phenol (2 ml) was added and the mix was vortexed for 30 s followed by centrifugation for 10 min at 4˚C and 4000 rpm. The aqueous phase (upper layer) was transferred to a fresh tube, the volume estimated and the same extraction step with one volume phenol/chloroform as above was repeated. Finally, the aqueous phase was mixed with one volume chloroform, vortexed for 30s followed by centrifugation for 10 min at 4˚C and 4000 rpm. Then 1.9 ml ethanol was added to 1.5 ml of sample. This lysate was then purified using a commercial RNA isolation Kit (Qiagen). Briefly, the lysate was added to a column, the RNA will bind to the matrix and the RNA was washed with several buffers until eluted in 30 µl RNAse free water and was then ready for DNAse treatment and quantification.

## 2.11 DNAse Treatment and RNA Quantification

The isolated RNA was treated with RNAse free DNase (Ambion, Life Technologies) as specified by the manufacturer. Briefly 0.1 volume of 10 x Buffer and 1 µl of DNAse were added to the RNA and incubated for 30 min at 37˚C, then the DNase inactivation reagent was added and incubated for 2 min at RT with occasional mixing. The mixture was then centrifuged and the RNA in the supernatant was transferred to a fresh tube.

Quantification of RNA was carried out using Qubit®, which uses specific fluorescent dyes for the respective biomolecule for accurate quantification. Briefly 1-20 µl of sample were mixed with the fluorescent reagent to a total volume of 200 µl and incubated for 2 min and then measured against two standards provided with the kit. A detailed protocol can be found in the supplier's manual called Qubit RNA-HS Assay (Life Technologies).

## 2.12 Quality Assessment of RNA Samples

The quality and integrity of the RNA was assessed by using the Bioanalyzer (Agilent). It is a chip based capillary electrophoresis machine using gel chips. The run was carried out according to the supplier's manual. Before preparing the chip, the appropriate program is loaded and the sample loading has to be assigned. Briefly, a gel was prepared by adding the fluorescent dye to it and left to equilibrate for 30 min. A ladder is provided with the Kit. After washing the electrodes of the instrument a new RNA chip was put into the station and was filled with prepared gel and the priming station was used to pressurize the gel across the chip. The marker and the ladder were then loaded to the respective well on the chip as well as 1 µL of sample to each sample wells. Unused wells were loaded with water. The chip was then vortexed at 2000 rpm (IKA vortex mixer) for exactly 60 s. The chip was then inserted to the machine and the run was started.

Furthermore the RNA samples were also used as templates for a generic PCR in order to check for DNA contamination using primers for *hrdB* (SCO5820, Table 2-7). This PCR was performed as described in section 2.7.6.

## 2.13 RT-PCR

RNA isolated from biomass grown on solid medium was analysed for gene expression of genes of interest over time using RT-PCR. The one-step RT-PCR kit (Qiagen) was used using 50 ng of total RNA as template. The PCR was carried out according to the supplier's instructions. The PCR consisted of the reverse transcription at 50˚C for 30 min and inactivation of the enzyme at 95˚C for 15 min. The PCR was performed by 30 cycles of denaturation at 94˚C for 45 s, annealing 45 s at 60˚C (see gene specific primer in Table 2-7 for each gene) and extension for 60 s at 72˚C and a final extension at 72 ˚C for 10 min. The resulting DNA was then run on 2 % agarose gel and analysed for the presence and absence of the genes in comparison to *hrdB* as reference gene constitutively expressed.

## 2.14 RNA SEQUENCING

This work was carried out in collaboration with Richard Reumerman. A general workflow can be found in Figure 2.3. Triplicate growth curves of *S. coelicolor* WT M145 in HMM with either glucose or Tween 40 as sole carbon source (0.33 M carbon) were carried out. Samples were taken at log phase for analysis.

### 2.14.1 ENRICHMENT OF MRNA BY RRNA DEPLETION

RNA samples from liquid cultures were used for sequencing after they were analysed for their purity and integrity. In order to enrich the mRNA, which was the RNA of interest, rRNA depletion was performed (rRNA depletion Kit Ribo Zero Magnetic Kit for Gram-positive bacteria from Epicentre (Illumina)). The kit uses biotinylated capture probes to hybridise to the rRNA, the rRNA-probe hybrid is bound to magnetic beads and removed by a magnet. The remaining solution contains the desired mRNA in solution.The protocol of the supplier was followed. Briefly, 2.7 µg of total RNA were resuspended in 26 µl mixed plus 10 µl rRNA removal solution and 4 µl of 10 x Reaction Buffer (total volume 40 µl). The reaction was gently mixed and incubated at 68°C for 10 min followed by 5 min at RT. The mix was then transferred to prepared tubes containing washed Magnetic Beads and mixed by pipetting at least 10 times followed by 10 sec vortexing. After 5 min incubation at RT the mix was vortexed for 10 sec and transferred to 50°C for 5 min, followed by being put on a magnetic stand to remove the beads and to recover the mRNA in the supernatant to a fresh tube. The rRNA depleted sample was then precipitated with ethanol as described in the manual. The quality and integrity of the samples were then analysed on the BioAnalizer (Agilent) and the concentration was determined using Qubit (Life Technologies, 2.11).

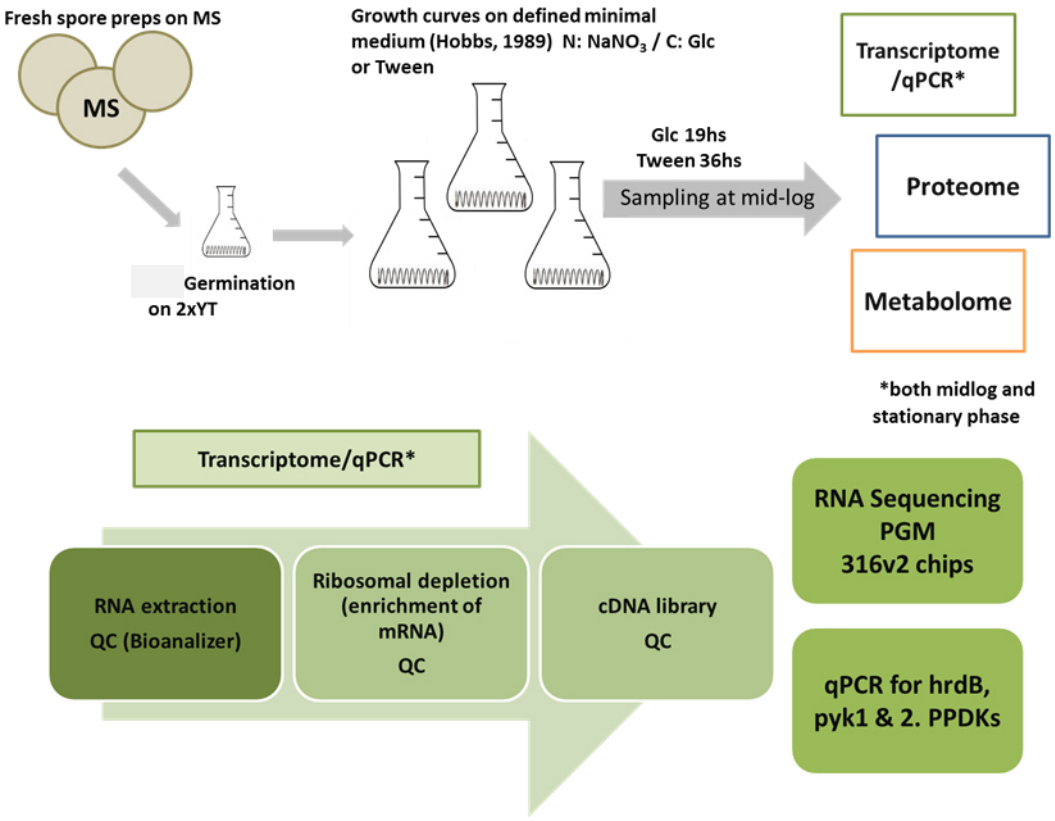



Figure 2.3 Workflow for gene expression analysis on different carbon sources: Spores from WT M145 were grown on MS agar for three days and used to inoculate 50 ml of 2xYT for germination. Germinated spores were washed and used to inoculate 400 ml of HMM medium supplemented with either glucose or Tween40. Samples were taken at log (19 h for glucose and 36 h for Tween) and stationary phase (36 h for glucose and 42 h for Tween) for RNA extraction. For the log phase samples rRNA was depleted and cDNA was prepared and a library for sequencing was prepared. RNA Seq was performed on Ion Torrent PGM (316v2) chips. For the remaining total RNA samples cDNA was synthesized and analysed by qPCR for five different genes (*hrdB*, *pyk1, pyk2*, *ppdK1*, *ppdK2,* Chapter 6). At the same time, there were also samples taken at log phase for metabolome analysis (Chapter 6)

### 2.14.2 cDNA SYNTHESIS AND LIBRARY PREPARATION

cDNA synthesis and library preparation was carried out according to the User Guide Ion Total RNA-Seq Kit v2 Revision E from Ion Torrent, Life Technologies. It is summarised briefly below. The samples were spiked with ERCC Spike In Mix Control, which is a set of 92 transcripts commercially available. After using megablast no signifcant similarities were found with the *S. coelicolor* genome. Thus it was decided to use it as an internal quality control. However the plug-in analysis tool was faulty and gave different and contradictory results depending on the chosen settings. Technical support from Life Technologies was contacted, however they were not able to explain the outcome after analysing all the data and quality checks of the runs. Their recommendation is now not to use this product in combination with any prokaryotic samples.

The enriched mRNA samples were fragmented enzymatically using RNAse III in order to obtain sequenceable fragment size. The samples were dried and resuspended in RNase-free water. The manufacturer's protocol for less than 100 ng rRNA depleted samples was followed. After fragmentation the RNA was purified with magnetic beads and then hybridized and ligated the adapters with an individual barcode for each sample to the fragments followed by RT-PCR to obtain cDNA. The cDNA was then purified using magnetic beads and amplified and afterwards again purified with magnetic beads. At this stage the yield and size distribution of the amplified cDNA was carried out using the BioAnalizer (Agilent). Each library sample was diluted to a concentration of 500 pM and then three samples were pooled since the samples contained barcodes and could be separated during the data analysis. The three glucose samples and the three Tween samples were pooled to a concentration of 20 pM as specified by the supplier.

### 2.14.3 SEQUENCING

Ion OneTouch 2 system was used for the template preparation which includes the steps of emulsification, amplification and enrichment of the library. Ion PGM Template OT200 Kit (Life Technologies) was utilised. The pooled libraries, each at 20pM concentration, were mixed with the reagents in order to bind the fragments to the Ion sphere particles and to amplify them, the mix was then loaded onto the Ion Touch Plus Reaction Filter connected to the Reaction Tube followed by reaction oil and it was inserted to the machine. At the end of the run, the sample was recovered carefully from the two tubes by discarding all but 50 µl of the supernatant. The library was deposited as a pellet at this stage composed of the amplified

ion sphere particles with each sphere carrying only one distinct molecule in many copies bound to it. The sample was then enriched with the enrichment system (ES) where the loaded spheres were bound to streptavidin beads and any empty spheres were washed off. The sample was quality checked by the quality control assay using Qubit (Life Technologies). There are two different dyes, each binds specifically to either one of the primers. One of the primers is only present on extended templates whereas the other is present on all ion sphere particles. By the ratio of the two fluorescent signals, it yields percentage of template ion sphere particles is determined. The optimal amount should be in the margin of 10 -30% as specified by the supplier.

The sequencing of the samples was carried out on Ion Torrent Personal Genome Machine (PGM) System using a 316v2 chip following the procedures in the manual Ion PGM Sequencing 200 Kit v2 User Guide Revision 3.0.

First of all the run has to be planned on the Torrent Browser and was carried out as specified in Table 2-9.

Table 2-9 Specification of RNA Sequencing Run on PGM

| **Type** | **Specification** |
|---|---|
| Application | RNA Seq |
| Run Type | Forward |
| Template Kit /Sequencing Kit | Template OT2 200 Kit/ Sequencing 200 Kit v2 |
| Flows | 500 |
| Barcode Set | IonXpressRNA BC01-06 |
| Reference library | *Streptomyces coelicolor* (NC_038888) |
| Plugins | Alignment, coverage analysis, ERCC-Analysis, FastQC, FastqCreator |
| Project | Streptomyces_RNA-seq_Jana_and_Richard |
| Name | R_2013_07_05_07_49_27_user_SYB-31-Glucose_RNA-Seq_midlog_23JUL14/ R_2013_07_05_11_45_31_user_SYB-32-Tween_RNA-Seq_midlog_23JUL14 |
| Sample Name | GlcA, GlcB, GlcC / TweA, TweB, TweC |

Before use, the machine was cleaned and initialized according to the manual. The sample was then mixed with Control Ion Sphere particles and followed by the annealing of sequencing primer to the sample. Then, the sequencing chip needs to be checked before loading the sample onto it by running the Chip Check option on the machine. It is important to be grounded and the chip should never be touched with gloves. After annealing the sequencing primer, the polymerase was added to the reaction. After 5 min of incubation the

sample was loaded according to the manual without introducing any air bubbles. The loaded chip was then inserted to the machine and the planned run was selected and run.

The Ion Torrent technology is one of the new generation sequencing (NGS) technologies and relies on the fragmentation of the molecules to a size of roughly 200 bp to which adapters are added and each molecule is bound to an ion sphere where the bound molecule is amplified in an emulsion. Each ion sphere after enrichment and getting rid of empty spheres is placed into a single well of the chip. The chip is then flooded with one of the dNTPs together with the respective buffer and polymerase. The pH is detected in each well, the addition of a base release a $H^+$ and therefore decreases the pH, thus the change of pH is a signal for the incorporation of that particular dNTP. The chip is flooded with another dNTP and this is repeated 500 times cycling through a random order of dNTPs. Therefore the sequence of each individual fragment in each of the wells will be known at the end of the run and can be analysed *in silico*.

### 2.14.4 Data Analysis

The sequencing data were downloaded from the Ion Torrent server version Torrent Suite 4.0.2 in Fastq format and were processed with different software packages: Rockhopper (McClure *et al.*, 2013), CLC Genomics Workbench 7.5 (Qiagen) and RNA Rocket (Warren *et al.*, 2015).

The general workflow was trimming the reads of low quality data at the ends of the sequence, then the reads were mapped to the reference genome and were counted. Statistical analysis was performed on the counts and comparisons of different conditions could be made. This usually also included a normalisation of the data to account for the different gene sizes. Detailed specification of each analysis pipeline can be found in the appendix (10.2).

## 2.15 qPCR

qPCR was performed on the same samples as the RNA Sequencing in order to confirm the results obtained. Additionally to the log measurement which was analysed by RNA Seq, the stationary phase of the same growth curves was analysed for the expression of five genes: SCO2014 (*pyk1*), SCO5423 (*pyk2*), SCO0208 (*ppdK1*) and SCO2494 (*ppdK2*) from central carbon metabolism and (SCO5820) *hrdB* as reference gene.

### 2.15.1 PRIMER

The recommended amplicon length for qPCR is of 80 – 200 bp, never exceeding 400 bp as recommened by the manufacturer. The melting temperature was of around 60˚C using default Primer 3 settings (http://frodo.wi.mit.edu/primer3/). Each primer pair was tested using genomic DNA as template and were dissolved in RNase free water.

The primer concentrations were optimised to decrease primer dimer formation and therefore increase PCR efficiency. For this five different concentrations of each primer were tested: 1000 nM, 500 nM, 250 nM, 125 nM and 62.5 nM.

### 2.15.2 cDNA SYNTHESIS AND QUANTIFICATION

First of all cDNA was synthesized from the RNA samples using qPCRBIO cDNA synthesis Kit (PCR Biosystems) following the instructions of the manual supplied. Total RNA at a concentration of 1.28 µg were used as template and running 2 reactions per sample. The synthesis was carried out at 50˚C for 30 min due to high GC-content, followed by 55˚C for 10 min and on hold at 12˚C until use. Quantification of the cDNA was carried out by using the QuantiFluor ssDNA system (Promega) with the Qubit Fluorometer (Life Technologies). The technique has a similar principle as the Qubit Kit from Life Technologies.

### 2.15.3 qPCR REACTIONS

All cDNA samples were diluted to a concentration of 10 ng/ µl. Each reaction contained 10 ng of cDNA (1 µl) and were then mixed with 10 µl MasterMix (Kit 2x qPCRBIO SyGreen Mix Lo-ROX from PCRBIOSYSTEMS), 2.5 µl of each primer and 4 µl $dH_2O$ to a final reaction volume of 20 µl. The PCR machine used was a Corbett Research 6000 (Qiagen) with specific tubes for its rotor obtained from Qiagen. The program was a Three Step with Melt composed of once 3 min at 95˚C followed by 40 cycles of 5 s at 95˚C and once 25 s at 60˚C. Each reaction was carried out in duplicate and a no template control was always run per each set of primers.

In order to quantify the amount of each gene of interest in the total sample it was necessary to obtain standard curves for each gene. As template served a purified PCR product from a PCR with genomic DNA, this template was diluted to get in total seven different standards ranging from $10^1$-$10^7$ molecules. The standard curves were prepared in triplicate and were used to calculate the concentration in the unknown samples and were then compared to the results obtained in the RNA Sequencing.

## 2.16 METABOLOMICS

An aliquot of 100 ml of *S. coelicolor* liquid culture grown in HMM plus glucose or Tween40 was filtered through glass fibre paper by vacuum filtration and the biomass was washed twice with HMM medium without any carbon source. The biomass was then transferred to pre-cooled -80˚C Methanol/Chloroform/d$H_2O$ (3:1:1) and was incubated with shaking for 1 h at 4˚C for the extraction of metabolites. The debris were then spun down for 20 min at 4,000 rpm and the supernatant was transferred to a fresh tube and stored at -80˚C before analysis. Furthermore spent and fresh medium were also analysed and were prepared using the same procedure as for the biomass to check for metabolite leakage as well as secretion into the medium. The samples were then analysed by LC-MS in-house by Dr Tong Zhang at the institute using a ZIC-pHILIC column (Zhang *et al.*, 2014).

The raw data of the metabolite standard solutions were processed using ToxID 2.1 (Thermo Fisher Scientific) with ± 3 ppm mass accuracy tolerance with both ESI positive and negative modes. The generated extracted ion chromatograms of metabolite standards on different columns were visually evaluated with respect to peak shapes and their retention times.

The raw data files were converted into mzXML open format and chromatograms were extracted using detection algorithm from XCMS (Smith *et al.*, 2006) and stored in corresponding PeakML files (Scheltema *et al.*, 2011). Those were then aligned and combined using mzMatch.R (Scheltema *et al.*, 2011) after filtering out all peaks that were not reproducibly detected. Finally these were imported into IDEOM for putative metabolite identification based on accurate mass (±3 ppm) and retention time prediction from the standards (Creek *et al.*, 2012). Metabolites of interest were extracted and analysed in the different samples and plotted in Prism (GraphPad).

# 3 Central Carbon Metabolism in *S. coelicolor* – Gene Expansions and the Phosphoenolpyruvate-Pyruvate-Oxaloacetate Node

## 3.1 Enzyme Expansion Analysis of Primary Metabolism in *Streptomyces*

To establish if gene expansion in precursor supplying pathways is a contributing factor to the diversity of specialised metabolites in the genus of *Streptomyces*, genomes of 614 Actinobacteria covering in total 83 genera, were analysed. On the date of search, 311 genera were listed in the NCBI Taxonomy database (accessed on 26.02.2015). Of those 138 had no genome sequence available and 7 categories were unclassified, leaving 166 genera from which to select genomes. Each selected genome (10.3) was re-annotated on RAST (Aziz *et al.*, 2008) to ensure consistency and the number of gene annotations per functional role in each genome was counted. The mean number of gene annotations per functional role was calculated per genus and those that had an equal or higher value than the mean plus its standard deviation were highlighted as a putative genetic expansion. Here the analysis is focussed on central carbon metabolism and amino acid metabolism, as they are the major supply routes for specialised metabolism.

Generally the analysis revealed that gene expansion events are more frequent in genera that are known to produce specialised metabolites (Tokovenko *et al.*, 2016), supporting the hypothesis of an expanded primary metabolism being linked to a rich specialised metabolite repertoire (Table 3-1). Pathway expansion was assessed by percentage of species per suborder showed expansions in the pathway of interest. Overall Catenulisporineae and Pseudonocardineae showed the most expansions throughout central carbon metabolism. When looking at each pathway in particular however there were also other groups showing up. Glycolysis showed highest expansion in Streptomycineae and Catenulisporineae with 23.3% and 25.0% respectively. Pseudonocardineae showed highest expansion for gluconeogenesis with 25.0%, 28.3% for TCA, as well as the amino acid metabolic pathways derived from AKG (20.1%), Pyr (31.3%), OAA (20.7%), 3-PGA (24%) and E4P/PEP (19.4%). Pentose phosphate pathway showed highest expansion in Catenulisporineae with 42.9%. Amino acid metabolism derived from ribose-5-phosphate was most expanded in

Micromonosporineae with 23.3%. The amino acid metabolisms derived from AKG was also expanded for 21.4% of genes in Catenulisporineae. Additionally metabolic functions from amino acids derived from OAA were also expanded in 21.1% of the genes in Micromonosporineae, Bifidobacteriales and Kineosporineae. In Catenulisporineae 25% of the genes in amino acid metabolism with 3PGA as precursor were expanded (Table 3-1).

Table 3-1 Percentage of metabolic expansions per pathway and suborder, the highest percentage for each pathway is highlighted in yellow- Legend: GNG = gluconeogenesis, PPP = pentose phosphate pathway, TCA = tricarboxylic acid cycle, AA = amino acids, AKG = α-ketoglutarate, PYR = pyruvate, OAA = oxaloacetate, 3PGA = 3-phosphoglycerate, R5P = ribose-5-phosphate, E4P = erythrose-4-phosphate, PEP = phosphoenolpyruvate

| Suborder | Glycolysis | GNG | PPP | TCA | AA AKG | AA PYR | AA OAA | AA 3PGA | AA R5P | AA E4P/PEP |
|---|---|---|---|---|---|---|---|---|---|---|
| **Streptomycineae** | 23.3 | 8.3 | 4.8 | 8.3 | 11.9 | 11.1 | 10.5 | 16.7 | 0.0 | 14.0 |
| **Catenulisporineae** | 25.0 | 12.5 | 42.9 | 25.0 | 21.4 | 16.7 | 13.2 | 25.0 | 10.0 | 15.8 |
| **Streptosporangineae** | 6.7 | 11.1 | 3.2 | 13.2 | 10.3 | 17.3 | 11.7 | 11.1 | 3.3 | 12.3 |
| **Frankineae** | 6.7 | 0.0 | 23.8 | 0.0 | 11.9 | 0.0 | 8.8 | 11.1 | 10.0 | 8.8 |
| **Pseudonocardineae** | 10.6 | 25.0 | 17.0 | 28.3 | 20.1 | 31.3 | 20.7 | 24.0 | 17.5 | 19.4 |
| **Corynebacterineae** | 7.5 | 9.4 | 5.4 | 10.2 | 8.7 | 8.3 | 6.6 | 6.3 | 7.5 | 5.3 |
| **Micromonosporineae** | 20.0 | 13.9 | 15.9 | 16.7 | 14.3 | 18.5 | 21.1 | 13.0 | 23.3 | 13.5 |
| **Glycomycineae** | 0.0 | 0.0 | 14.3 | 0.0 | 7.1 | 0.0 | 5.3 | 0.0 | 0.0 | 0.0 |
| **Micrococcineae** | 6.5 | 5.9 | 5.0 | 5.5 | 4.6 | 2.6 | 8.4 | 6.9 | 3.5 | 4.0 |
| **Bifidobacteriales** | 0.0 | 0.0 | 14.3 | 0.0 | 0.0 | 11.1 | 21.1 | 0.0 | 10.0 | 0.0 |
| **Actinomycineae** | 0.0 | 0.0 | 7.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 5.3 |
| **Kineosporineae** | 20.0 | 0.0 | 28.6 | 0.0 | 14.3 | 0.0 | 21.1 | 0.0 | 20.0 | 5.3 |
| **Propionibacterineae** | 13.8 | 12.5 | 7.1 | 9.4 | 7.1 | 18.1 | 11.8 | 12.5 | 7.5 | 4.6 |

For a detailed overview of the expansions organised by pathway in each genus can be found in section 10.4 and a summary for the genus of *Streptomyces* can be found in Table 3-2.

The expansion table was organised according to an underlying species tree based on the β-subunit of RNA polymerase (RpoB, Figure 3.1), giving a higher resolution than 16SrRNA for closely related species (Case *et al.*, 2007).

In total, 14 genes in *Streptomyces* were highlighted as gene expansions, meaning that they were overrepresented in this genus compared to the rest of the genera in the database. The expanded enzyme families were as follows - phosphofructokinase, pyruvate kinase, pyruvate phosphate dikinase, malic enzyme, pyruvate dehydrogenase complex E1, chorismate mutase, acetylglutamate kinase, diaminopimelate decarboxylase, aspartate aminotransferase, aspartate-semialdehyde dehydrogenase, serine hydroxymethyltransferase, glutamine synthetase, arginiosuccinate lyase and methionine synthetase (see Appendix 10.4, Table 3-2).

In addition to the 14 expansions identified, another three potential gene expansions that were close to the set threshold were identified from the evolutionary analysis. These are included in the analysis as they represent key gene expansions in several species and may represent species-specific expansion events: enolase, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and transaldolase (Table 3-2).

Table 3-2 Overview of expansion hits by comparing mean of number of annotation per functional role for the genus *Streptomyces* and *S. coelicolor* to the overall mean plus standard deviation (Std) of the whole database in central carbon and amino acid metabolism

| Enzyme | Overall Mean+Std | *Streptomyces* / *S.coelicolor* A3 (2) | Gene ID (SCO) in *S.coelicolor* genome |
|---|---|---|---|
| **Phosphofructokinase (EC 2.7.1.56)** | 1.39 | 1.76/ 3 | 1214/2119/5426 |
| **Pyruvate kinase (EC 2.7.1.40)** | 1.34 | 2.00/ 2 | 2014/5423 |
| **Pyruvate Phosphate Dikinase (EC 2.7.9.1)** | 1.36 | 1.37/2 | 0208/2494 |
| **Malic enzyme (EC 1.1.1.38)** | 1.12 | 1.50/2 | 2951/5621 |
| **$PDHC_{E1}$ (EC 1.2.4.1)** | 2.04 | 2.18/3 | 7124/2183/2731 |
| **Chorismate mutase (EC 5.4.99.5)** | 1.86 | 2.66/2 | 1762/2019 |
| **Acetylglutamate kinase (EC 2.7.2.8)** | 1.31 | 1.31/1 | 1578 |
| **Diaminopimelate decarboxylase (DPD, EC 4.1.1.20)** | 3.35 | 3.57/3 | 0315/5353/6438 |
| **Aspartate aminotransferase (AAT, EC 2.6.1.1)** | 10.88 | 10.93/10 | 1284/1577/3306/3658/4645/ 5941/6222/6412/6769/7537 |
| **Aspartate-semialdehyde dehydrogenase (ASAD, EC 1.2.1.11)** | 1.57 | 1.92/2 | 2640/3614 |
| **Serine hydroxymethyltransferase (SHMT, EC 2.1.2.1)** | 2.37 | 2.94/3 | 4837/5364/5470 |
| **Glutamine synthetase (GS, EC 6.3.1.2)** | 3.69 | 4/ 5 | 1613/2210/2241/2198/6962 |
| **Arginiosuccinate lyase (ASL, 4.3.2.1)** | 1.85 | 1.97/1 | 1570 |
| **Methionine synthase (5HMT, EC 2.1.1.13)** | 1.7 | 1.7/2 | 1476/6137 |
| **Enolase (EC 4.2.1.11)** | 1.58 | 1.50/2 | 3096/7638 |
| **GAPDH (EC 1.2.1.12)** | 1.86 | 1.81/3 | 1947/7040/7511 |
| **Transaldolase (EC 2.2.1.2)** | 1.56 | 1.49/2 | 1936/6662 |

## 3.2 Phylogenetic Analysis of Members of the PEP-PYR-OAA node and the Metabolic Expansions in *S. coelicolor*

To further characterise the gene expansion events identified, a phylogenetic analysis of each of the genes was undertaken to examine the relationship between the expanded gene families and the overall phylogeny of the phylum. This will allow the identification of specific duplication events to be separated from horizontal transfer events. To achieve this, a species level phylogeny was constructed using the β-subunit of RNA polymerase (RpoB) which is a good marker for overall phylogeny in closely related bacteria (Case *et al.*, 2007; Kämpfer and Glaeser, 2012; Vos *et al.*, 2012). This phylogenetic analysis provided a quality check of each specific genome annotation and a reference tree for comparison of evolutionary congruence. The RpoB tree could be divided in three distinct phylogenetic branches, one composed of Streptomycineae and Catenulisporinae, second including Propionibacterineae, Actinomycineae, Bifidobacteriales and Micrococcineae and the third one with Micromonosporineae, Glycomycineae, Corynebacterineae, Pseudonocardineae, Frankineae and Streptosporangineae (Figure 3.1). Five species were misplaced on the tree and grouped with other species rather than the rest of their genus: *Streptomyces carneus* (ID: 103200_819) was in *Nocardia* genus cluster (suborder Corynebacterineae), *Streptomyces NRRL F3213* (ID: 104209_3190) showed in the *Amycolatopsis* cluster (suborder Pseudonocardineae), *Rhodococcus rhodni* (ID:104518_2496) as *Kibdelsporangium* (suborder Pseudonocardineae), *Actinospica acidiphila* (ID: 104678_6451) and *Microtetraspora glauca* (ID:104669_4572) both as *Streptomyces* (suborder Streptomycineae).

Out of the 17 possible expansion hits detected for *Streptomyces*, some were not an expansion in *S. coelicolor* A3 (2) which had one gene annotated for this functional role, as in the case of acetylglutamate kinase (AGK) and arginiosuccinate lyase (ASL). Additionally, the alignments of aspartate aminotransferase (AAT), pyruvate dehydrogenase and glutamine synthase (GS) were found to be poor representing significant divergence in sequence identity across the representatives and were thus not used for a tree construction. In addition to the expansion hits, all members of the metabolic node were analysed for their evolutionary history. Table 3-3 gives an overview of the number of hits found by blastp compared to number of genes found in the RAST annotated across the whole database, as well as the e-value utilised for the blast search.

Table 3-3 Overview of number of genes per functional role using blastp and RAST of the expansion analysis and the corresponding e-value which was used as cutoff, Legend RpoB = β-subunit RNA polymerase, Pyk= pyruvate kinase, Ppdk = pyruvate phosphate dikinase, PDHC = pyruvate dehydrogenase complex, ME = malic enzyme, PEPCk = phosphoenolpyruvate carboxykinase, PEPCx = phosphoenolpyruvate carboxylase, Pyc= pyruvate carboxylase, GAPDH = glyceraldehyde-3-phosphate dehydrogenase, CM = chorismate mutase, AKG = acetylglutamate kinase, ASAD = Aspartate-semialdehyde dehydrogenase, SHMT = Serine hydroxymethyltransferase, ASL= arginiosuccinate lyase , 5HMT = methionine synthase, GS=glutamine synthase , DPD= Diaminopimelate decarboxylase , AAT = aspartate aminotransferase *best hit edge used instead of e-value, because it did not seem to be stringent enough

| Enzyme | Query (SCO) | EC | e-value | # hits in blastp | # of genes from RAST |
|---|---|---|---|---|---|
| **RpoB** | 4654 | 2.7.7.6 | 0.001 | 618 | 636 |
| **Pyk1** | 2014 | 2.7.1.40 | 0.001 | 922 | 928 |
| **Ppdk2** | 2494 | 2.7.9.1 | 1E-16 | 621 | 622 |
| **PDHC E1** | 7124 | 1.2.4.1 | 0.0000001 | 1090 | 1084 |
| **ME** | 2951 | 1.1.1.38 | 0.0001 | 1057 | 471 |
| **PEPCk** | 4979 | 4.1.1.32 | 0.0001 | 608 | 660 |
| **PEPCx** | 3127 | 4.1.1.31 | 0.0001 | 438 | 477 |
| **Pyc** | 0546 | 6.4.1.1 | 0.25* | 227 | 97 |
| **$PDHC_{E1\alpha}$** | 1270 | 1.2.4.1 | 0.0001 | 2851 | 1258 |
| **$PDHC_{E1\beta}$** | 1269 | 1.2.4.1 | | | |
| **GAPDH** | 7040 | 1.2.1.12 | 0.001 | 1291 | 929 |
| **transaldolase** | 1936 | 2.2.1.2 | 0.001 | 818 | 821 |
| **enolase** | 3096 | 4.2.1.11 | 0.001 | 806 | 800 |
| **CM** | 2019 | 5.4.99.5 | 0.0001 | 662 | 629 |
| **AKG** | 1578 | 2.7.2.8 | 0.0001 | 659 | 691 |
| **ASAD** | 2640 | 1.2.1.11 | 0.0001 | 873 | 913 |
| **SHMT** | 4837 | 2.1.2.1 | 0.0001 | 1246 | 1360 |
| **ASL** | 1570 | 4.3.2.1 | 0.0001 | 923 | 980 |
| **5HMT** | 1476 | 2.1.1.13 | 0.0001 | 741 | 766 |
| **GS** | 1613 | 6.3.1.2 | 0.0001 | 1961 | 2094 |
| **DPD** | 5353 | 4.1.1.20 | 0.0001 | 1560 | 1713 |
| **AAT** | 1577 | 2.6.1.1 | 0.0001 | 5451 | 5228 |

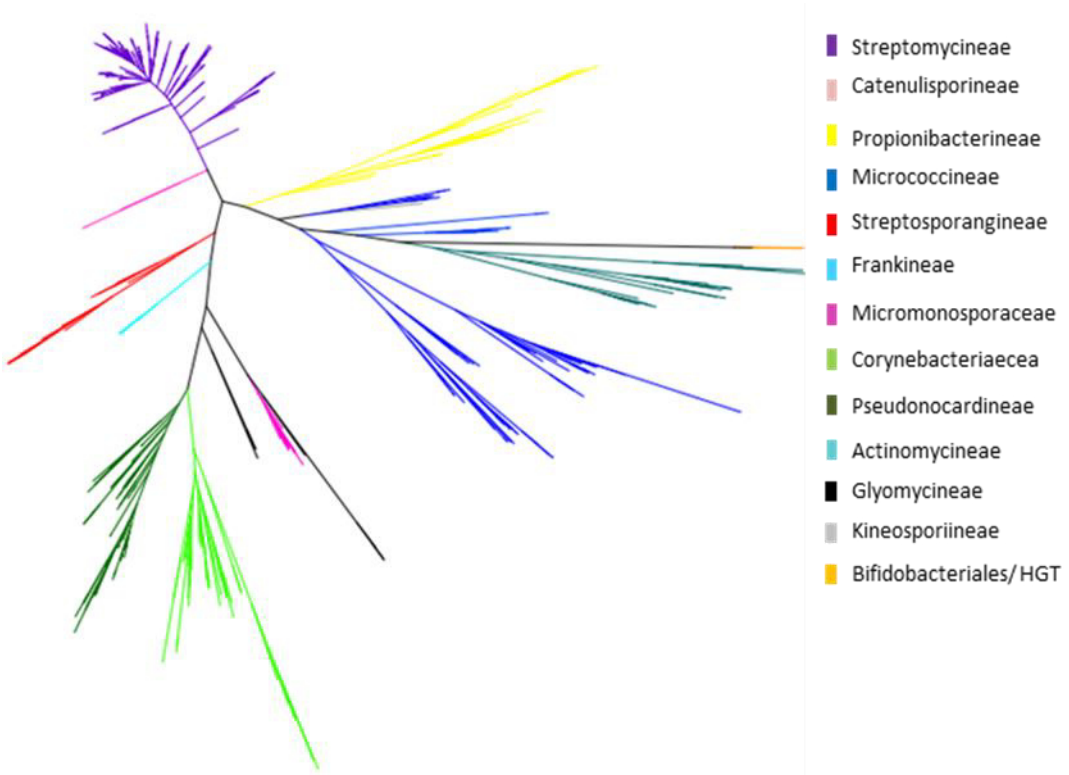



Figure 3.1 RpoB based phylogenetic tree from all actinobacterial species present in the database. The tree can be divided in three distinct phylogenetic branches, one composed of Streptomycineae (purple) and Catenulisporinae (rose), one including Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange) and Micrococcineae (blue) and the third one with Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red). Five species were misplaced on the tree and grouped with other species rather than the rest of their genus: *Streptomyces carneus* (ID: 103200_819) was in *Nocardia* genus cluster (suborder Corynbacterineae), *Streptomyces NRRL F3213* (ID: 104209_3190) showed in the *Amycolatopsis* cluster (suborder Pseudonocardineae), *Rhodococcus rhodni* (ID:104518_2496) as *Kibdelsporangium* (suborder Pseudonocardineae), *Actinospica acidiphila* (ID: 104678_6451) and *Microtetraspora glauca* (ID:104669_4572) both as *Streptomyces* (suborder Streptomycineae).

### 3.2.1 Phylogenetic Analysis of Gene Expansions in Central Carbon Metabolism in *S. coelicolor*

**Pyruvate kinase**

Pyruvate kinase was one of the hits of the expansion analysis with two genes for the genus of *Streptomyces* compared to 1.34 as the overall average across all analysed genera (Table 3-2). The phylogenetic tree of pyruvate kinase (Figure 3.2) showed a high level of congruence with RpoB (Figure 3.1). Interestingly in the Streptomycineae (*Kitasatospora*, *Streptacidiphilus* and *Streptomyces*) family the presence of two genes encoding for pyruvate kinase was the most frequently observed situation. These data indicate that a gene duplication event has occurred within this group. Analysis of 286 *Streptomyces* species showed that 281 species have two copies of *pyk*, three species with one copy, two species with three copies and one species with four copies (Figure 3.2). The species with a single pyruvate kinase are *S. somaliensis , S. sp NRRL F5135* and *S. scrabrisporus,* the first two were in the branch of Pyk2 and the latter was in the Pyk1 branch. *S. olinedensis* had two copies in the Pyk2 branch and one within the Pyk1 cluster. *S. sp AcH505* had one copy of Pyk in each branch of the tree and an additional copy, which was distant phylogenetically from the other copies present in the genome suggesting that this copy was acquired through horizontal gene transfer (HGT). *S. resistomycificus* showed four annotations for pyruvate kinases, one in each Pyk1 and Pyk2 lineages and two additional copies in the phylogenetic branch are from the HGT.

In the Actinobacteria, 302 of the 327 species that were not from the genus *Streptomyces* had a single pyruvate kinase (Figure 3.2). Notable exceptions showing multiple gene annotations for pyruvate kinase with one ancestral copy and a second copy that claded within the HGT branch were: *Longispora albida* , *Micromonospora sp CNB394*, *Rhodococcus rhodochrous* ATCC21198 and *R. rhodnii*, *Corynebacterium glutamicum* and *C. jejeikum, Mycobacterium smegmatis, Kocuria rhizophila* and *Kocuria sp. UCDDTCP, Citriococcus CH26A, Dietzia UCDTHP, Kutzneria albida* and *Frankia* CN3. With the exception of *Micromonospora, Longispora* and *Streptomyces,* the BLAST hits did not agree with the genome annotation in RAST. *R. rhodnii* was one of the species that showed up as *Kibdelsporangium* in the species tree.

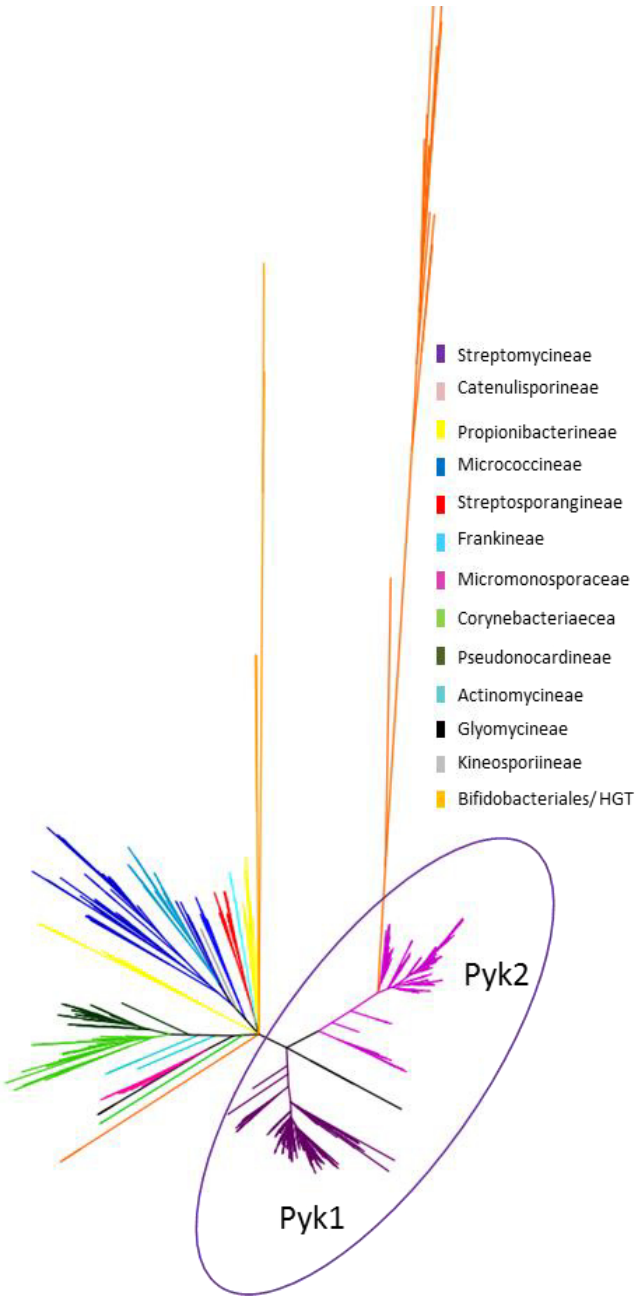



Figure 3.2 Pyruvate kinase phylogenetic tree from all actinobacterial species present in the database. Purple cycle is drawn around the group of Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

**Purifying selection is observed for pyruvate kinases gene**

The ratio of non-synonymous changes to synonymous changes is an indicator as to the evolutionary selection pressure present on a protein. Through comparison of several sequences of orthologous or paralogous genes. Non-synonymous changes are those that will result in the change of the current amino acid sequence of the protein encoded, whereas a synonymous change is a so-called silent mutation, where the change in the DNA sequence does not lead to a change in the amino acid sequence (Hurst, 2002; Rocha *et al.*, 2006). If the dN/dS ratio is below 1, the synonymous changes are more abundant than non-synonymous changes. Indicating purifying or stabilizing selection is present, this means any negative mutations are counter selected and the protein is likely to be maintained. If the dN/dS ratio is above 1, this represents positive selection, where the non-synonymous changes would be more abundant than synonymous, more variability is tolerated in the protein and is likely to be evolving rapidly (Hurst, 2002; Rocha *et al.*, 2006; Clark and Hoskisson, 2011).

Twenty pyruvate kinase sequences from ten *Streptomyces* genomes were chosen for the assessment of selection pressure present in pyruvate kinase. The Nei-Gojobori method was utilised for estimation of the synonymous (dS) and non-synonymous (dN) changes, this method assumes that equal nucleotide frequencies and random substitution apply. A change in the second position of a codon will always result in a non-synonymous change, whereas the first and third position only a fraction, with the above assumption this 5 % for the first and 72 % for the third position. By this the number of synonymous (s) and non-synonymous (n) changes among two homologues sequences can be assessed codon by codon. The fraction of synonymous changes ($f_i$) at the position i of a given codon with i = 1, 2 or 3 (Nei and Gojobori, 1986).

$$s = \sum_{i=1}^{3} f_i \text{ and } n = 3 - dS$$

For a sequence with r codons, the total number of synonymous and non-synonymous changes can be calculated as following:

$$dS = \sum_{j=1}^{r} S_i \text{ and } dN = 3r - dS$$

The standard error was estimated by including a bootstrap of 1000 replicates (Tamura *et al.*, 2013).

The dN, dS and dN/dS were obtained between the two pyk genes per species as well as the overall values for all *pyk1*-like and all *pyk2*-like, additionally to an overall *pyk* value. For comparison the values for *E. coli pykA* and *pykF* sequence were also obtained in order to compare it to the values obtained from *Streptomyces*. The dN/dS ratio of comparing each of the two pyk sequences between themselves on each of the genomes yielded ratios from 0.407 to 0.500, suggesting that pyruvate kinases in *Streptomyces* are under strong purifying selection. Comparing all pyk1-like and all pyk2-like among themselves ratios of 0.161 and 0.216 were observed respectively suggesting even stronger purifying selection. Analysing all twenty sequences the dN/dS ratio was of 0.332. Comparing the *pykA* and *pykF* genes from *E. coli* the dN/dS ration was of 0.631, which was higher than the values obtained for *Streptomyces* (Table 3-4). These high levels of purifying levels are further evidence that the duplication event in *Streptomyces* is ancient.

A phylogenetic tree of the selected pyruvate kinase genes was also constructed, showing a similar topology as the phylogenetic tree of the entire actinobacterial database (Figure 3.2 and Figure 3.3)

Table 3-4 Synonymous and non-synonymous changes in pyruvate kinase gene sequences from 10 *Streptomyces* species and *E. coli* Legend: dS = synonymous changes, dN = nonsynonymous changes

| | Gene # | | | | | | |
|---|---|---|---|---|---|---|---|
| **Strain** | ***pyk1*** | ***pyk2*** | **dS** | **Std. Err** | **dN** | **Std. Err** | **dN/dS** |
| ***S. coelicolor*** | 2014 | 5423 | 0.509 | 0.044 | 0.242 | 0.020 | 0.476 |
| ***S. lividans*** | 2336 | 5692 | 0.503 | 0.043 | 0.241 | 0.020 | 0.479 |
| ***S. avermitilis*** | 6217 | 2825 | 0.572 | 0.051 | 0.260 | 0.021 | 0.455 |
| ***S. hygroscopicus*** | 3488 | 6499 | 0.523 | 0.046 | 0.251 | 0.021 | 0.480 |
| ***S. sviceus*** | 0633 | 1713 | 0.611 | 0.054 | 0.268 | 0.021 | 0.438 |
| ***S. griseus*** | 5516 | 2113 | 0.592 | 0.055 | 0.252 | 0.021 | 0.425 |
| ***S. clavuligerus*** | 1203 | 4329 | 0.641 | 0.057 | 0.260 | 0.022 | 0.407 |
| ***S. tsukubaensis*** | 9939 | 27199 | 0.574 | 0.053 | 0.256 | 0.022 | 0.447 |
| ***S. rimosus*** | 19389 | 25850 | 0.591 | 0.051 | 0.259 | 0.021 | 0.438 |
| ***S. venezuelae*** | 1640 | 5075 | 0.473 | 0.046 | 0.237 | 0.020 | 0.500 |
| **All *pyk1*** | | | 0.504 | 0.023 | 0.081 | 0.007 | 0.161 |
| **All *pyk2*** | | | 0.365 | 0.017 | 0.079 | 0.08 | 0.216 |
| **Overall** | | | 0.512 | 0.011 | 0.17 | 0.018 | 0.332 |
| ***E. coli*** | *pykA* | *pykF* | 1.062 | 0.122 | 0.67 | 0.043 | 0.631 |

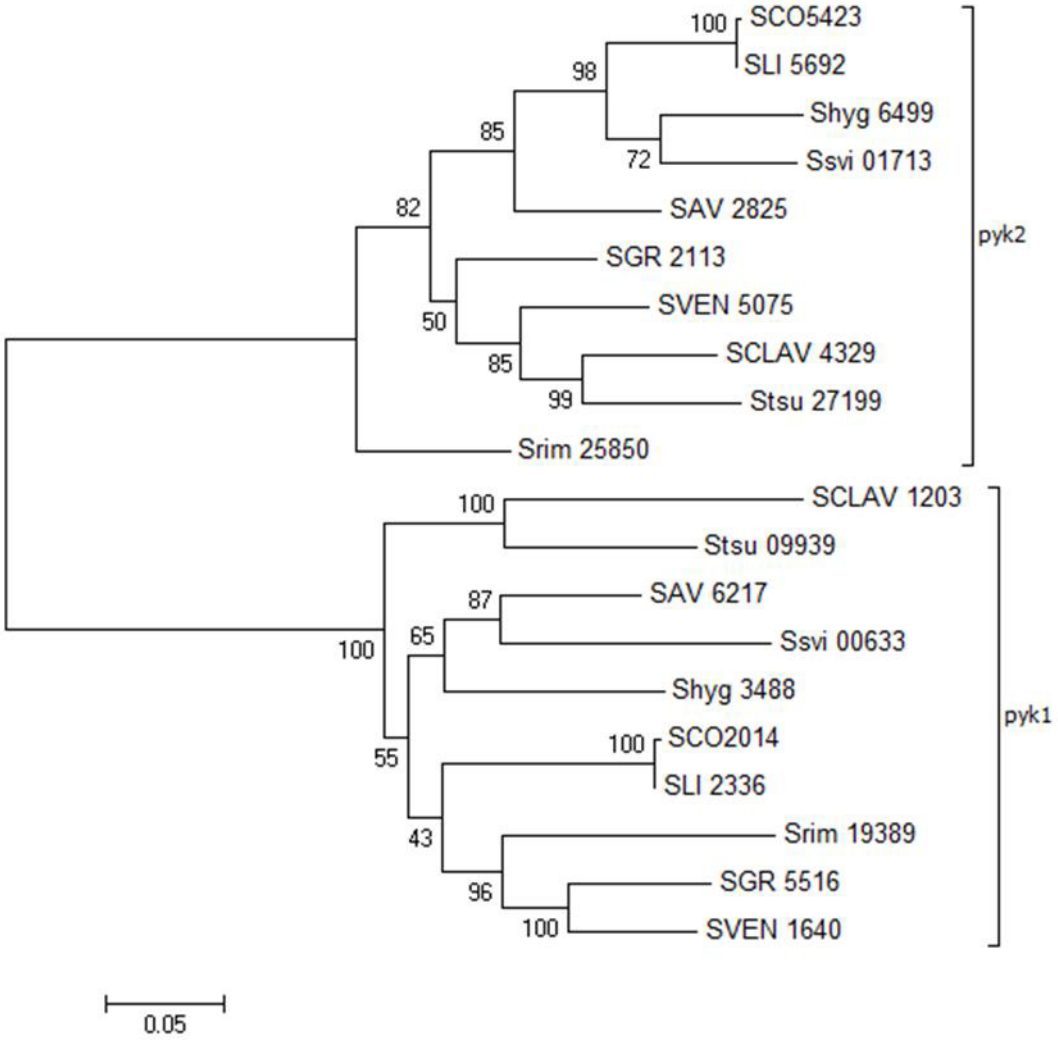



Figure 3.3 Phylogentic tree of pyruvate kinase genes from ten different *Streptomyces* species which were seleceted to study the selection pressure on the gene. Tree was constructed using MEGA6 software with Maximum likelihood algorithm with 1000x bootstrap. Legend: Sco = *S. coelicolor*, Sli = *S. lividans*, Shyg = *S. hygroscopicus*, Ssvi = *S. sviceus*, Sav = *S. avermitilis*, Sgr = *S. griseus*, Sven = *S. venezuleae*, Sclav = *S. clavuligerus*, Stsu= *S. tsukubaensis*, Srim = *S. rimosus*

**Phosphofructokinase**

A further example of a gene expansion event mediated through gene duplication is the family of genes encoding phosphofructokinase. In the *S. coelicolor* genome, three copies of PFK are found, compared to an average plus standard deviation of 1.39 in the overall database and 1.76 copies in the genus of *Streptomyces* (Table 3-2). The topology of the tree suggested that PfkA2 and PfkA3 are duplicates. However PfkA1 was highly divergent of the two others, being located at the other end of the tree. Overall the tree did not reflect the species tree. Each of the clades of PfkA1, PfkA2 and PfkA3 also showed another branch emerging, which were mainly Actinomycineae, some Propionibacterieae, Micromonosporeae (only in PfkA2) and Micrococcineae (only in PfkA1). Most Corynebacterieae were located at the centre of the tree, wherease Pseudonocardiaeae were closer to PfkA1 than to PfkA3 or PfkA2 (Figure 3.4).

**Pyruvate Phosphate Dikinase (PpdK)**

Overall the mean number of occurences plus standard deviation was of 1.36 for PpdK. In *Streptomyces* this was 1.37 and two copies in the case of *S. coelicolor* A3(2) (Table 3-2). The phylogenetic tree reveals a highly divergent tree, which does not follow the species tree. Especially Micrococcineae, Propionibacterineae and Corynebacteriaceae show highly divergent PpdK copies across the different genera. The two genes of pyruvate phosphate dikinases have not arisen by gene duplication as is clearly visible from the topology of the phylogenetic tree (Figure 3.5).

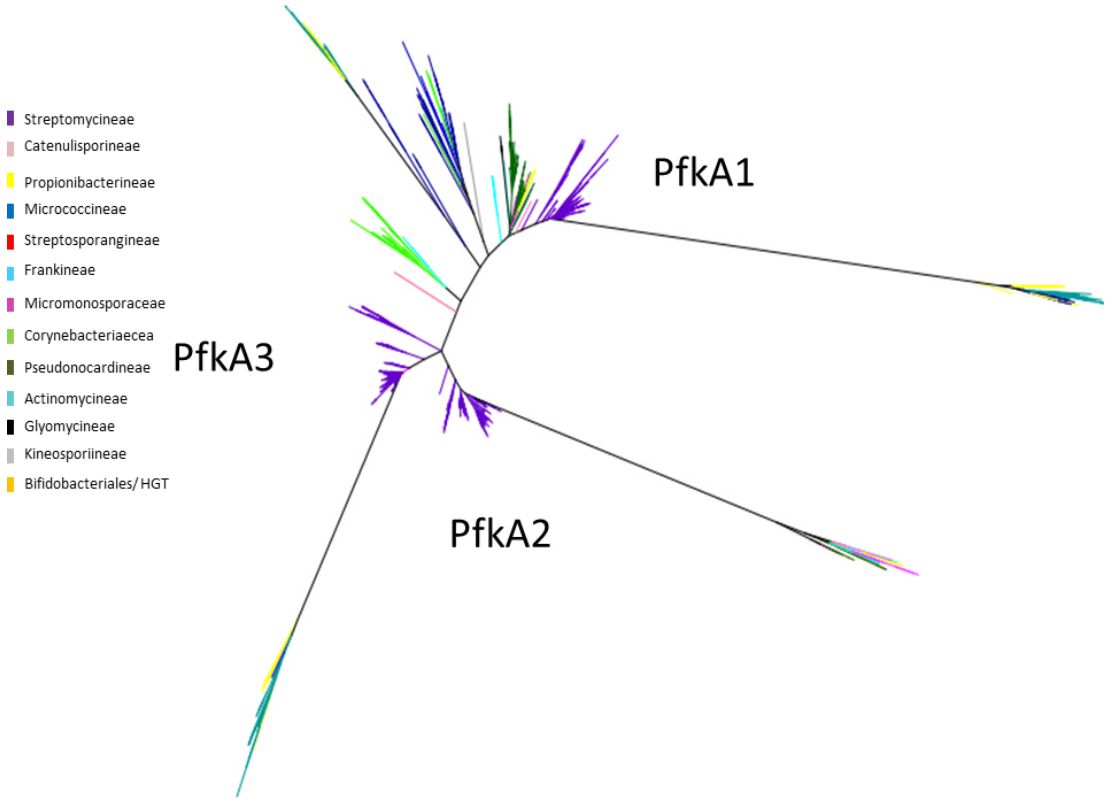



Figure 3.4 Phosphofructokinase phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

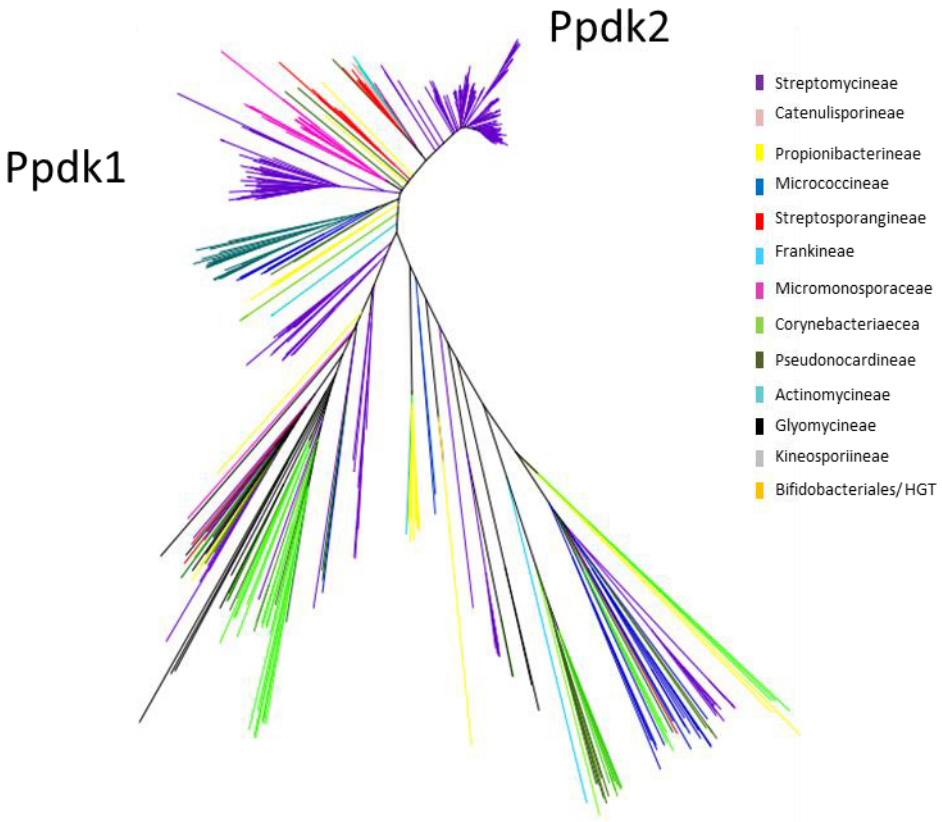



Figure 3.5 Pyruvate phosphate dikinase phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

**Pyruvate dehydrogenase complex (PDHC)**

For the PDH complex, only $PDHC_{E1}$ was analysed since it is the rate limiting step in the reaction closely working together with $PDHC_{E2}$. $PDHC_{E3}$ is often shared with other enzyme complexes such as succinyl dehydrogenase or keto acid dehydrogenases (de Kok *et al.*, 1998). $PDHC_{E1\alpha}$ and $PDHC_{E1\beta}$ yielded very poor alignments across the entire database and were thus not included in the phylogenetic analysis. On average there were 2.04 genes annotated for $PDHC_{E1}$ across the entire database compared to 2.18 in the genus *Streptomyces* and three genes can be found in *S. coelicolor* A3(2) (Table 3-2). The phylogenetic tree revealed that SCO2183 and SCO7124 are much closer related to each other than either of them to SCO2371. There are three big clades to be distinguished on the tree, one branch containing SCO2371 together with most other species, on a second branch there were SCO2183 and SCO7124 and a third branch with another Streptomycineae cluster as well as some Corynebactericeae (Figure 3.6).

Due to the presence of both Gram-positive and Gram-negative bacteria like sequences in *Streptomyces* genome an alignment of the three $PDHC_{E1}$ with the *E. coli* $PDHC_{E1}$ was carried out. This was performed in order to test the hypothesis that the E1 subunit of PDHC (SCO7124) is more closely related to Gram-negative bacteria. A motif 'E-R—HQMKL' which has been reported in the *E. coli* E1 for a loop trapping the substrate into the active centre has been searched in the *Streptomyces* sequences and has been found to be conserved (Figure 3-7, Kale *et al.*, 2007). This motif was not detected in α or β chains of $PDHC_{E1}$. The general perception that all Gram- positive bacteria have the heteromeric α/β E1 and Gram negative bacteria have the homomeric E1 seems to be wrong as it has been found that *C. glutamicum* for instance has only the Gram negative form (Schreiner *et al.*, 2005) and in *Streptomyces* there are actually several copies of each present. The situation thus seems to be more complex.

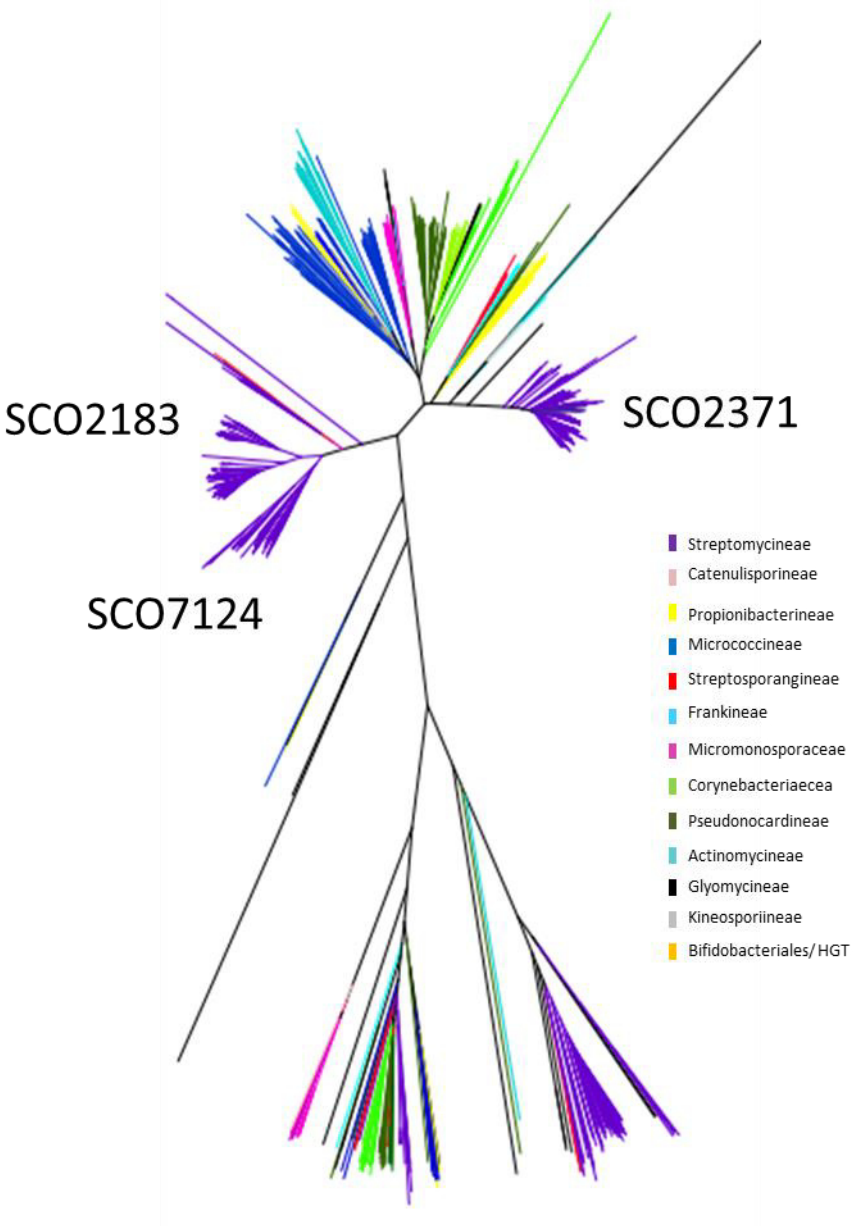



Figure 3.6 Pyruvate dehydrogenase complex $PDHC_{E1}$ phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

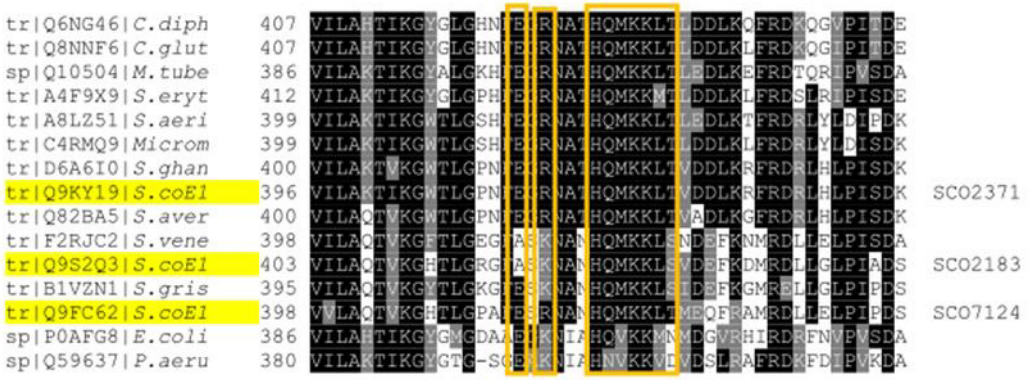



Figure 3.7 Part of $PDHC_{E1}$ alignment using ClustalW and Boxshade looking for the motif "E-R—HQMKL" reported in $PDHC_{E1}$ from *E. coli* in predicted $PDHC_{E1}$ from *Streptomyces* (Kale *et al.*, 2007)

**Phylogenetic analysis for the other enzymes of the metabolic node: Malic enzyme, PEPCk, PEPCx and PCx**

The average number plus standard deviation of annotations for malic enzymes in the database was of 1.12, in *Streptomyces* this was 1.5 and two annotations are found in the *S. coelicolor* A3(2) genome (Table 3-2). The two malic enzymes were on distinct clades of the tree and furthest apart. It is known from the literature, that they show different cofactor dependencies - SCO2951 is $NAD^+$ dependent, whereas SCO5261 is $NADP^+$- dependent (Rodríguez *et al.*, 2012). Pseudonocardineae, Micrococcineae, Micromonosporaceae and Streptosporangineae showed clusters in two separate clades similar to Streptomycineae. Actinomycineae showed as a single clade, closer to the $NAD^+$ dependent clade. Corynebacteriaceae showed one clade only, closely located to one of the clades from Pseudonocardineae. Furthermore there was one outbranch, which contained a clade from Propionibacterineae. However there were also two clades from this group similarly occurring to Streptomycineae. This outbranch seemed equally distant from the $NAD^+$-dependent clade as from the $NADP^+$-dependent clade (Figure 3.8).

PEPCk (SCO4979) and PEPCx (SCO3127) both have only a single copy in the *S. coelicolor* genome and did not come up as an expansion in the analysis, however as they are members of the metabolic node of interest, they were included in the analysis. PEPCk was present in almost the species (608 hits for 614 species, see Table 3-3) and was very similar to the species tree with the exception for Micrococcineae, Actinomycineae and Propinionibacterineae which each showed two distinct clades (Figure 3.9). PEPCx was present in most but not all, with overall 438 hits for 614 species (Table 3-3). It was less prominent in Pseudonocardineae and Micrococcineae than in other genera. The tree reflected the species tree, with the exception of Propionibacterineae which was distributed over several clades (Figure 3.10).

The gene encoding Pyc (SCO0546) was absent in many species. There were 227 hits for 614 species present in the database (Table 3-3) and there is a single copy present in *S. coelicolor* A3(2), it is also present in Pseudonocardineae, Corynebactericeae, Propionibacterineae and Streptosporangineae. Apart from Pseudonocardineae the other families were all in a single clade. There were a few outliers from Streptomycineae, which were present in a second clade (Figure 3.11).

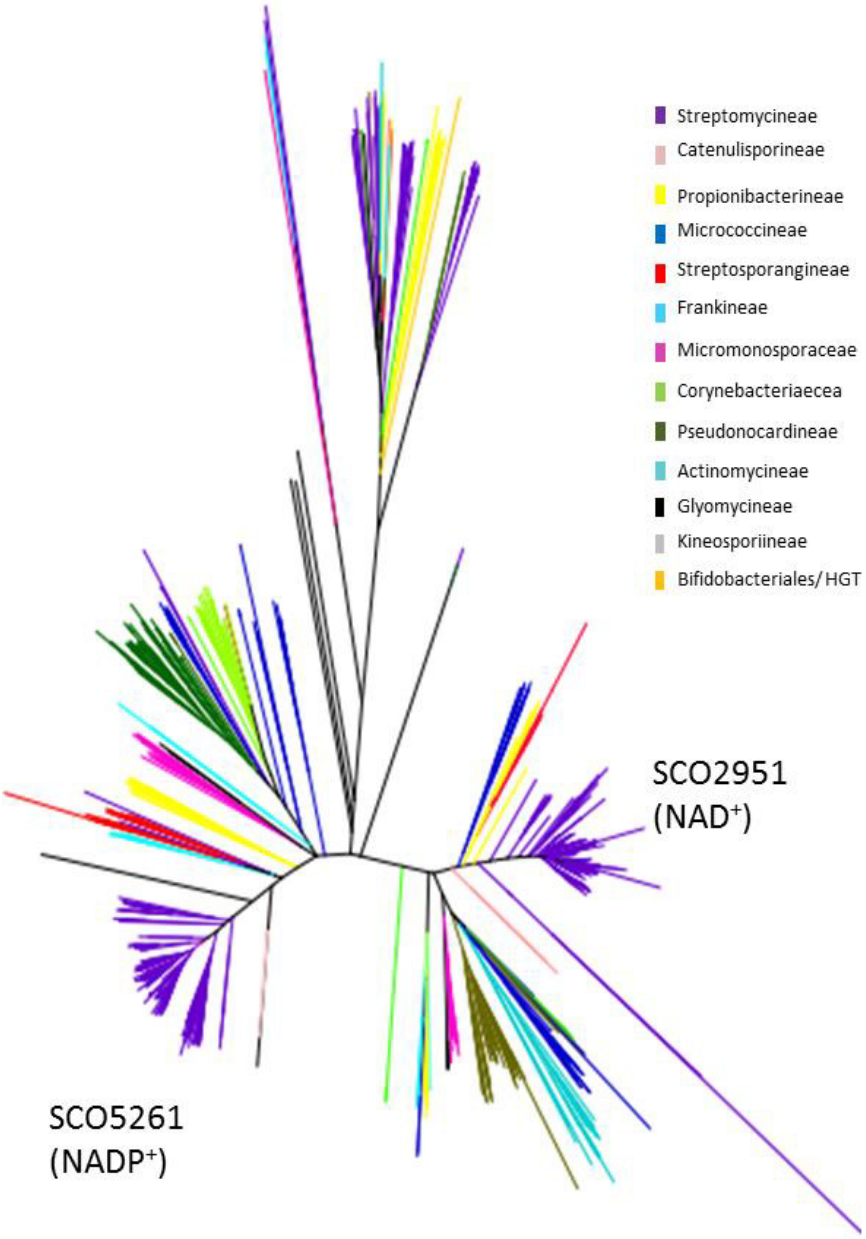



Figure 3.8 Malic enzyme phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

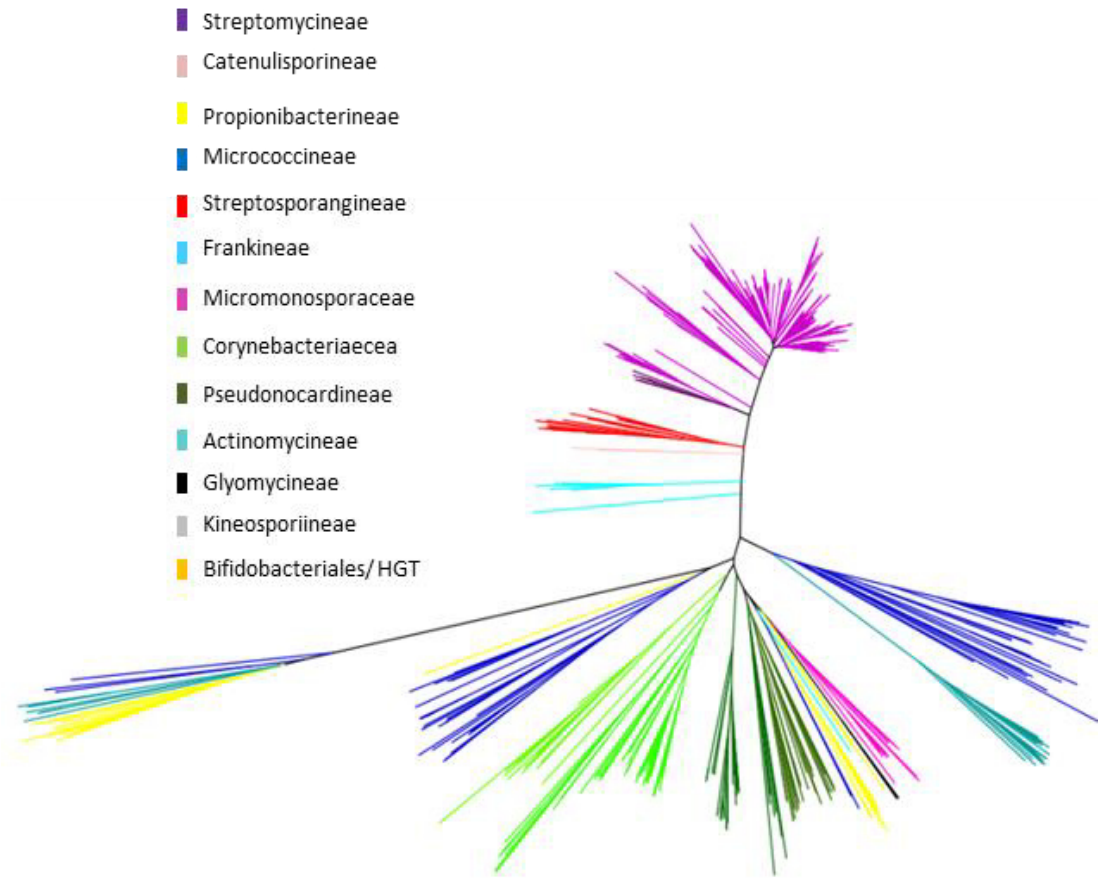



Figure 3.9 PEP carboxykinase phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

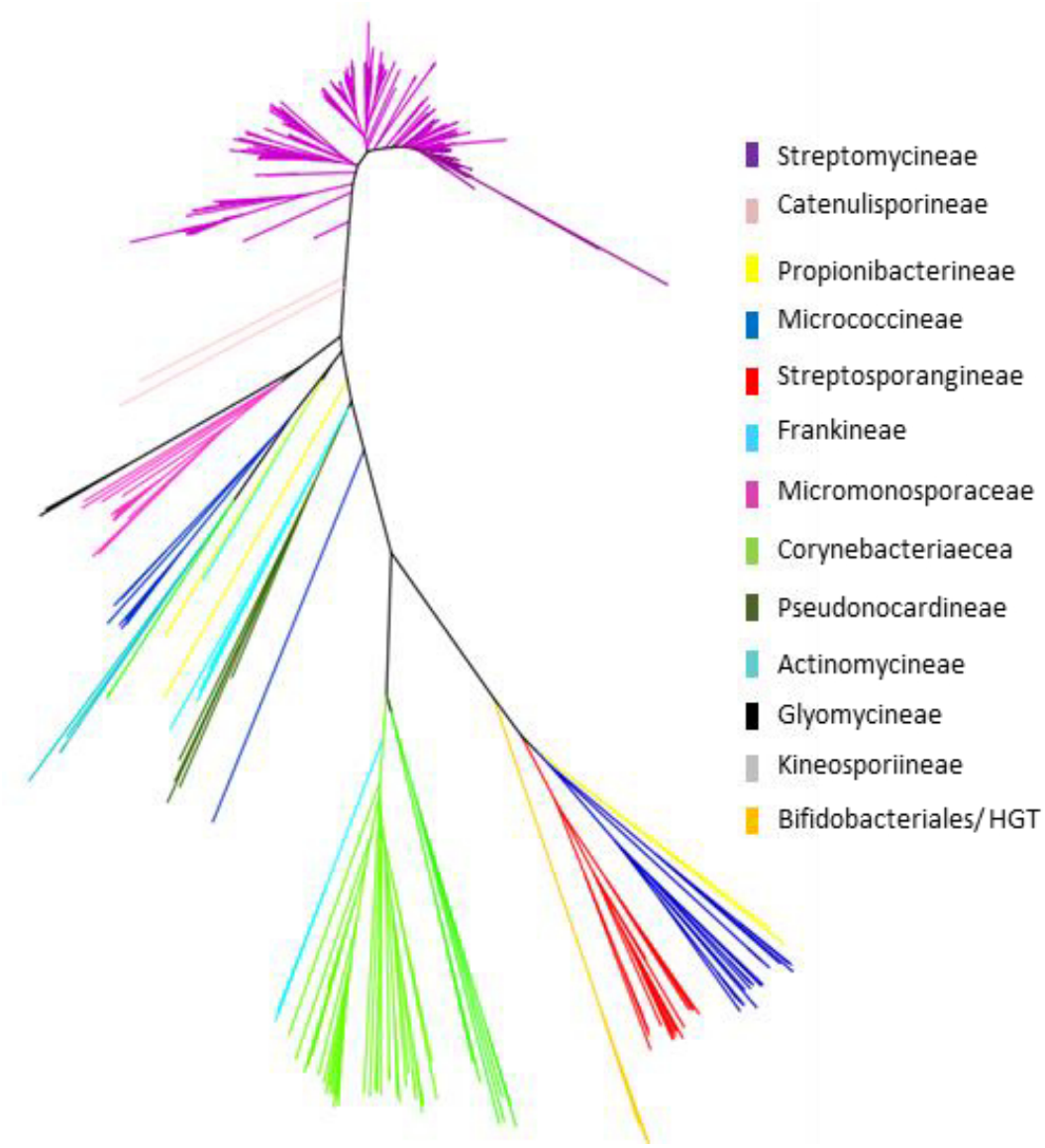



Figure 3.10 PEP carboxylase phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

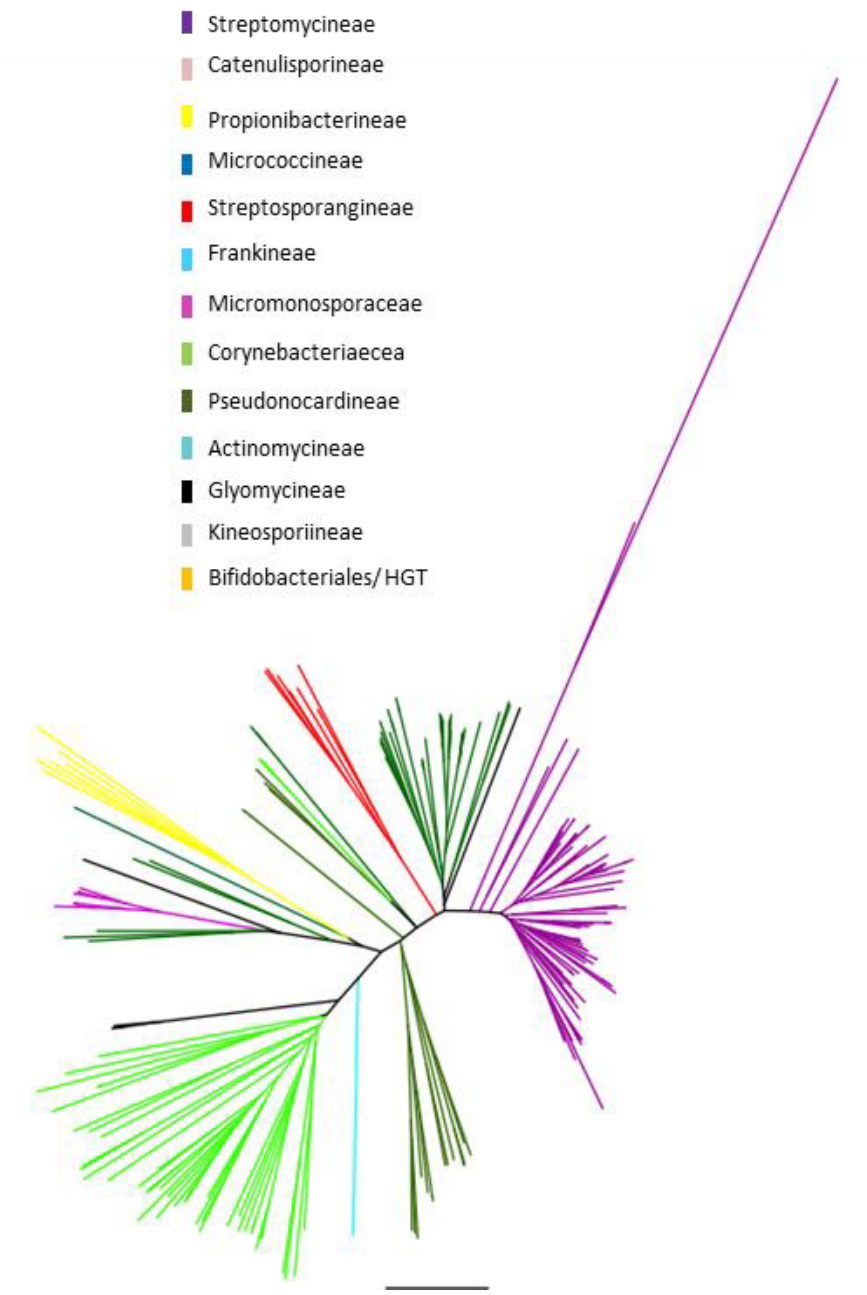



Figure 3.11 Pyruvate carboxylase (Pyc) phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

### 3.2.2 METABOLIC EXPANSIONS IN AMINO ACID METABOLISM OF *STREPTOMYCES*

Amino acid carbon skeletons are also often used as building blocks for specialised metabolites (Rokem *et al.*, 2007; Stirrett *et al.*, 2009). Arginine for example is one of the precursor for clavulanic acid produced by *S. clavuligerus* (Bascarán *et al.*, 1989; Ives and Bushell, 1997). The specialised metabolite teicoplanin uses tyrosine as precursor and in general glycopeptide antibiotics use aromatic amino acids as precursors (Gunnarsson *et al.*, 2004).

Thus it was investigated if there are metabolic expansions present for the amino acid metabolism as well. From the expansion analysis it was found that at least one expansion was present in amino acid metabolism per precursor. There were four expansions detected for amino acids derived from oxaloacetate, which were asparate-semialdehyde dehydrogenase (ASAD), diaminopilimelate decarboxylase (DPD) and aspartate amino transferease (AAT) and methionine synthase (5HMT).

ASAD occured with a mean number of genes of 1.92 in *Streptomyces* compared to 1.57 plus standard deviation across the entire database. In *S. coelicolor* A3(2) two genes can be found coding for ASAD (Table 3-3). The two genes were in different clades and on opposite poles of the tree. A similar phenomenon could be observed for Pseudonocardineae and Micromonosporaceae. Propionibacterineae showed two clusters, but were closer together. All hits from Actinomycineae and Corynebacteriaceae showed one clade only respectively (Figure 3.12).

No phylogenetic tree was constructed for DPD and AAT as the alignments were too poor and the sequences too diverse.

Methionine synthase (5HMT) showed an overall mean number of occurences plus standard deviation of 1.7. For *Streptomyces* it was also 1.7 and two genes are encoded in the *S. coelicolor* A3(2) genome (Table 3-3). The phylogenetic tree did not reflect the species tree. Propionibacterineae were found spread in different clades of the tree. Streptomycineae genes were found in two different clades. Overall the genes were divergent even among closely related species (Figure 3.13).

3-Phosphoglycerate (3PG) was as precursor for the amino acids serine, glycine and cysteine. One expansion was detected for the enzyme serine hydroxymethyltransferase (SHMT). The overall average with standard deviation was of 2.37 and 2.94 for *Streptomyces*. *S. coelicolor*

A3(2) has three genes coding for SHMT in its genome (Table 3-3). The phylogenetic tree showed three different clades for each of the copies from *S. coelicolor* A3(2). The tree did not mimic the species tree, therefore indicating that the gene had a different evolutionary history compared to the speciation. The most populated clade was the one containing SCO5364. The clade containing SCO5470 was dominated by Streptomycineae and SCO4878 was in the smallest clade with enzymes from Propionibacterineae, Frankineae and Micromonosporaceae (Figure 3-14).

α-ketoglutarate is the precursor for glutamate from which proline, glutamine, histidine and arginine are derived. The enzyme acetylglutamate kinase (AGK) was found to be expanded with 1.31 average with standard deviation in the overall database and in the genus of *Streptomyces* where the mean number of genes with this role was of 1.31. However only one copy is annotated for this enzyme in the *S. coelicolor* A3(2) genome (Table 3-3). The tree was similar to the species tree with a few exceptions. Actinomycineae and Propionibacterineae were present in multiple clades. Furthermore a number of Micrococcineae were present in the clade with Micromonosporaceae and Frankineae (Figure 3-15).

Ariginiosuccinate lyase (ASL) is involved in the metabolism of alanine, leucine and valine, which are derived from pyruvate. In the database it was expanded with a mean occurrence of the enzyme plus standard deviation of 1.85. In *Streptomyces* this was 1.97 and in *S. coelicolor* A3(2) there was only one annotation (Table 3-3). The phylogenetic tree mimicked in parts the species tree. However three clades were composed of hits from several different families (Streptomycineae, Micromonosporaceae, Pseudonocardineae, Frankineae, as well as some Micrococcineae and Propionibacterineae). Streptomycineae and Catenulisporineae were in one clade, as were Propionibacterineae, Bifidobacteriales, Actinomycineae, some Streptosporangineae and Micrococcineae. Corynebacteriaceaea and Pseudonocardineae formed another distinct clade, as well as Glycomycineae, Frankineae, some Streptosporangineae and Micromonosporaceae (Figure 3.16).

Furthermore chorismate mutase was expanded, which is part of the shikimate pathway. *S. coelicolor* A3(2) had two genes for this role, overall the database had on average 1.86 genes with standard deviation for this function while in *Streptomyces* the average was 2.66 (Table 3-3). The two chorismate mutases in *S. coelicolor* A3(2) were in two distant clades on either pole of the phylogenetic tree, which did not follow the species tree. It showed some similarity to the malic enzyme tree (Figure 3.17).

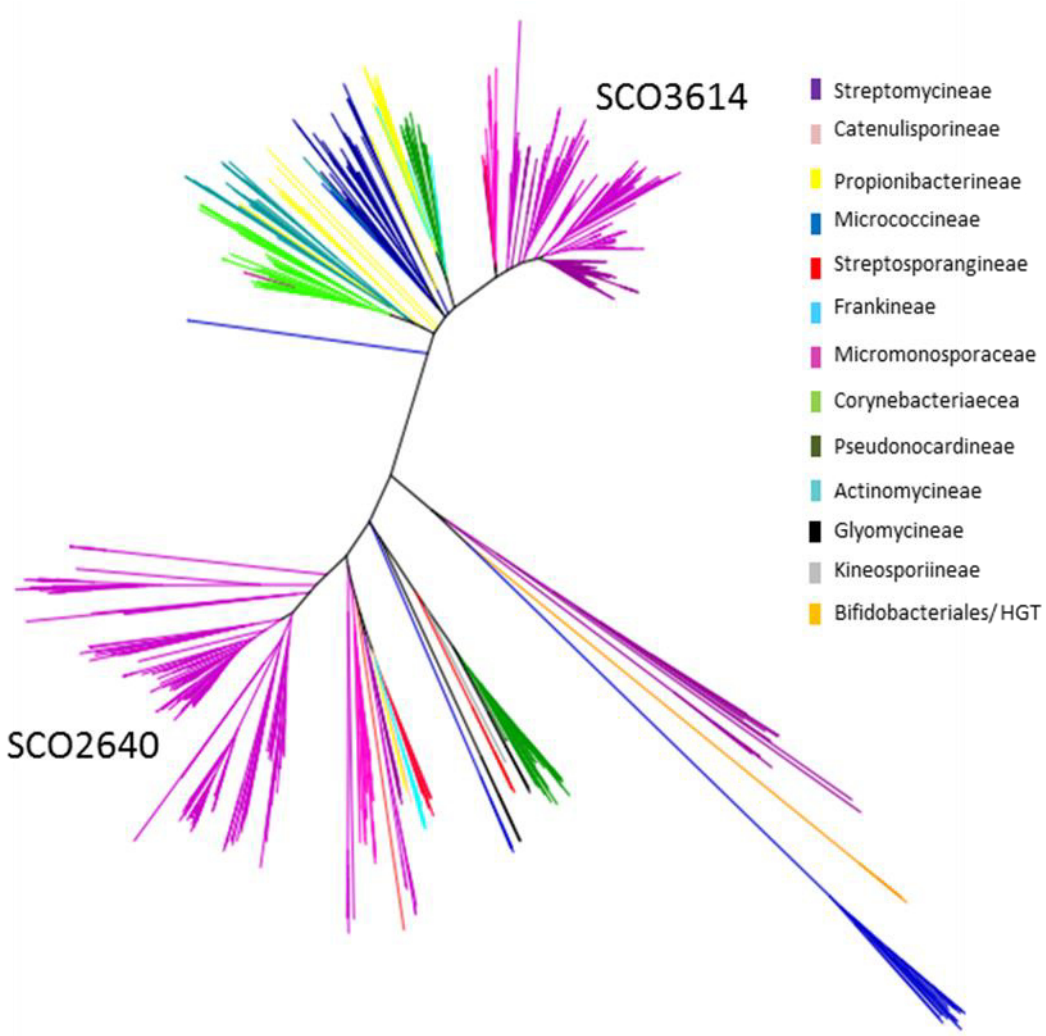



Figure 3.12 Aspartate-semialdehyde dehydrogenase (ASAD) phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

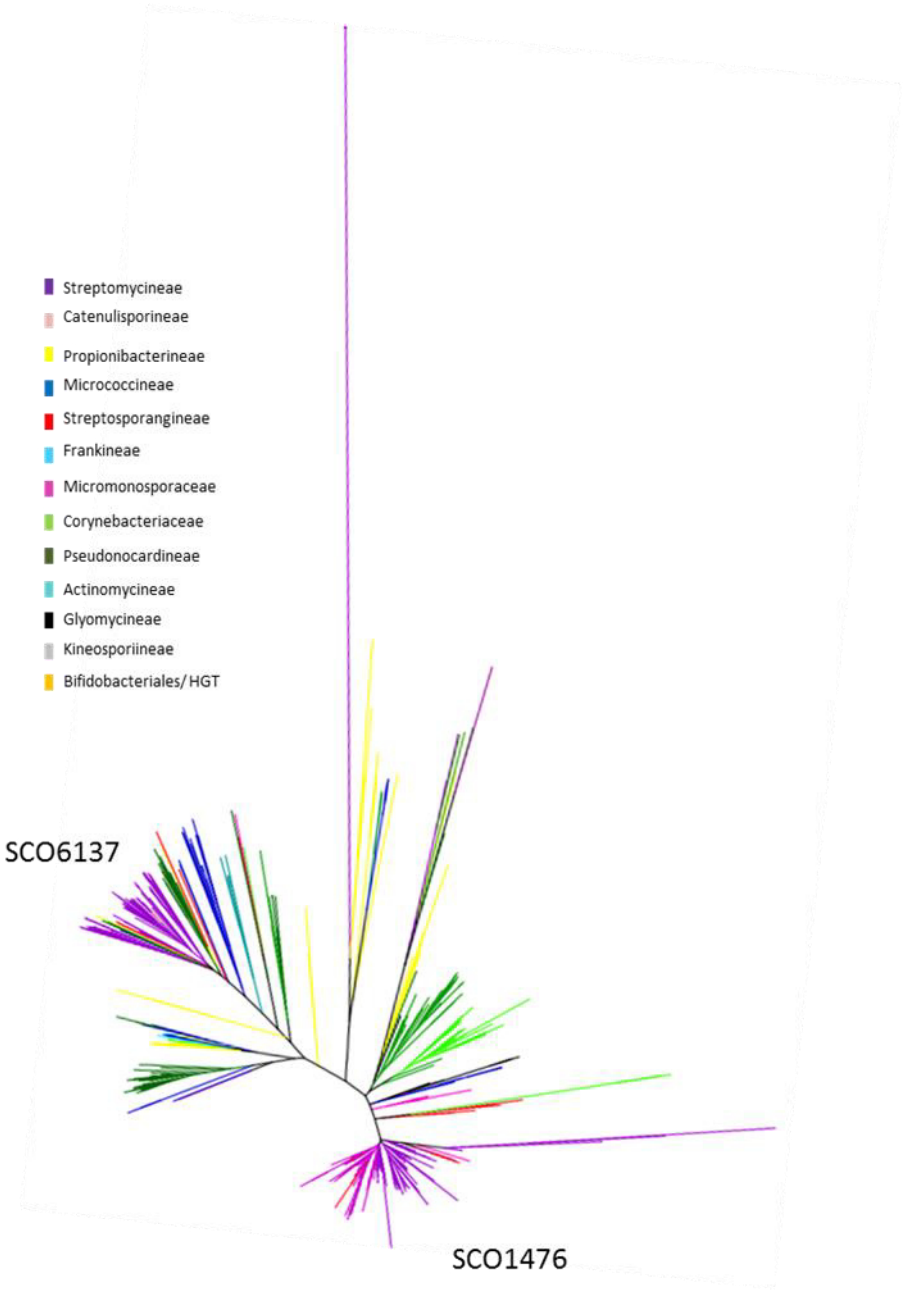



Figure 3.13 Methionine synthase (5HMT) phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

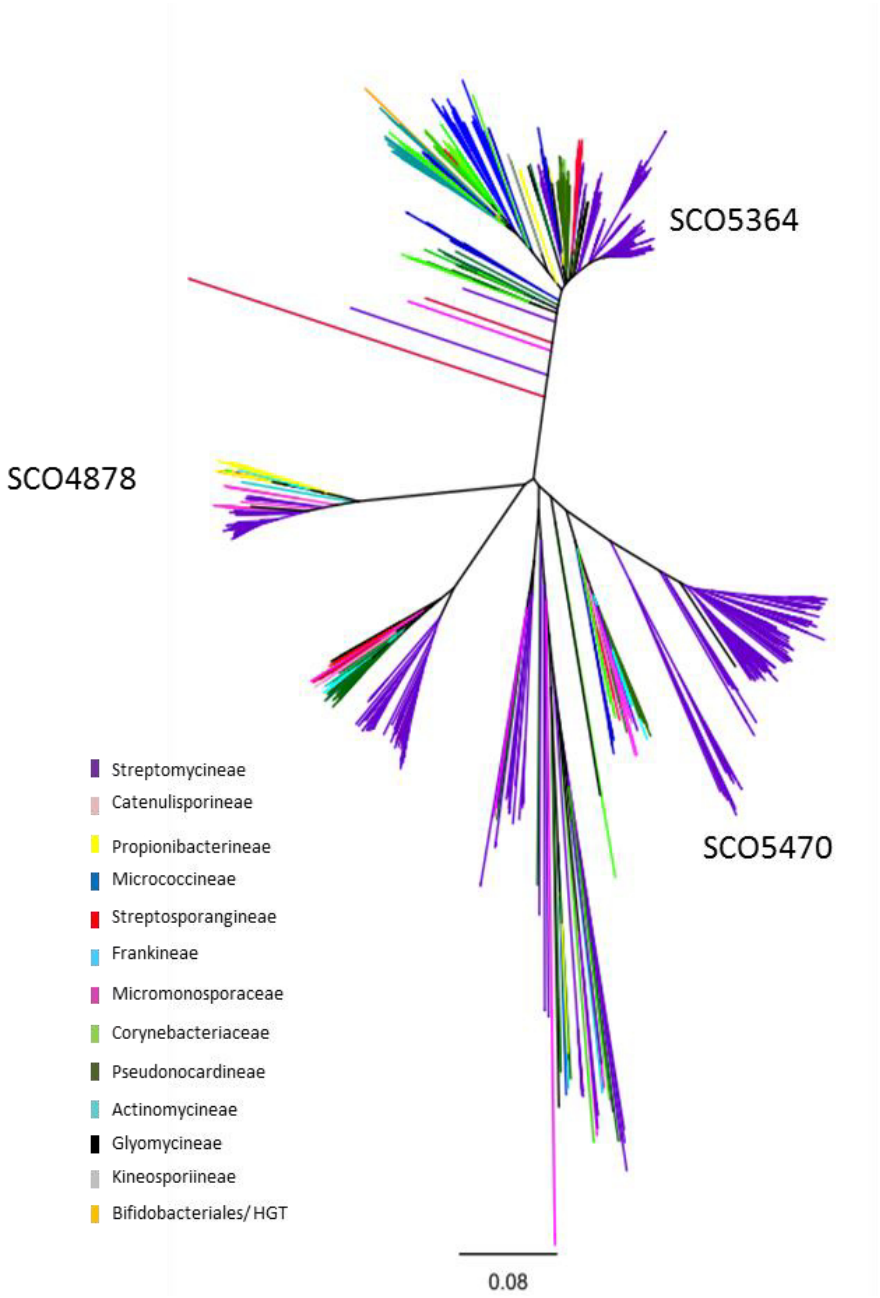



Figure 3.14 Serine hydroxymethyltransferase (SHMT) phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

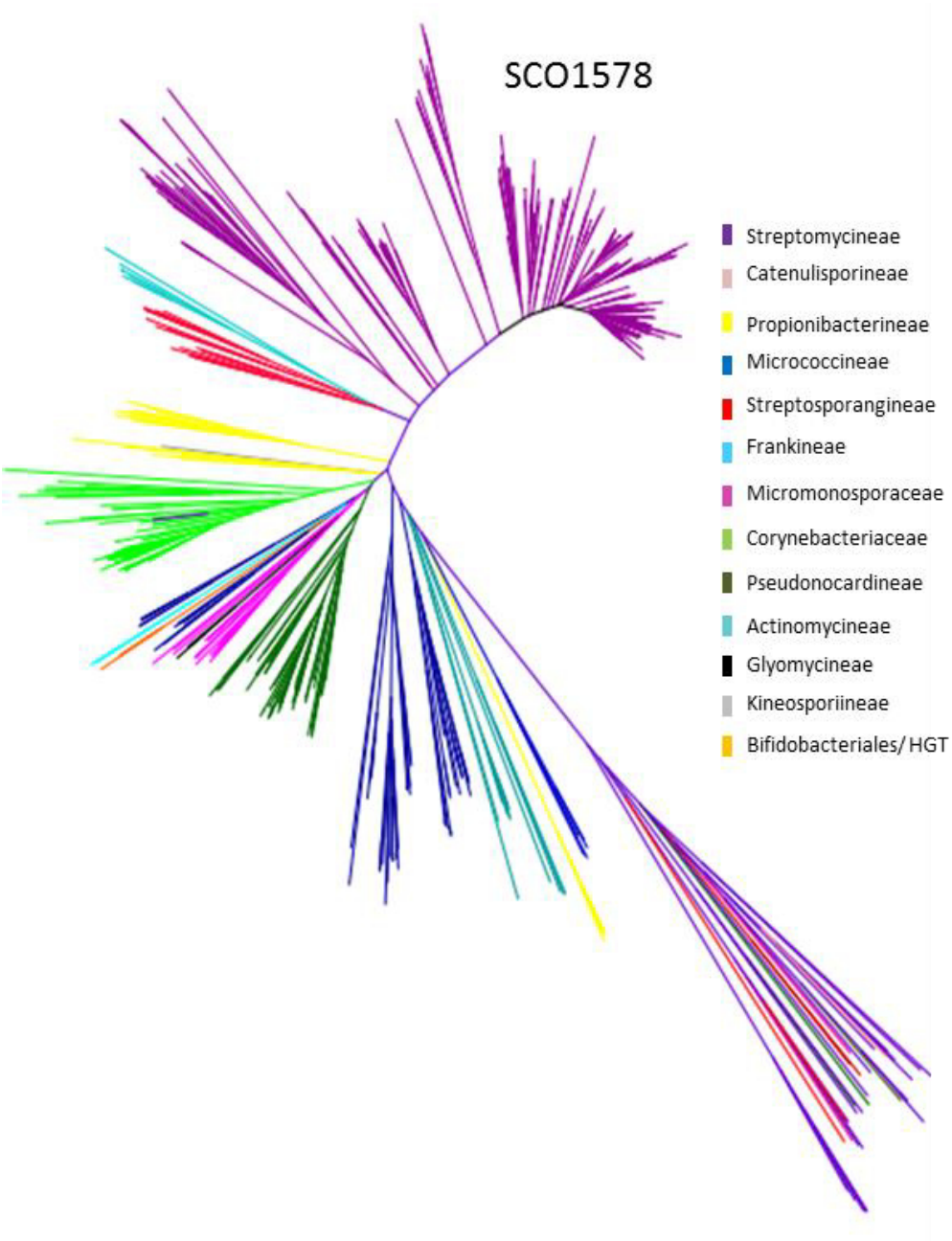



Figure 3.15 Acetylglutamate kinase (AGK) phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

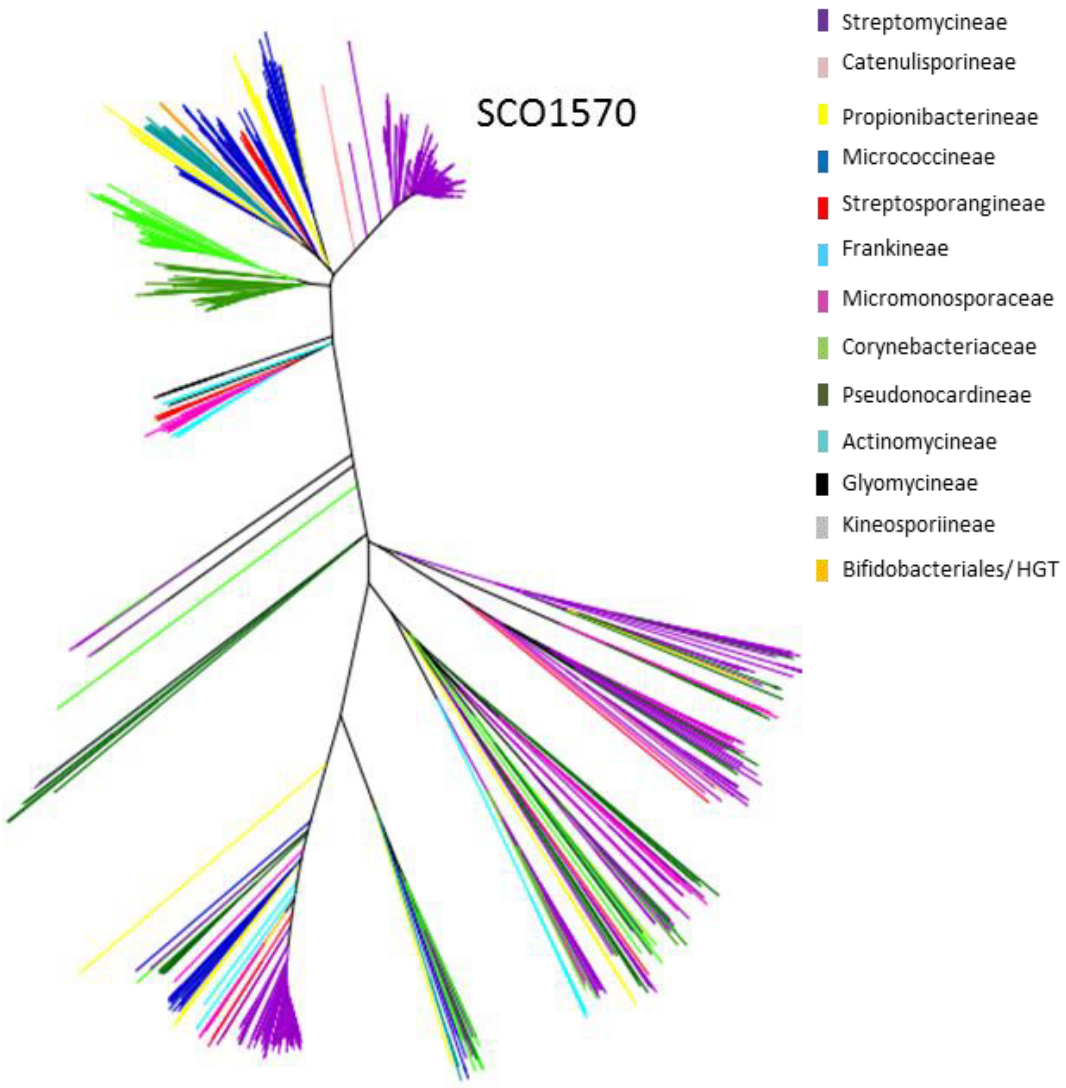



Figure 3.16 Arginiosuccinate lyase (ASL) phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

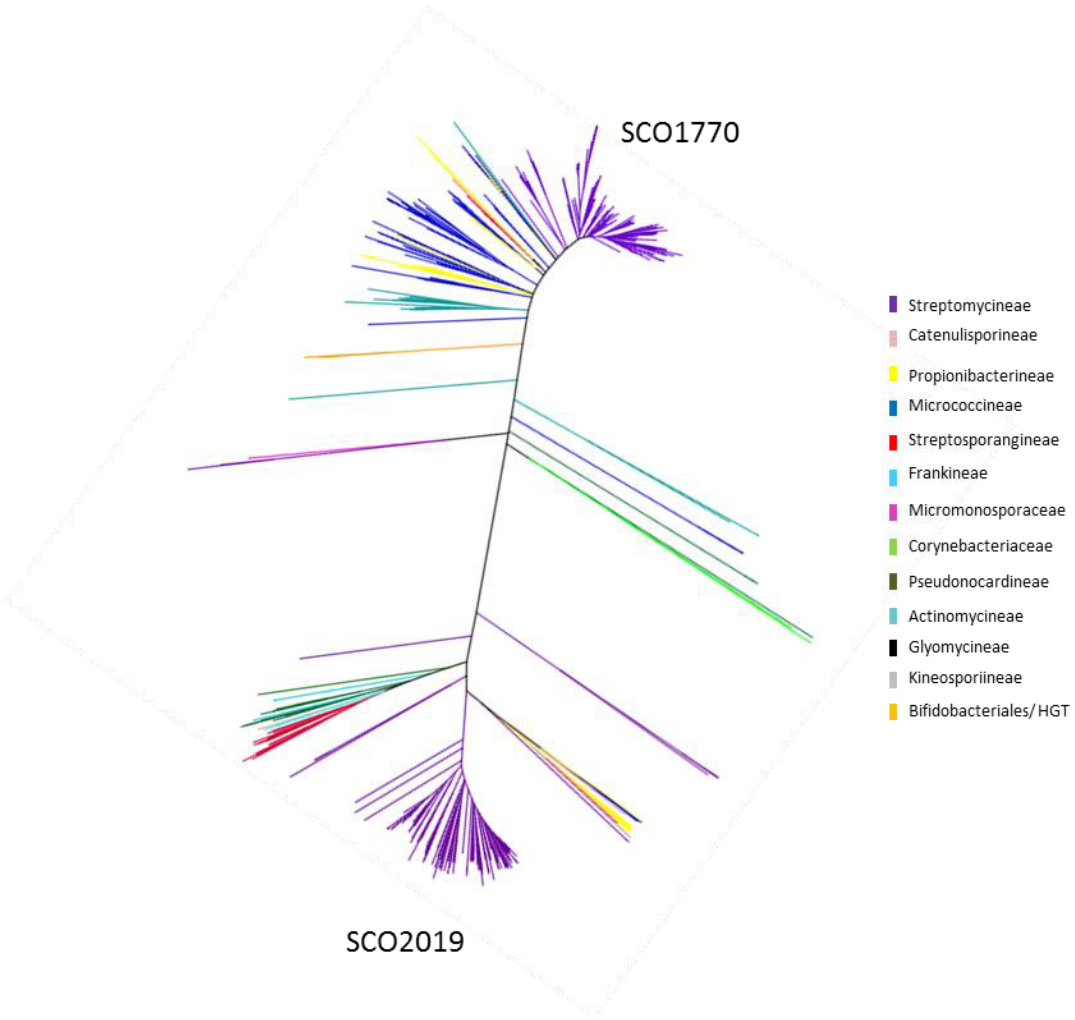



Figure 3.17 Chorismate mutase phylogenetic tree from all actinobacterial species present in the database. Legend: Streptomycineae (purple), Catenulisporinae (rose), Propionibacterineae (yellow), Actinomycineae (turquoise), Bifidobacteriales (orange), Micrococcineae (blue), Micromonosporineae (pink), Glycomycineae (black), Corynebacterineae (light green), Pseudonocardineae (dark green), Frankineae (light blue) and Streptosporangineae (red).

## 3.3 Gene Context of the Enzymes of the PEP-PYR-OAA-Node

The context of genes which are members of the PEP-PYR-OAA node were analysed in order to look for eventual recruitments when several copies were present. Additionally, it is known that adjacent genes are often functionally related in bacterial genomes (Ciria *et al.*, 2004).

Both *ppc* coding for PEPCx and *ppck* coding for PEPCk are standalone genes, not organised with the context of an operon and the surrounding genes are functionally unrelated (Figure 3.18).

The gene SCO2951 encoding a $NAD^+$-dependent malic enzyme is not in an operon and its direct neighbours encode proteins involved in DNA organisation, with HU DNA-binding protein which is responsible for restraining DNA supercoiling and condensing the genome. On the other site a gene encoding helicase is present, which is responsible for DNA denaturation to make it accessible for replication and transcription. SCO5621, which encodes a $NADP^+$-dependent malic enzyme, is not within an operon. The next gene downstream encodes a dehydrogenase, but many of its surrounding genes encode hypothetical proteins (Figure 3.18).

The gene SCO0546 (*pyc*) encoding pyruvate carboxylase (PCx) is also not present within an operon. However there is an α-galactosidase present, as well as three genes encoding acyltransferases and two acyl carrier proteins (Figure 3.18).

The two genes encoding pyruvate oxidase (or pyruvate dehydrogenase) are SCO7412 and SCO6155. The gene upstream of SCO6155 has an unknown annotation and the gene downstream is annotated as a regulatory protein. The remaining genes in the close vicinity are not related to carbohydrate metabolism. SCO7412 has a LacI family regulator gene upstream and an oxidoreducatase downstream. Furthermore a β-galactosidase encoding gene and a cytochrome P450-family protein gene are present close by (Figure 3.18). The direct organisation around SCO6155 and SCO7412 with one gene up- and one downstream look similar. However since the sequences of these enzymes were too diverse and the alignment was too poor for the construction of a phylogenetic tree, it remains unknown how these genes have evolved. The alignment of these two proteins showed only 30.5 % of identity.

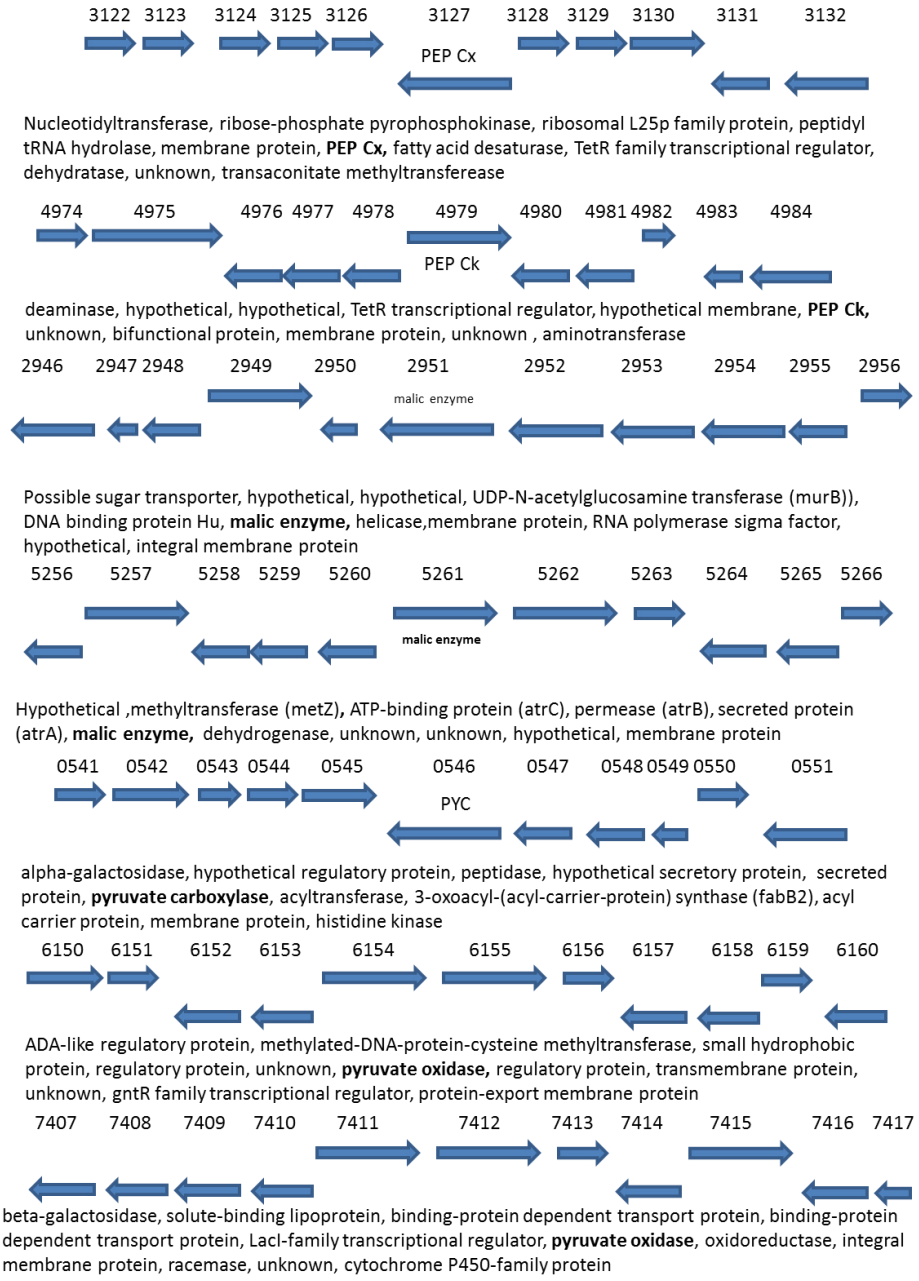



Figure 3.18 Gene context analysis of members of the PEP-PYR-OAA Node in *S. coelicolor* genome for SCO3127 (PEPCx), SCO4979 (PEPCk), SCO2951 (me-NAD), SCO5261 (me-NADP), SCO0546 (PCx), SCO6155 (Odx), SCO7412 (Odx)

The genes encoding PpdK1 (SCO0208) and PpdK2 (SCO2494) show a dissimilar gene context. SCO0208 is adjacent to some sensors and regulatory genes, as well as pseudogenes and a few unknowns. SCO2494 has a *tetR* family regulator gene, a dehydrogenase, oxidoreductase as well as three membrane proteins (one which is a hypothetical efflux protein) and a possible transport ATPase in its vicinity (Figure 3.19).

The pyruvate kinases Pyk1 (SCO2014) and Pyk2 (SCO5423) are both not present in an operon. Adjacent to SCO2014 are four branched chain amino acid transport genes, a two component response regulatory gene and genes encoding a nucleotidase, a monooxygenase, an amino acid decarboxylase and one aminopeptidase as well as a gene encoding a chorismate mutase (SCO2019). SCO5423 is adjacent to a transcriptional regulator, a thioredoxin, a chlosterol esterase, one hypothetical, an acetate kinase gene, phosphate acetyltransferase, phosphofructokinase (*pfkA2*, SCO5426), an unknown and an integral membrane protein. The *pfkA2* and *pfkA3* are an example for gene duplication as are the pyruvate kinases (Figure 3.19).

Some of the pyruvate dehydrogenase complex genes are in close vicinity, such as SCO2180 (dihydrolipimade dehydrogenase encoding gene) and SCO2183 ($PDHC_{E1}$). There is also a gntR-family regulator upstream (SCO2182) and SCO2181 encodes a dihydrolipoamide succinyltransferase. Another regulatory gene is present downstream, some unknowns, a peptidase and one aminopeptidase and a α-polygalactosaminidase. SCO2371, which is another gene encoding $PDHC_{E1}$, is surrounded by a number of hypotheticals, an integral membrane protein, a thiol-specific antioxidant protein, a small hydrophobic protein, a tetracenomycin C efflux protein, a transcriptional regulator and a secreted protein. The gene is not within an operon (Figure 3.19).

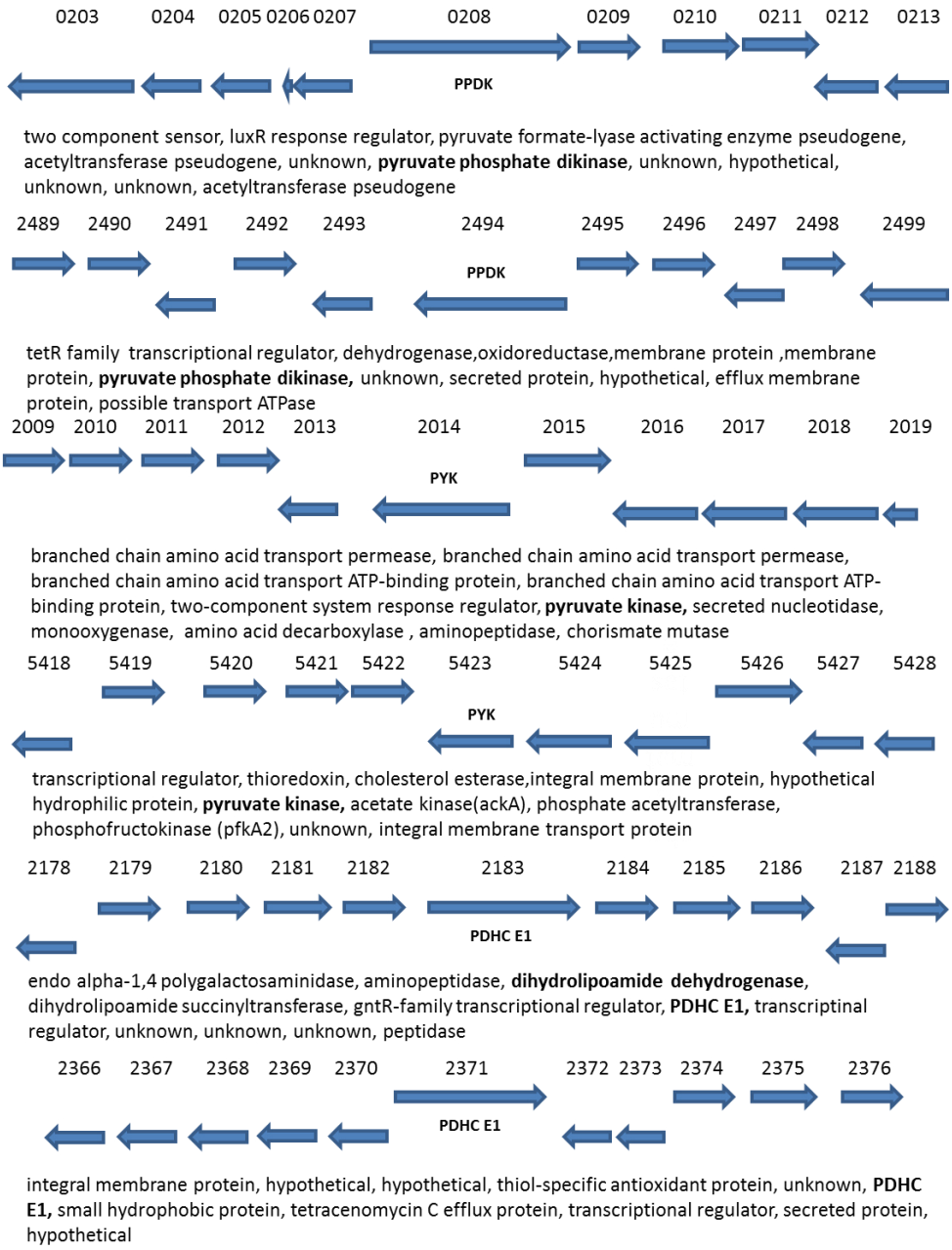



Figure 3.19 Gene context analysis of members of the PEP-PYR-OAA Node in *S. coelicolor* genome for the genes encoding for PpdK1 (SCO0208) and PpdK2 (SCO2494), pyruvate kinase with SCO2014 and SCO5423 and two genes of $PDHC_{E1}$ with SCO2183 and SCO2371

The third gene encoding a $PDHC_{E1}$, SCO7124, has an acyltransferase adjacent to it. There is also a dehydrogenase, ubiquinol-cytochrome C reductase and cytochrome B subunit and a secreted protein nearby. Most interesting is the presence of a transmembrane efflux protein, which is predicted to transport actinorhodin. Additionally there are three hypotheticals and one unknown gene in the neighbourhood (Figure 3.20).

The $PDHC_{E1\alpha/\beta}$ are always present in a cluster together with the $PDHC_{E2}$. In the case of SCO1269/70 the neighbouring genes apart from the $PDHC_{E2}$ (SCO1268) are mostly acyl carrier proteins and 3-oxoacyl synthases, reductases as well as a lipase and a halogenase (Figure 3.20).

SCO3816/17 is adjacent to SCO3815 which encodes $PDHC_{E2}$ (dihydrolipamide acyltransferase). There are also several regulatory genes adjacent: two GntR- family regulators, and a two-component system transcriptional regulator. Additionally two serine/threonine protein kinases are present, a membrane protein and a DNA-binding protein (Figure 3.20).

SCO3829/30 is adjacent to SCO3831 which encodes $PDHC_{E2}$ and has genes adjacent for an ABC transport transmembrane component, a membrane protein, two genes for molybdenum cofactor biosynthesis upstream, downstream there are two transcriptional regulators, two dehydrogenases and a rRNA methylase (Figure 3.20).

$PDHC_{E3}$, SCO0884 and SCO4919, both encode an oxidoreductase and are not part of an operon. SCO0884 has genes adjacent encoding for an ABC-transport protein, a membrane protein, two unknowns and polypeptide deformylase upstream and downstream a gene for a thioredoxin, LacI-family transcriptional regulator and a TetR-family regulator, a secreted protein and a LysR-family transcriptional regulator. SCO4919 has a genes adjacent downstream that encode deoxyribose-phosphate aldolase, an integral membrane protein, a phosphomannomutase, a purine nucleoside phosphorylase and one hypothetical. Downstream the gene has a deoR-family transcriptional regulatory gene, an acyl-CoA carboxylase, two membrane proteins and one hypothetical adjacent (Figure 3.20).

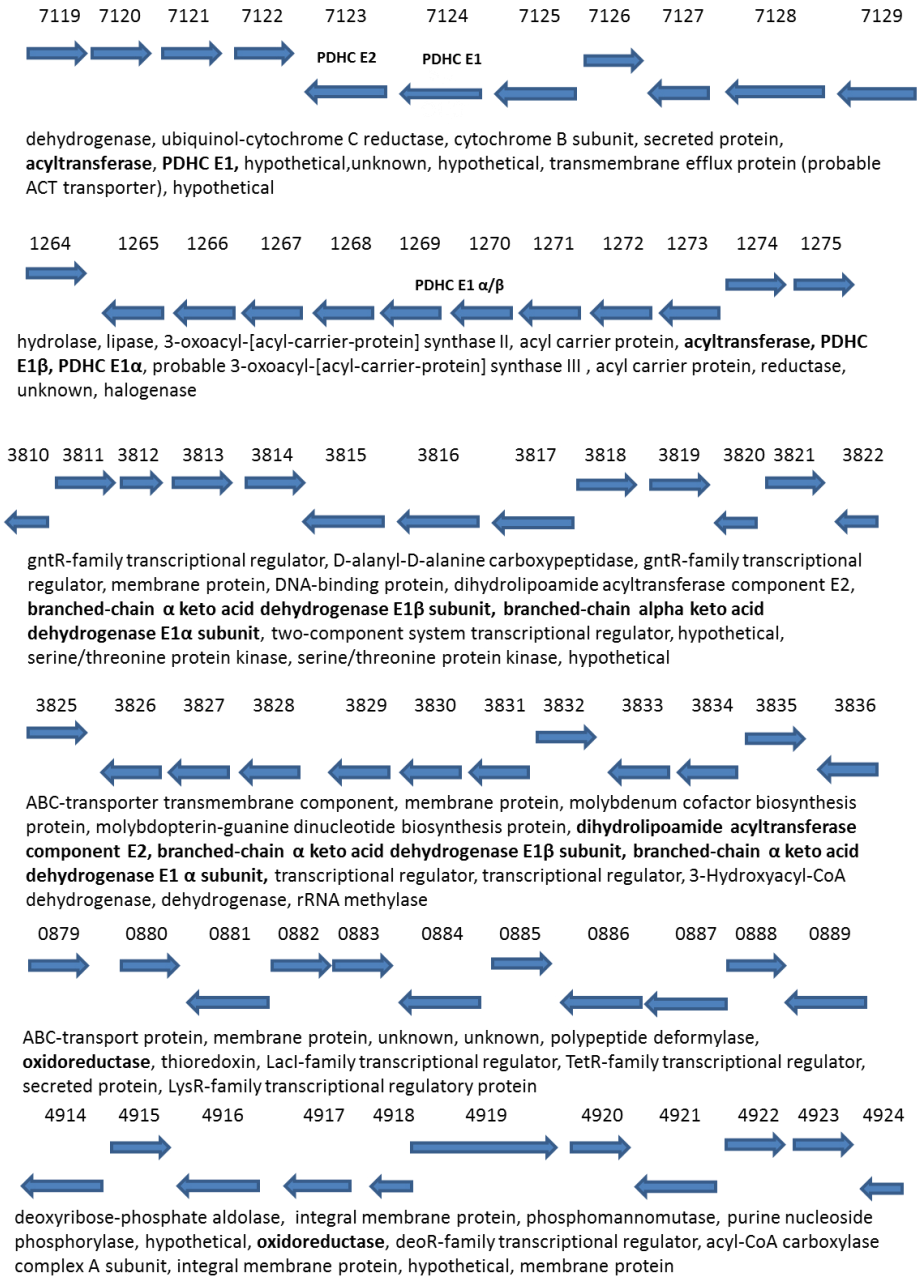



Figure 3.20 Gene context analysis of members of the PEP-PYR-OAA Node in *S. coelicolor* genome for the genes encoding the pyruvate dehydrogenase complex $PDHC_{E1}$ (SCO7124), $PDHC_{E1\alpha/\beta}$ (SCO1269/70/3816/17/3829/30) $PDHC_{E2}$ (SCO1269/SCO3815/SCO3831) and $PDHC_{E3}$ (SCO0884/4919)

## 3.4 SUMMARY

The PEP-PYR-OAA node shows all potential reactions, which can be part of it and many are expanded in *S. coelicolor* A3(2). The heat plot of the metabolic expansion revealed that those genera showing a higher number of expansions are those known to produce specialised metabolites (see Appendix 10.4).

From the in total 14 expansions detected, it was only possible to analyse 11 of the 14 expansions as the alignments were too diverse for AAT, DPD and GS and too few positions could be used to calculate a meaningful tree. Of the remaining 11, two actually only had a single copy, which was the case for AGK and ASL; two were due to a duplication (PFK and Pyk); and malic enzyme, chorismate mutase, methionine synthase and ASAD were divergent enzymes, which was also true for PPDK, $PDHC_{E1}$ and SHMT. Furthermore the pyruvate dehydrogenase, which is part of the metabolic node could also not be analysed as its alignment was of insufficient quality.

The most intriguing examples are the pyruvate kinases and phosphofructokinase that show distinctive gene duplication events for these enzymes. It is known from the literature that the deletion of PFK has an effect on actinorhodin production (Borodina *et al.*, 2008; Siebring, 2010). Thus pyruvate kinase was identified as an interesting target to study in more detail. Duplication is generally considered to be less common than HGT for acquiring new enzymatic function and maintenance of redundant copies, especially in essential genes where it seems to be even rarer (Treangen and Rocha, 2011; Noda-García and Barona-Gómez, 2013). In case of malic enzyme a divergence regarding the cofactor was observed. As these have been studied previously (Rodríguez *et al.*, 2012), they were not further analysed in this thesis.

The enzymes for amino acid metabolism were not further investigated as the focus of the thesis was on the metabolic node connecting the major carbon flux routes. However they were included in this analysis for completeness and to test the hypothesis if precursor supply routes are a good target for metabolic engineering efforts.

To assess the importance of each member of the PEP-PYR-OAA node and any role of their expansions their role in regard to growth and specialised metabolite production phenotypic studies were undertaken- which is the topic of the following chapter. This was done in order to identify interesting targets for metabolic engineering and improving specialised metabolite production.

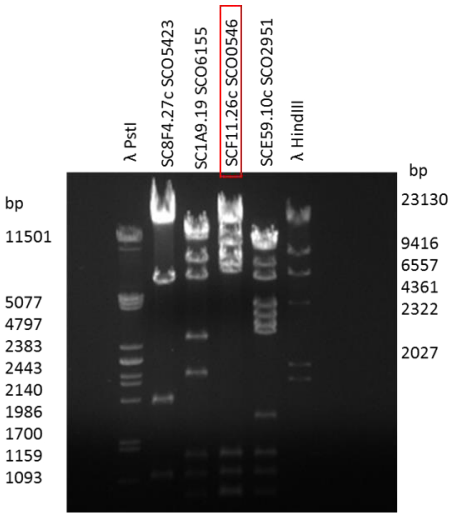
# 4 INFLUENCE OF GENES IN THE PHOSPHOENOLPYRUVATE-PYRUVATE-OXALOACETATE NODE ON GROWTH AND SPECIALISED METABOLITE PRODUCTION IN *S. COELICOLOR* A3(2)

This chapter analyses the influence of the genes from the metabolic node of interest on the growth and specialised metabolite production in *S. coelicolor* by gene disruption and physiological experiments. The work was made possible through assistance from Dr Lorena Fernández-Martinez, who made the following transposon insertion mutants; SCO2014 (*pyk1*), SCO3127 (*ppc*), SCO4979 (*ppck*), SCO7412 (*poxB2*) and SCO7124 (*aceE3*) and Dr Pablo Cruz-Morales, who provided ReDirect deletion mutants of SCO2014 (*pyk1*) and SCO5423 (*pyk2*). Some strains were verified at the University of Strathclyde and the transposon insertion strains for SCO5423 (*pyk2*), SCO2951 (*me-NAD*), SCO6155 (*poxB1*) and SCO0546 (*pyc*) were constructed as part of this thesis (Table 2-4).

## 4.1 CONSTRUCTION OF TRANSPOSON INSERTION MUTANTS

The cosmids SC8F4.27 (SCO5423), SCF11.26 (SCO0546), SCE59.10 (SCO2951) and SC1A9.19 (SCO6155) were obtained from the transposon insertion library curated by the University of Swansea (Bishop *et al.*, 2004). The cosmid carries a Supercos-1 backbone with a kanamycin resistance marker (Redenbach *et al.*, 1996). A map was constructed for each cosmid *in silico* and a restriction digest was carried out using *Xho*I and *Hind*III for verification and to obtain the orientation of the transposon in the cosmid. After verification each cosmid was used to transform the non-methylating *E.coli* strain ET12567/pUZ8002 and the obtained strain was then conjugated with WT M145 to transfer the cosmid. Plates were overlayed with 50 µg/ml of apramycin to maintain selection of the mutation and 25 mg/ml nalidixic acid to remove *E. coli*. Primary excongugants were selected and streaked on to fresh medium containing 50 µg/ml of apramycin and at the same time on an identical plate containing 50 µg/ml of apramycin and kanamycin. Colonies with apramycin resistance, but kanamycin sensitive were used to prepare spore stocks from single colonies. The mutant *pyc*::Tn5062 was characterised by a Masters student (João Cruz, MSc thesis, 2012) and will thus not be further discussed in this thesis. The other strains were screened for changes in actinorhodin production, and if a

difference was detected, were utilised in further experiments and were verified by PCR (Figure 4.1-Figure 4.4).




| Gene of interest/ Cosmid | Restriction enzyme | Expected Digest pattern (bp) |
|---|---|---|
| SCO0546 (pyc) SCF11.26c | *Xho*I, *Hind*III | a.21927,10877,7687,6664,1022,841,699,694 b.21927,13212,6664,5352,1022,841,699,694 |

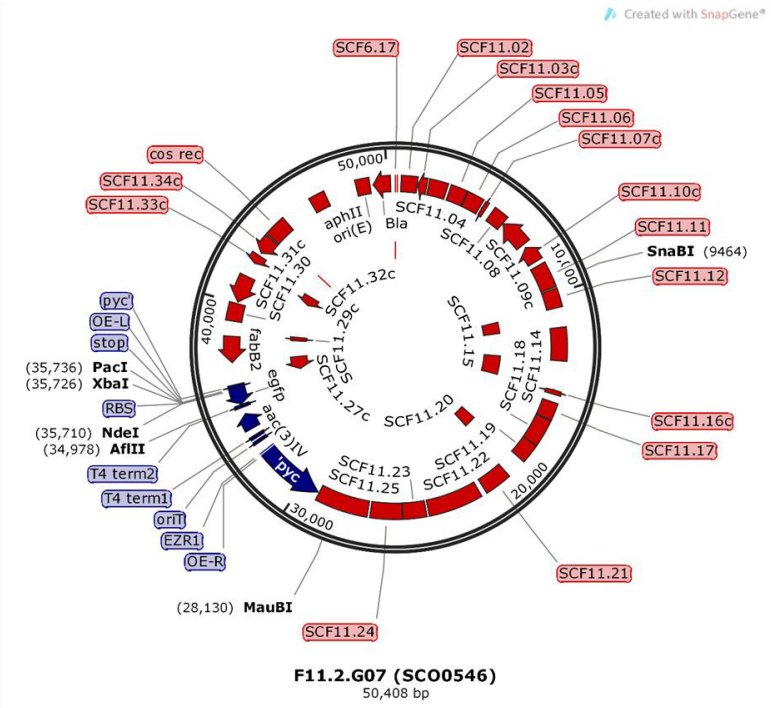
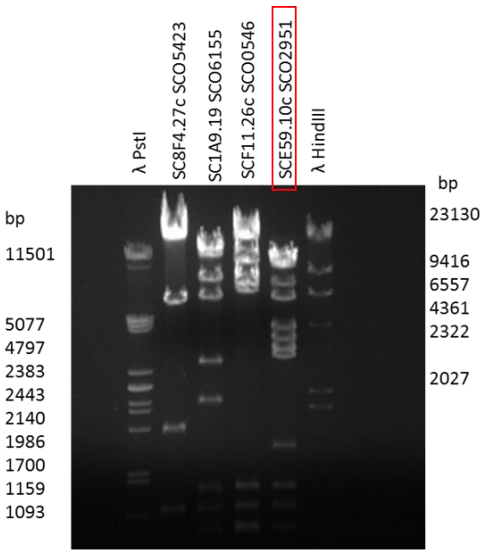



Figure 4.1 Cosmid F11.2 map and restriction digest verification and establishment of orientation, blue colour in the map indicates location of Tn5062 disrupting SCO0546 (*pyc*)




| Gene of interest/ Cosmid | Restriction enzyme | Expected Digest pattern (bp) |
|---|---|---|
| SCO2951 (me) SCE59.10c | *Xho*I, *Hind*III | a.11228,10421,6092,4221,3727,3338,3147,1423,1022,841,699 b.11228,8086,6092,5482,4221,3727,3338,1423,1022,841,699 |

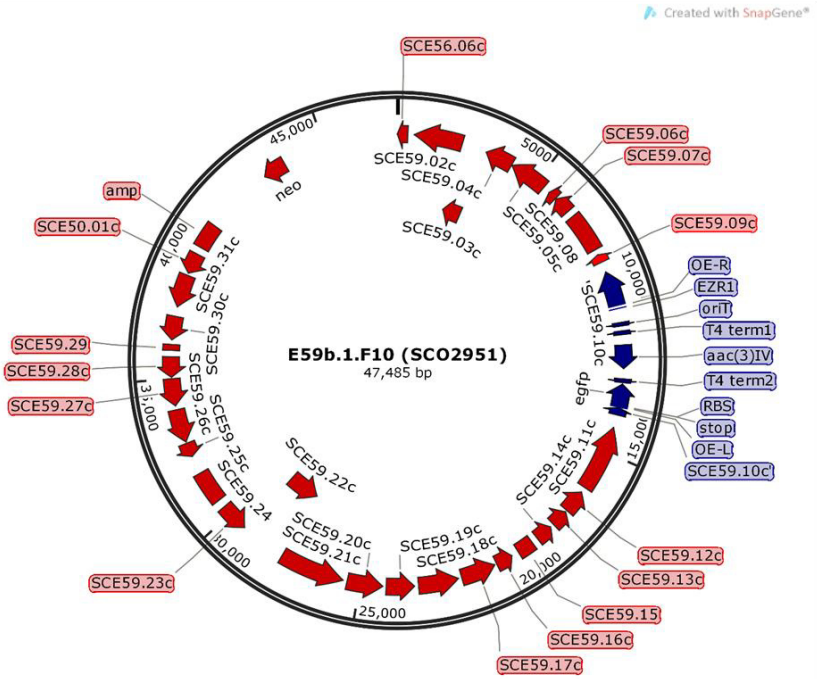
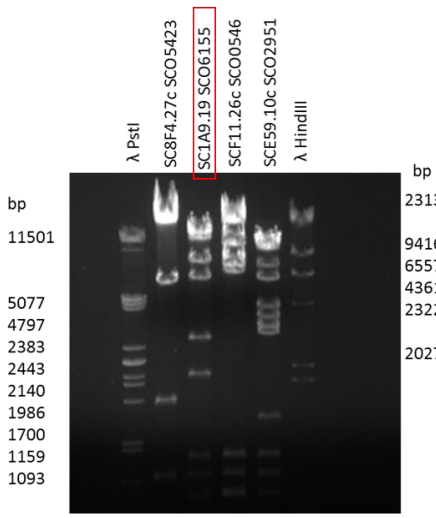



Figure 4.2 Cosmid E59.1 map and restriction digest verification and establishment of orientation, blue colour in the map indicates location of Tn5062 disrupting SCO2951 (*meNAD*)




| Gene of interest/ Cosmid | Restriction enzyme | Expected Digest pattern (bp) |
|---|---|---|
| SCO 6155 (poxB1) SC1A9.19 | *Xho*I, *Hind*III | a.16120,11037,8373,6 300,5377,2120,1022,8 41,699<br>b.16120,13372,8373,6 300,3042,2120,1022,8 41,699 |



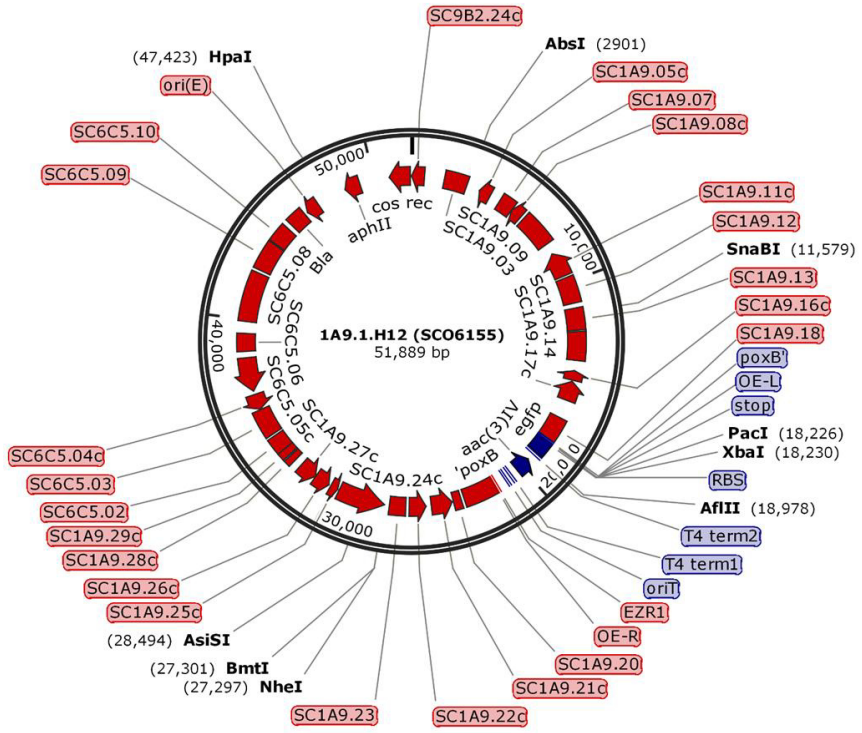
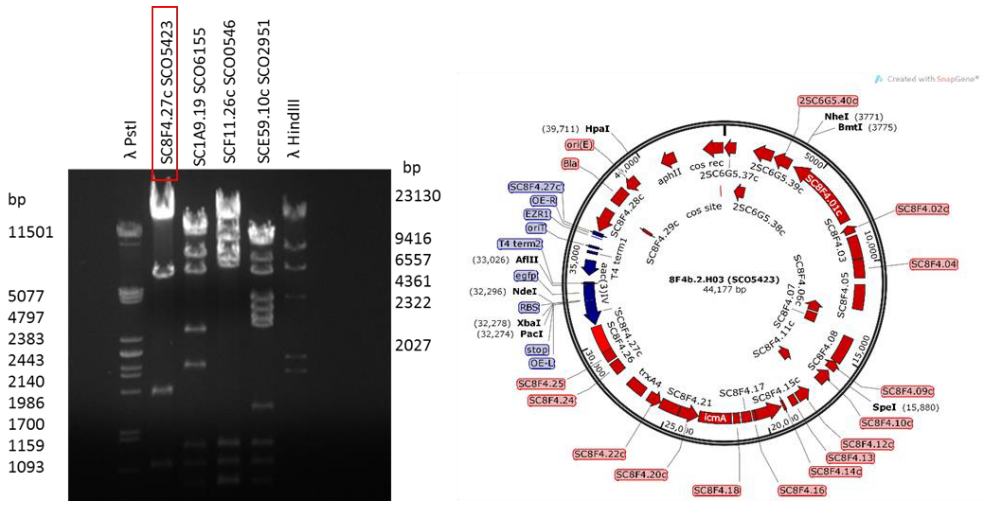



Figure 4.3 Cosmid 1A9.1 map and restriction digest verification and establishment of orientation, blue colour in the map indicates location of Tn5062 disrupting SCO6155 (*poxB1*)




| Gene of interest | Cosmid | Restriction enzyme | Expected digest pattern (bp) |
|---|---|---|---|
| SCO5423 (*pyk2*) | SC8F4.H03 | *Hind*III | a. 37665, 3950, 1721, 841<br>b. 35330, 6285, 1721, 841 |
| | | *Xho*I | 15867,12903,7571,4769,2152,915 |
| SCO5423 (*pyk2*) | SC8F4H03::Tn5066 | *Xho*I | 15867,15780,7571,4769,915 |

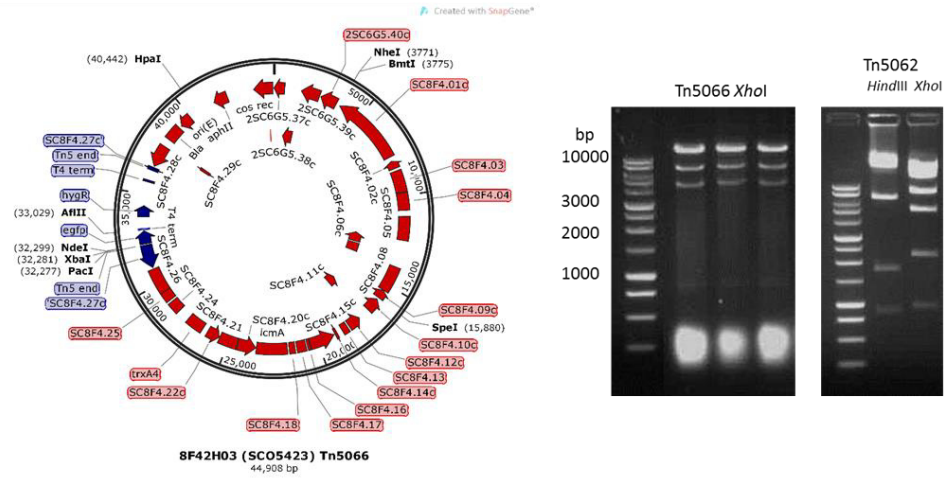



Figure 4.4 Top: Cosmid 8F4.H03 map and restriction digest verification and establishment of orientation, blue colour in the map indicates location of Tn5062 disrupting SCO5423. Bottom: Cosmid 8F4.H03 map with Tn5066 disrupting SCO5423 (*pyk2*), exchanged resistance marker and verified by restriction digest

## 4.2 Target Selection through Phenotypic Screening of PEP-PYR-OAA Node Mutants

Strains that were mutated in the genes of interest from the PEP-PYR-OAA node were subjected to a rapid screen of their phenotype in liquid minimal medium supplemented with glucose and casamino acids at small scale with a focus on selecting targets for metabolic engineering for increased polyketide production. A minimal medium was used for this as the genes encode proteins that function within primary carbon metabolism and it was expected that rich media might mask phenotypes due to the inherent robustness of primary metabolism. Each mutant was grown in 1.5 ml of supplemented minimal medium with glucose as carbon source for seven days. Cultures were screened on the basis of altered actinorhodin production (endpoint determination) when compared to *S. coelicolor* M145.

This initial screen showed that several mutants in the PEP-PYR-OAA node exhibited altered actinorhodin production. The *aceE3*::Tn*5062* (SCO7124) mutant exhibited an approximate 4 fold increase in actinorhodin yield compared to the WT and *pyk1*::Tn5062 (SCO2014) and *pyk2*::Tn5062 (SCO5423) with approximately two-fold increase in yield. The malic enzyme *meNAD*::Tn5062 (SCO2951) showed a three-fold yield increase and Δ*poxB2* (SCO7412) approximately two-fold. Disruption of *poxB1*::Tn5062 (SCO6155), *ppck*::Tn5062 (SCO4979) and *ppc*::Tn5062 (SCO3127) appeared to have no effect of the yield of actinorhodin compared to M145 (Figure 4.5). Whilst variation was high in this small-scale cultivation system, it proved to be a reasonable rapid screen for phenotypic analysis of large numbers of strains, to identify strains for more detailed analysis. Those were *aceE3*::Tn5062, *pyk1*::Tn5062, *pyk2*::Tn5062 and *meNAD*::Tn5062. Two malic enzymes from *S. coelicolor* are already characterised in the literature, so they were not pursued further (Rodríguez *et al.*, 2012).

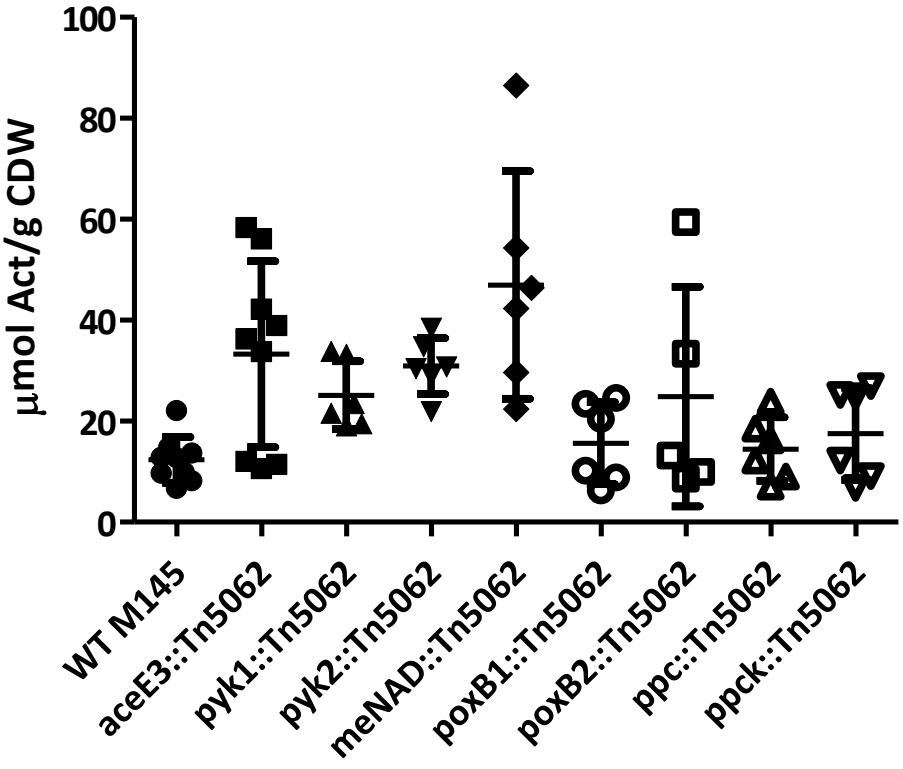



Figure 4.5 Small scale screening (1.5 ml volume) for change in actinorhodin production (given as yield per cell dry weight) of the transposon insertion mutants of the metabolic node as endpoint measurement

## 4.3 Phenotypic Characterisation of Mutants on Solid Medium

In addition to screening mutant strains in liquid minimal medium, the strains were also analysed phenotypically on a range of solid agar media. Strains were screened and compared to M145 on in total 15 different media. However only those media that resulted in phenotypic differences between M145 and the mutant strains are presented for clarity. As in the liquid initial screen for interesting targets, several of the mutants exhibited a different phenotype compared to M145.

The mutant deficient in $PDHC_{E1}$ (*aceE3*::Tn5062) showed an increase in actinorhodin production on YEME and on minimal medium with either pyruvate or glucose as carbon source. On Tween80 and N-acetylglucosamine this strain had an increase in undecylprodigiosin production compared to M145 (Figure 4.6).

The *pyk1*::Tn5062 mutant showed antibiotic overproduction on minimal medium containing malate, pyruvate, N-acetylglucosamine and glucose as the sole carbon source (Figure 4.6). Intriguingly, the *pyk2*::Tn5062 mutant exhibited fewer differences from the WT M145 and overproduced undecylprodigosin rather than actinorhodin on malate, pyruvate and N-acetylglucosamine (Figure 4.6).

The *meNAD*::Tn5062 mutant had less actinorhodin production on minimal medium with malate than the WT (Figure 4.6).

The *ppck*::Tn5062 mutant had increased actinorhodin production on N-acetylglucosamine and YEME (Figure 4.7). The *ppc*::Tn5062 mutant showed an increased actinorhodin production on pyruvate, N-acetylglucosamine, glucose (KMM) and YEME (Figure 4.7).

The *poxB2*::Tn5062 mutant showed an increased undecylprodigiosin production on R2 and a higher actinorhodin production on pyruvate and YEME (Figure 4.7).The *poxB1*::Tn5062 mutant showed increased specialised metabolite production on mannitol-soya-agar (MS) and N-acetylglucosamine (Figure 4.7).

Solid medium showed greater differences between strains than liquid medium. However as *Streptomyces* industrial processes are liquid culture based, the focus was laid onto *aceE3* and the two pyruvate kinase for further detailed characterisation.

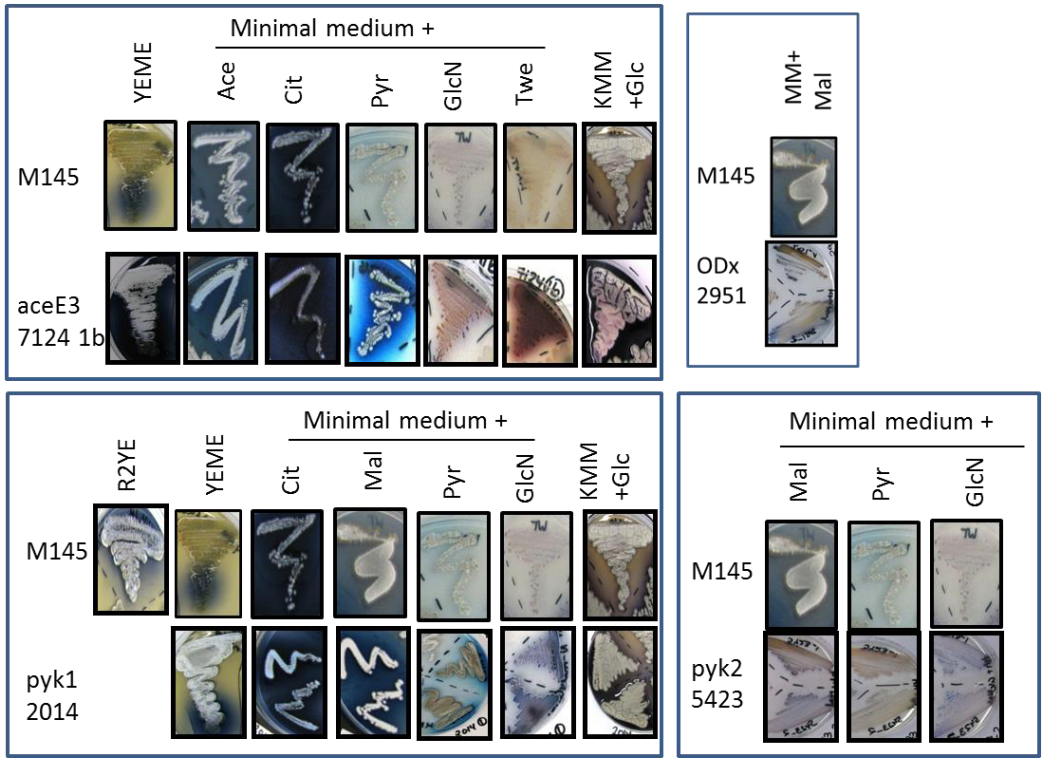



Figure 4.6 Phenotypic characterisation of transposon insertion mutants *aceE3*::Tn5062, *pyk1*::Tn5062 and *pyk2*::Tn5062 on solid medium using different carbon sources in minimal media as well as complex media. The strains were streaked together with the WT M145 for comparison. Legend: MM = Minimal Medium ($NH_4^+$ as N-source), KMM = MM from (Karandikar *et al.*, 1997, $NaNO_3$ as N-source), GlcN = N-Acetylglucosamine, Ace = acetate, Cit = citrate, Pyr = Pyruvate, Twe = Tween, Glc = Glucose, Mal = Malic acid

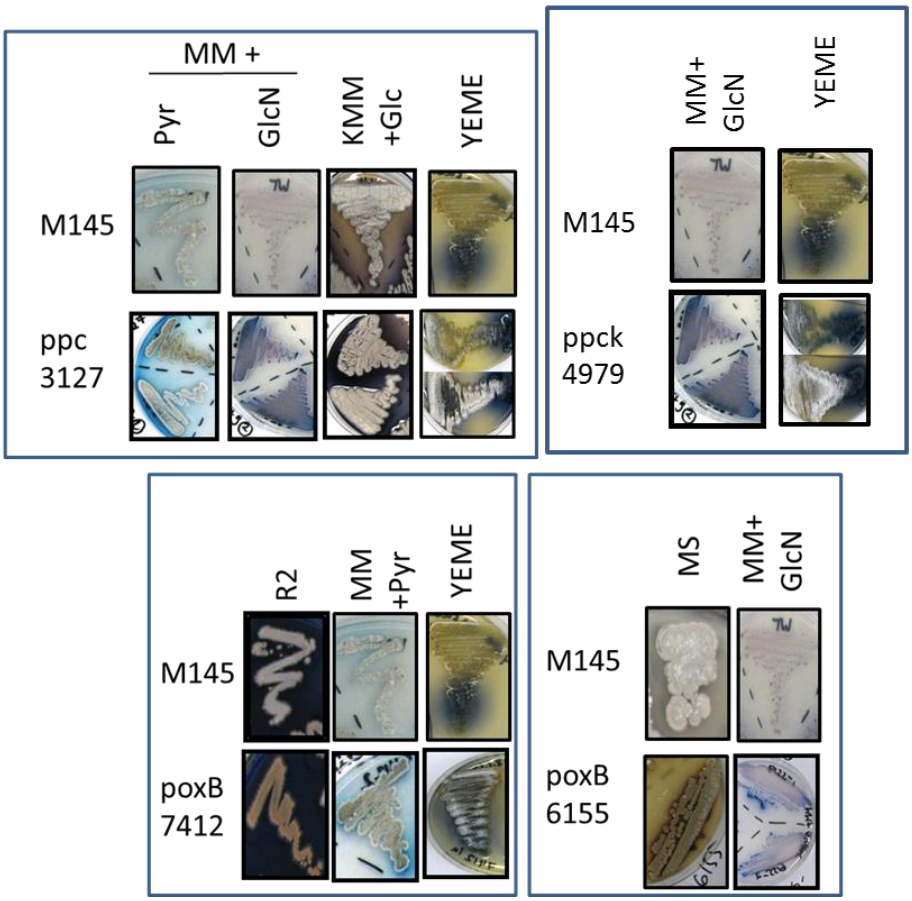



Figure 4.7 Phenotypic characterisation of transposon insertion mutants *ppc*::Tn5062, *ppck*::Tn5062, *poxB1*::Tn5062 and *poxB2*::Tn5062 on solid medium using different carbon sources in minimal media as well as complex media. The strains were streaked together with the WT M145 for comparison. Legend: MM = Minimal Medium ($NH_4^+$ as N-source), KMM = MM from (Karandikar *et al.*, 1997, $NaNO_3$ as N-source), GlcN = N-Acetylglucosamine, Ace = acetate, Cit = citrate, Pyr = Pyruvate, Twe = Tween, Glc = Glucose, Mal = Malic acid

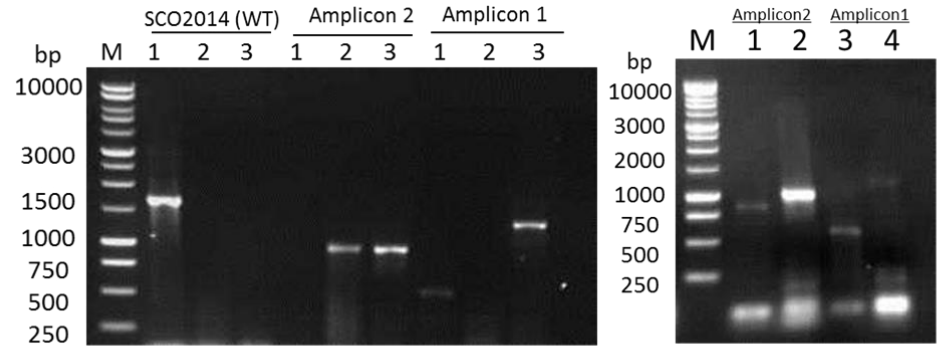
## 4.4 Construction of a Pyruvate Kinase Double Mutant

As there are two pyruvate kinase genes in the genome, a double mutant in *S. coelicolor* M145 with no functional pyruvate kinase was constructed. The apramycin resistance marker (Tn5062) was exchanged on the cosmid 8F4.H03 to hygromycin (Tn5066) by restriction digest of PQM5066 with *Pvu*I digestion (Fernández-Martínez *et al.*, 2011). This was necessary in order to have two different resistance markers on each of the two pyruvate kinase. The successful exchange was verified by restriction digests (Figure 4.4). The strain Δ*pyk1* was then conjugated with *E. coli* 12567/pUZ8002 carrying the obtained cosmid disrupting *pyk2* by Tn5066. This last step was performed by Dr Karina Verdel (member of Dr Francisco Barona-Gomezs' lab in Irapuato, Mexico). The strain has not been verified yet.

## 4.5 Confirmation of single Pyruvate Kinase Mutants in *S. coelicolor*

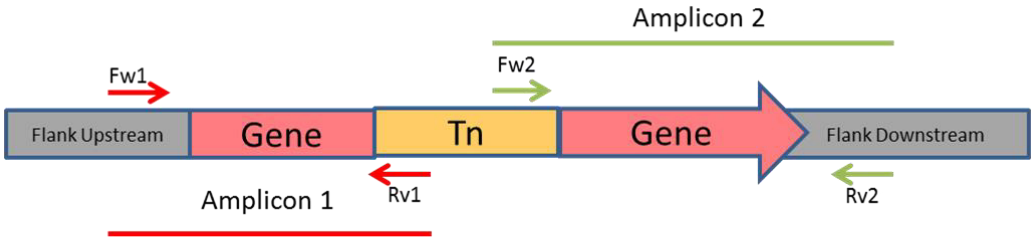
The pyruvate kinase mutants (*pyk1*::Tn5062, *pyk2*::Tn5062, Δ*pyk1*, Δ*pyk2*) were confirmed by PCR. For the transposon insertion mutants, this was carried out using two different PCR reactions. Each one had one primer in the flanking region, either upstream or downstream to the gene and one primer inside the transposon. Using this approach two unique PCR products were obtained from the mutant strain, which cannot be generated from the parental (WT) strain as it does not contain the transposon and thus the primer inside the transposon will not be able to bind. For *pyk1*::Tn5062 amplicon 1, the forward primer was 182-202 bp upstream of the gene and the reverse primer inside the Tn5062 at position 197-217 bp into the transposon with expected size of 1522 bp. For amplicon 2 the forward primer was inside the transposon, at 2944 -2964 bp and the reverse primer was binding to 44-64 bp downstream of *pyk1* with an amplicon size of 896 bp (Figure 4.8). For *pyk2*::Tn5062 amplicon 1, the forward primer was 212-227 bp upstream of *pyk2* and the reverse primer inside the Tn5062 at position 197-217 bp with expected size of 1719 bp. For amplicon 2, the forward primer was inside the transposon at 2944 -2964 bp and the reverse primer was binding to 241-263 bp downstream of *pyk2* with an amplicon size of 917 bp (Figure 4.9).

The ReDirect deletion strains Δ*pyk1* and Δ*pyk2* constructed by Dr Pablo Cruz Morales were verified by PCR using Taq polymerase (MyTaq, Bioline) to check for the presence of *pyk2* and *pyk1* and absence *pyk1* and *pyk2* respectively in the two strains. Primers inside each gene were chosen for this (RT-PCR primer, Table 2-7). Primers for *pyk1* gave an amplicon of 489 bp and 438 bp in the case of *pyk2* (Figure 4-10).




1 = WT gDNA
2 + 3 = *pyk1*::Tn5062

1 + 3 = WT
2 + 4 = *pyk1*::Tn5062




Fw1: 182-202 bp upstream of pyk1
Rv1: 197-217 bp of Tn
Amplicon 1 size: 1522 bp

Fw2: 44-64 bp downstream of pyk1
Rv2: 2944--2964 bp of Tn
Amplicon 2 size: 896 bp

Figure 4.8 Verification of *pyk1*::Tn5062 by PCR using a forward primer at 182-202 bp upstream of the pyk1 and the reverse primer inside the Tn5062 at position 197-217 bp into the transposon, the expected size was of 1522 bp. Amplicon 2 the forward primer was inside the transposon, at 2944 - 2964 bp inside the transposon and the reverse primer was binding to 44-64 bp downstream of *pyk1*, the amplicon size was of 896 bp.

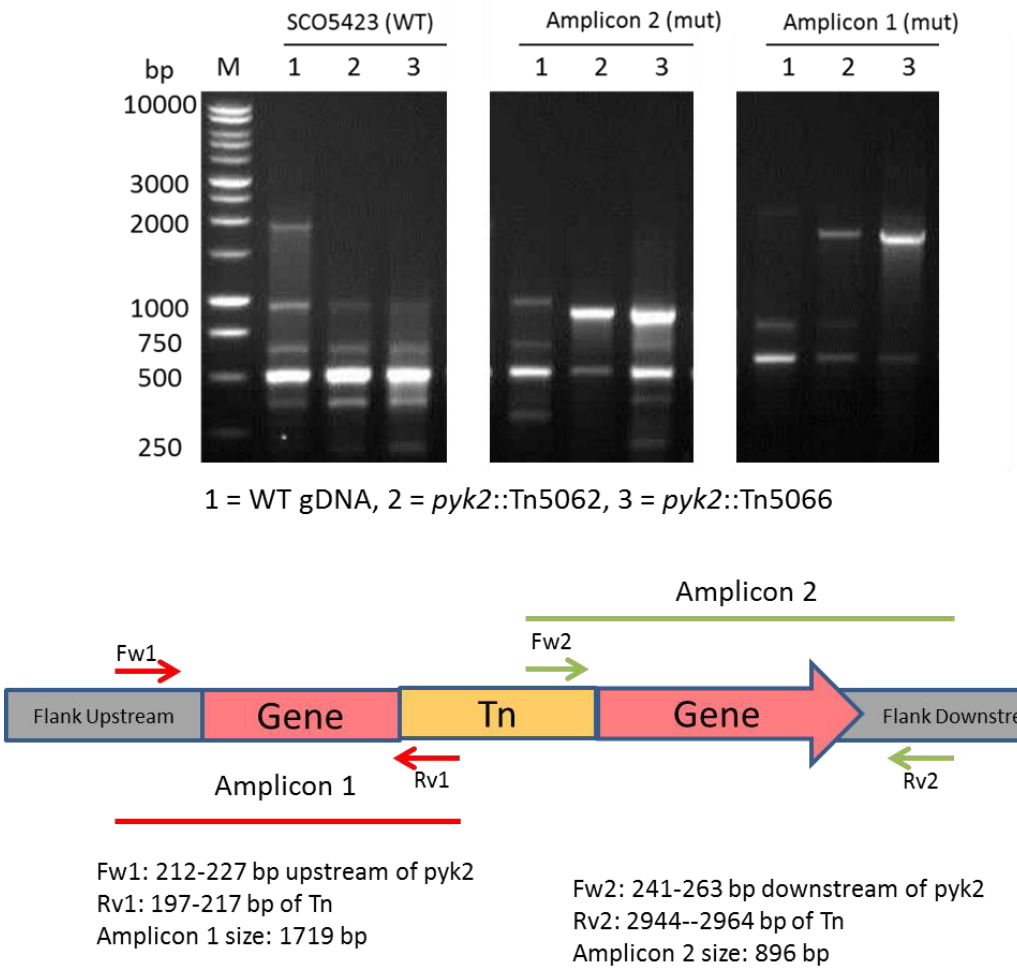



Figure 4.9 Verification of *pyk2*::Tn5062 by PCR using a forward primer at 212-227 bp upstream of *pyk2* and the reverse primer inside the Tn5062 at position 197-217 bp into the transposon, the expected size was of 1719 bp. Amplicon 2 the forward primer was inside the transposon, at 2944 -2964 bp inside the transposon and the reverse primer was binding to 241-263 bp downstream of *pyk2*, the amplicon size was of 917 bp.

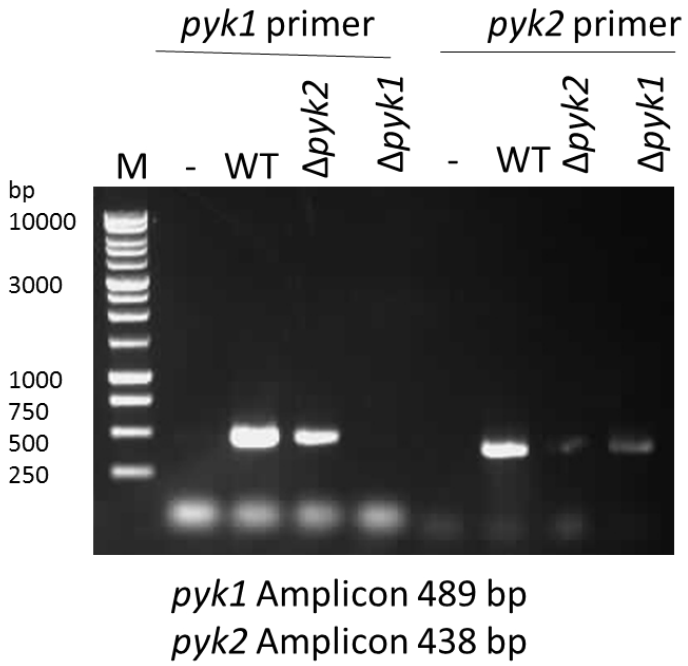



Figure 4.10 Verification of Δ*pyk1* and Δ*pyk2* by PCR to check for the presence of *pyk2* and absence *pyk1* in case of Δ*pyk1* and presence of *pyk1* and absence *pyk2* for the strain Δ*pyk2*. Primer inside each gene were chosen for this (RT-PCR primer, Table 2-7). The primer for *pyk1* yields an amplicon of 489 bp and 438 bp in case of *pyk2*

## 4.6 COMPLEMENTATION OF *PYK1*::TN5062, *PYK2*::TN5062 AND *ACE3*::TN5062

To demonstrate linkage of the gene disruption to the observed phenotype, genetic complementation was undertaken. To achieve this the complete coding sequence of the gene (*pyk1*, *pyk2* or *aceE3*) was amplified from genomic DNA of *S. coelicolor* M145 (WT) using Taq Polymerase (MyTaq; Bioline) and gene specific primers (Table 2-7). The amplicons included 202/227/152 bp of DNA upstream of the predicted translational start codon to ensure that any putative promoter sequence was captured respectively for *pyk1, pyk2* and *aceE3*. Amplicon (1703 bp for *pyk1*, 1921 bp for *pyk2* and 3092 bp for *aceE3*) were blunt-end cloned into pGEM-T-Easy using the manufacturer's instructions. Primary clones were selected using blue-white selection in *E. coli* DH5a on LB containing 25 µg/ml X-Gal, 100 µg/ml carbenicillin and 0.1 mM ITPG. Clones were confirmed by restriction analysis using *Sma*I (*pyk1*), *Pvu*I (*pyk2*) or *Sal*I (*aceE3*) which cut inside the insert (Figure 4.11-Figure 4.13) and the correct clones were sent for sequencing (MWG Eurofins) to ensure integrity of the insert and to ensure that no mutations were acquired during PCR amplification. The insert encoding the gene and its putative promoter region was sub-cloned into the integrative vector pIJ6902. The insert was released from pGEM-T-Easy using *EcoR*I and ligated into pIJ6902 (Figure 4.11-Figure 4.13) that was previously linearised with *EcoR*I and dephosphorylated using alkaline phosphatase (Promega) according to the manufacturer's instructions. Following ligation *E. coli* DH5α was transformed with the ligation mixture and plated onto LB agar with 100 µg/ml carbenicillin. Clones were checked and confirmed by restriction digest (Figure 4.11-Figure 4.13).

The plasmid pIJ6902 (Huang *et al.*, 2005) contains the ϕC31-integrase gene and the ϕC31 *attP* sites for integration at the specific *attB* sites within the *S. coelicolor* genome, an origin of transfer to facilitate conjugation via the non-methylating *E. coli* strain ET12567/pUZ8002 which facilitates conjugation, and a thiostrepton resistance marker. The resulting plasmids pJH06 (*pyk1*), pJH08 (*pyk2*) or pJH07 (*aceE3*) were introduced into *E. coli* ET12567/pUZ8002 and from the resulting colonies screened for apramycin and thiostrepton resistance was conjugated into *S. coelicolor* strain *pyk1*::Tn5062, *pyk2*::Tn5062, *aceE3*::Tn5062 respectively according to Kieser *et al*., (2000). Additionally, the strains *pyk1*::Tn5062 and *pyk2*::Tn5062 were cross complemented with an additional copy of *pyk1* or *pyk2*. Conjugation plates were overlayed with 50 µg/ml of apramycin to maintain selection of the mutation, with 25 µg/ml thiostrepton to maintain selection for plasmid pJH06/pJH08/pJH07 respectively and 25 µg/ml

nalidixic acid to remove *E. coli*. Primary exconjugants were selected and streaked onto fresh medium containing 50 µg/ml of apramycin and 25 µg/ml thiostrepton. Spore stocks were made from single colonies and subjected to phenotypic screening.

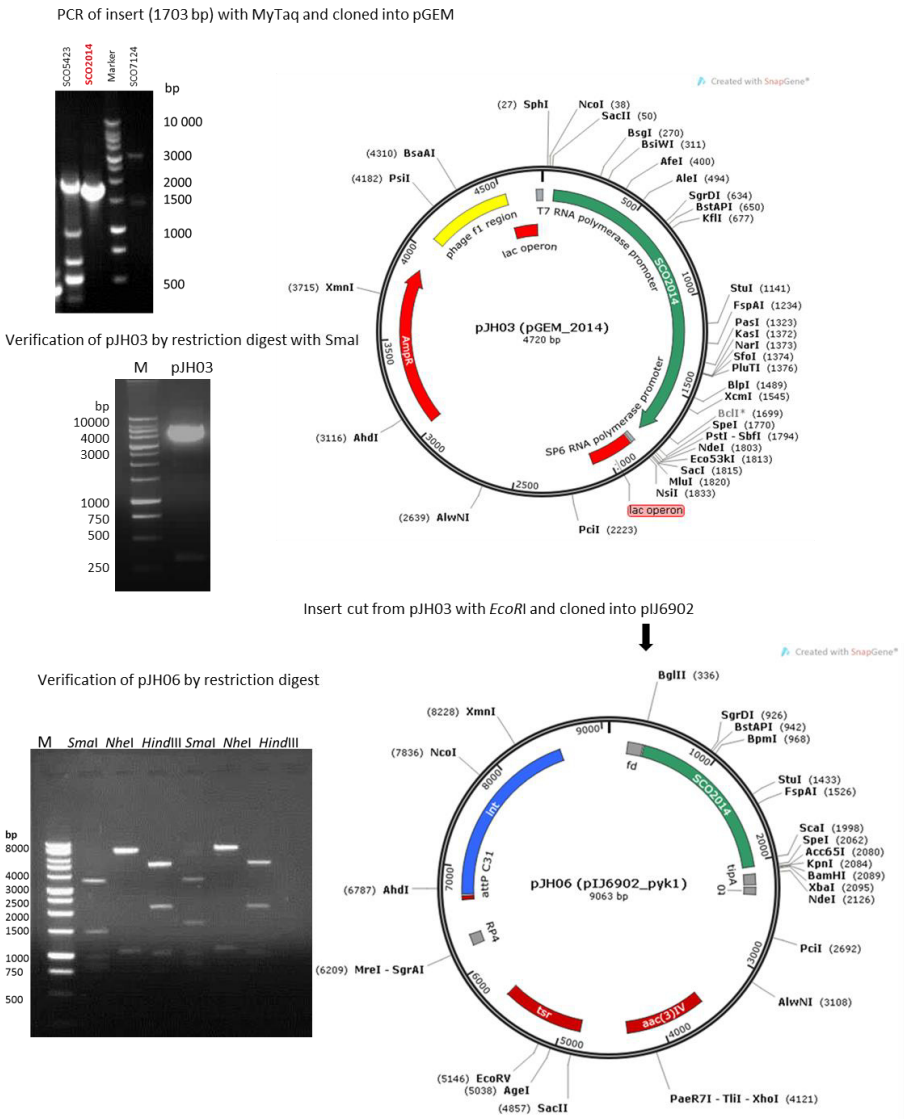
## 4.7 CHARACTERISATION OF *ACEE3*::TN5062

The strain *aceE3*::Tn5062 has a disrupted $PDHC_{E1}$ which is part of the pyruvate dehydrogenase complex responsible for the conversion of pyruvate into acetyl-CoA. In both the screening on liquid and solid medium, this mutant showed several fold increase in actinorhodin production compared to the WT M145. In order to link the observed phenotype to disruption of the gene, it was necessary to complement the mutant, which would if successful show the same phenotype as the WT M145.

Controls of both the WT and the mutant transformed with an empty vector (pIJ6902) were done to check for the influence of the plasmid without any insert. Neither the mutant nor the WT M145 with the empty vector showed a difference from the respective strain.

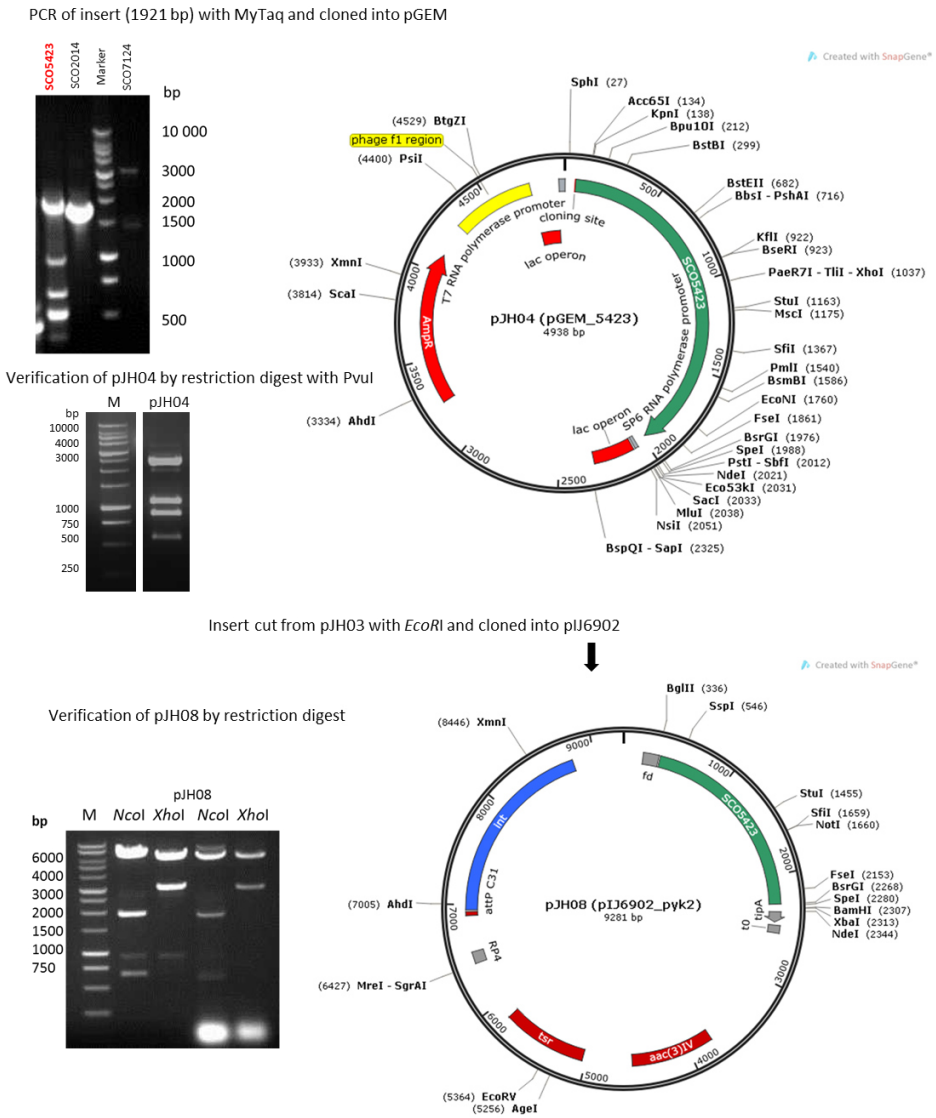
When complementing sJH *aceE3::*Tn5062 with pJH07 (pIJ6902_*aceE3*), the resulting strain sJH23 looked the same as the mutant on all media tested: minimal media with pyruvate, Tween80, acetate, citrate, GlcNAc and YEME. Thus complementation was not accomplished (Figure 4-14).

A possible explanation of the unsuccessful complementation is that it was an ectopic complementation, meaning that insertion was not at the same position as the gene is located normally. There also might be other mutations present in the strain or there might be a pleiotropic effect of the gene disruption.



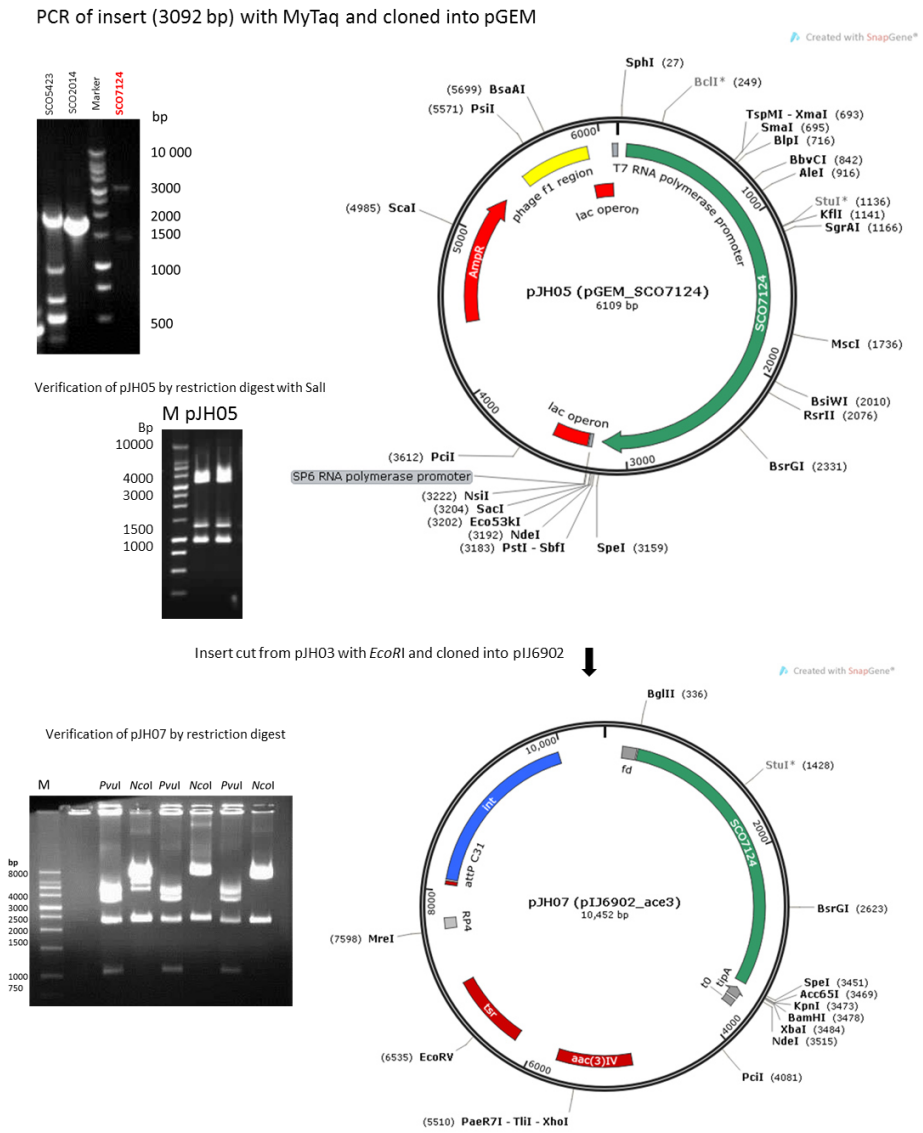

| Plasmid | Restriction enzymes (# sites) | Expected band sizes (bp) |
|---|---|---|
| pJH03 (pGEM_2014) | SmaI (2) | 4414, 306 |
| pJH06 (pIJ6902_2014) | SmaI (9) | 3598,1452,949,846,806,771,306,282,53 |
| | NheI (2) | 8027,1037 |
| | HindIII (4) | 5227, 2144, 969, 723 |

Figure 4.11 Construction of pJH06 for complementation of *pyk1*::Tn5062, *pyk2*::Tn5062, *Δpyk1* and *Δpyk2*. The gene SCO2014 (*pyk1*) was amplified by PCR yielding a 1703 bp fragment including 202 bp upstream of the gene to include a putative promoter region. This insert was blunt end cloned into the commercial vector pGEM-T-easy at the *Eco*RI site. The resulting plasmid (pJH03) was checked by restriction digest using *Sma*I, which cuts inside the insert and after positive result it was sent for sequencing (MWG eurofins). The insert was released from the pGEM insert using *Eco*RI and sub-cloned by ligation into pIJ6902, which was digested with EcoRI and dephosphorylated prior ligation. The resulting clones carrying the plasmid pJH06 were checked by restriction digest using *Sma*I, *Nhe*I and *Hind*III in three separate digestions.




| Plasmid | Restriction enzymes (# sites) | Expected band sizes (bp) |
|---|---|---|
| pJH04 (pGEM_5423) | PvuI (4) | 2459, 1096, 855,526 |
| pJH08 (pIJ6902_5423) | NcoI (3) | 6712,1912,653 |
| | XhoI (2) | 6271, 3010 |

Figure 4.12 Construction of pJH08 for complementation of *pyk1*::Tn5062, *pyk2*::Tn5062, *Δpyk1* and *Δpyk2*. The gene SCO5423 (*pyk2*) was amplified by PCR yielding a 1921 bp fragment including 227 bp upstream of the gene to include a putative promoter region. This insert was blunt end cloned into the commercial vector pGEM-T-easy at the *Eco*RI site. The resulting plasmid (pJH04) was checked by restriction digest using *Pvu*I, which cuts inside the insert and after positive result it was sent for sequencing (MWG eurofins). The insert was released from the pGEM insert using *Eco*RI and sub-cloned by ligation into pIJ6902, which was digested with *Eco*RI and dephosphorylated prior ligation. The resulting clones carrying the plasmid pJH08 were checked by restriction digest using *Nco*I and *Xho*I in two separate digestions.




| Plasmid | Restriction enzymes (# sites) | Expected band sizes (bp) |
|---|---|---|
| pJH05 (pGEM_7124) | SalI (4) | 3882, 1227, 943,57 |
| pJH07 (pIJ6902_7124) | NcoI (2) | 8108,2344 |
| | PvuI (4) | 3854,3353,2249,996 |

Figure 4.13 Construction of pJH07 for complementation of *aceE3*::Tn5062. The gene SCO7124 (*aceE3*) was amplified by PCR yielding a 3092 bp fragment including 152 bp upstream of the gene to include a putative promoter region. This insert was blunt end cloned into the commercial vector pGEM-T-easy at the *Eco*RI site. The resulting plasmid (pJH05) was checked by restriction digest using *Sal*I, which cuts inside the insert and after positive result it was sent for sequencing (MWG eurofins). The insert was released from the pGEM insert using *Eco*RI and sub-cloned by ligation into pIJ6902, which was digested with *Eco*RI and dephosphorylated prior ligation. The resulting clones carrying the plasmid pJH07 were checked by restriction digest using *Nco*I and *Pvu*I in two separate digestions.

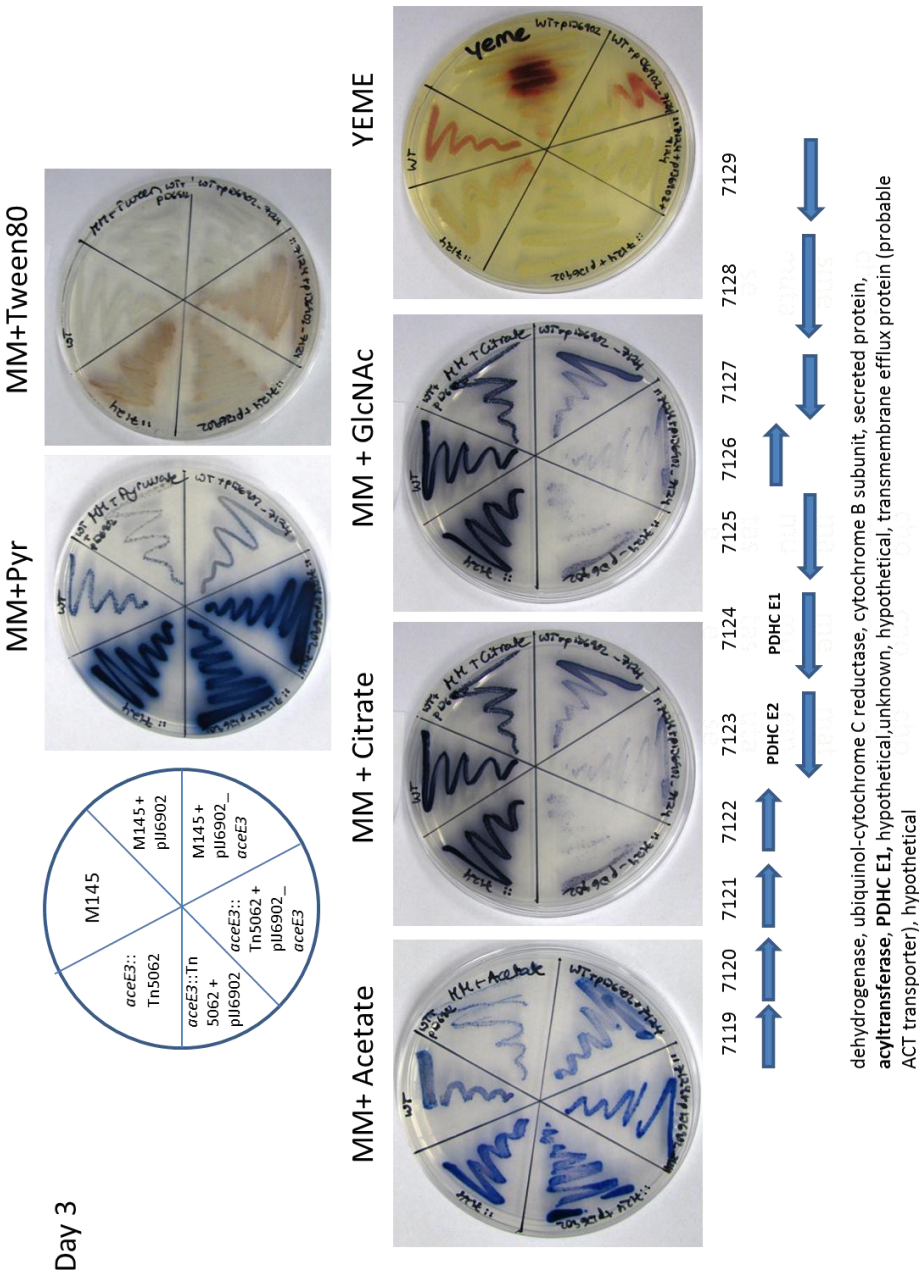



Figure 4.14 Phenotypic analysis of *aceE3*::Tn5062, *aceE3*::Tn5062 + pIJ6902, *aceE3*::Tn5062 + pJH07 in comparison with the WT M145, WT M145 + pIJ6902, WT + pJH07 on different solid minimal media with carbon sources pyruvate, Tween 80, citrate, acetate, N-acetylglucosamine and on YEME

## 4.8 Characterisation of Pyruvate Kinase Mutants

Since the pyruvate kinase mutants both showed differences to the WT phenotype on several carbon sources in the initial screening, these were investigated in more detail.

**Characterisation of transposon insertion mutants of pyruvate kinase in *S. coelicolor***

The transposon insertion mutants of the genes encoding pyruvate kinase were streaked on three different carbon sources in minimal media (glucose, GlcNAc and pyruvate) and on four complex media which were nutrient agar, R2YE, MS and YEME.

On minimal medium with 1% glucose as carbon source, the transposon insertional mutant of *pyk1* (*pyk1*::Tn5062) appeared less affected in growth than *pyk2*::Tn5062, which showed increased production of undecylprodigiosin. The provision of either *pyk1* or *pyk2* to either strain led to a similar appearance to the WT M145 (Figure 4.15).

Minimal medium with pyruvate led to an increased actinorhodin production by *pyk1*::Tn5062, which was not observed in *pyk2*::Tn5062. An additional copy of *pyk1* resulted in a decrease in actinorhodin production in the WT M145. Increased actinorhodin production on nutrient agar compared to the WT was observed for *pyk1*::Tn5062 but not for *pyk2*::Tn5062. Cross-complementation with the other pyruvate kinase seemed to lead to a bald phenotype in both cases on nutrient agar (Figure 4.15).

No apparent differences to the WT M145 were observed on N-acetylglucosamine. This was different than from the initial screening. However L-asparagine was used as nitrogen source and here it was $NH_4^+$. GlcNAc is a preferred carbon and nitrogen source for *Streptomyces*, as it is one of the building blocks of the polymer chitin and also bacterial peptidoglycan which is highly abundant in the habitat soil and upon vegetative mycelium lysis. It has been shown that under nutrient limiting conditions GlcNAc triggers the production of both actinorhodin and undecylprodigiosin by the pleiotropic GntR regulator DasR (Rigali *et al.*, 2008; Chater *et al.*, 2010; Świątek *et al.*, 2012).

Quantification of actinorhodin production on solid medium was not easy. Therefore antibiotic production in the transposon insertion pyruvate kinase mutants was studied in liquid media. Comparison of growth curves indicated that growth was unaffected in complex medium (YEME) or in SMM with pyruvate as the sole carbon source. Providing pyruvate, which is the end product of the pyruvate kinase reaction, negates the need for *pyk*.

Remarkably, disruption of *pyk1* resulted in a two-fold increase in production of the two polyketide metabolites actinorhodin and coelimycin. This increase in antibiotic yield correlated well with the phenotypes observed on solid agar medium (Figure 4.16-Figure 4.17).

**Characterisation of pyruvate kinase deletion strains in *S. coelicolor***

Pyruvate kinase deletion strains were kindly gifted by Dr Francisco Barona-Gómez and Dr Pablo Cruz-Morales (Langebio, Mexico) and these were screened and compared to the transposon insertional mutants. To assess the effect of providing additional copies of pyruvate kinase to M145, the plasmids pJH06 (*pyk1*) and pJH08 (*pyk2*) were introduced into *S. coelicolor* M145. On nutrient agar, an additional copy of either *pyk1* or *pyk2* led to impaired aerial hyphae formation and reduced actinorhodin production, whereas on minimal medium with glucose both showed increased undecylprodigiosin production. Δ*pyk2* strain did not impact on the actinorhodin production of the strain. The Δ*pyk1* strain had a higher actinorhodin production on nutrient agar and when complemented (with either *pyk1* or *pyk2*) this effect was abolished. Δ*pyk2* mutant showed no phenotype different to the WT M145. However when it was complemented with *pyk1* more aerial hyphae were made. Complementation with *pyk2* did not show this effect and resembled the WT (Figure 4.18).

On minimal medium with glucose, an additional copy of either *pyk1* or *pyk2* in the WT M145 lead to an increase in undecylprodigiosin production when compared to the WT. The Δ*pyk1* strain showed no phenotype on glucose whereas Δ*pyk2* strain showed a smaller colony size compared to the WT. This WT colony size was restored by complementation with either of the two pyruvate kinases (Figure 4.18).

During routine sub-culturing of these strains, it was observed that spores of *pyk1* mutants were less viable. To further characterise the strains, a quantitative assessment of their germination ability was undertaken. During germination the breakdown of carbon storage is essential (Suarez *et al.*, 1980) and the question was whether Pyk1 or Pyk2 play a role here. Fresh spore stocks were diluted to contain the same number of spores for WT M145, Δ*pyk1,* Δ*pyk1+ pyk1*, Δ*pyk2,* Δ*pyk2+ pyk2* and were pre germinated at 50˚C and then serially diluted and spotted on agar plates and cfu/ml was determined by counting. Both mutants were affected. However Δ*pyk1* was much more affected than Δ*pyk2* and the WT phenotype could be established after complementation for both mutant strains (Figure 4.19).

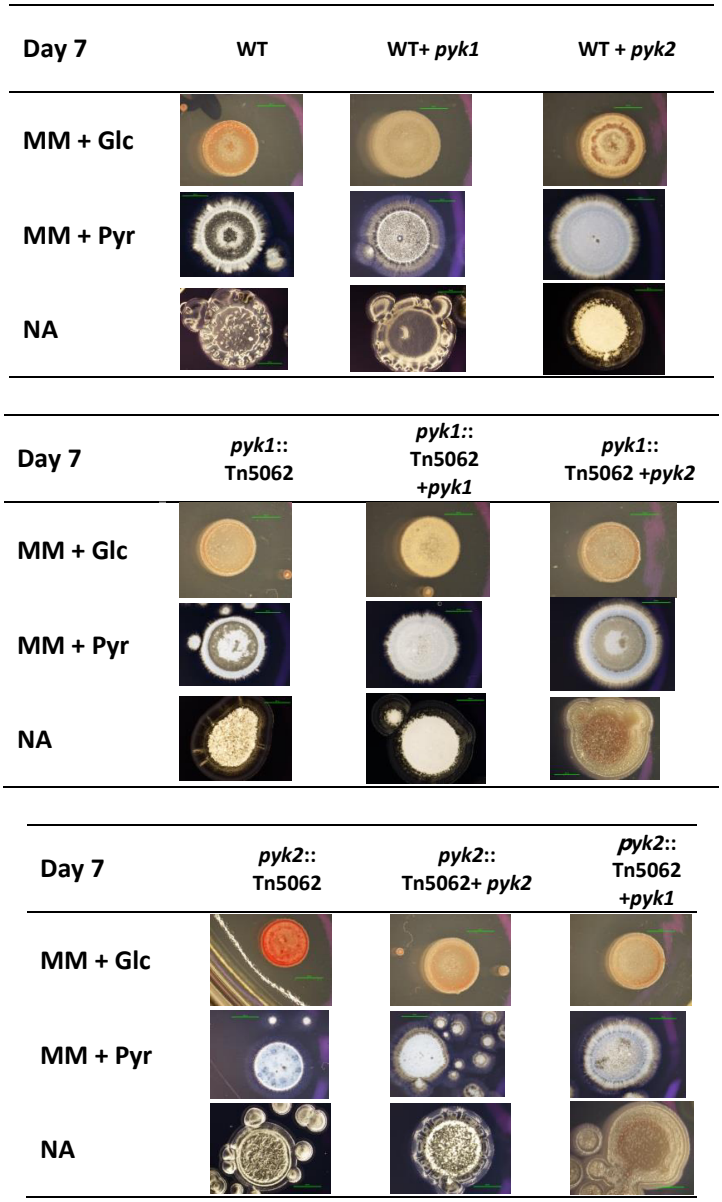



Figure 4.15 Phenotypes of pyruvate kinase transposon insertion mutants and the complemented mutants on different agar media

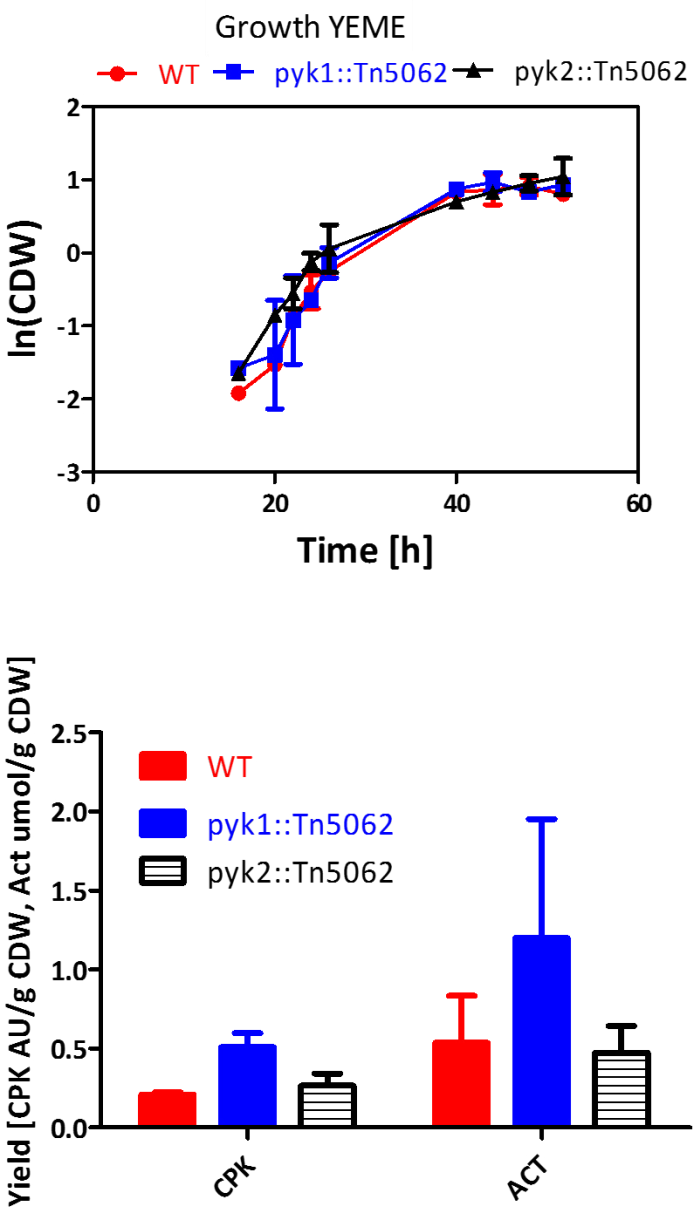



Figure 4.16 Growth curves of pyruvate kinase transposon insertion mutants on YEME medium in 400 ml medium in shaking flasks and specialised metabolite measurement of actinorhodin (ACT) and coelimycin (CPK) at 48 h of growth. Error bars represent standard deviation (σ) calculated by $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$ where N is the sample size, x each of the value and μ the average of all sample values.

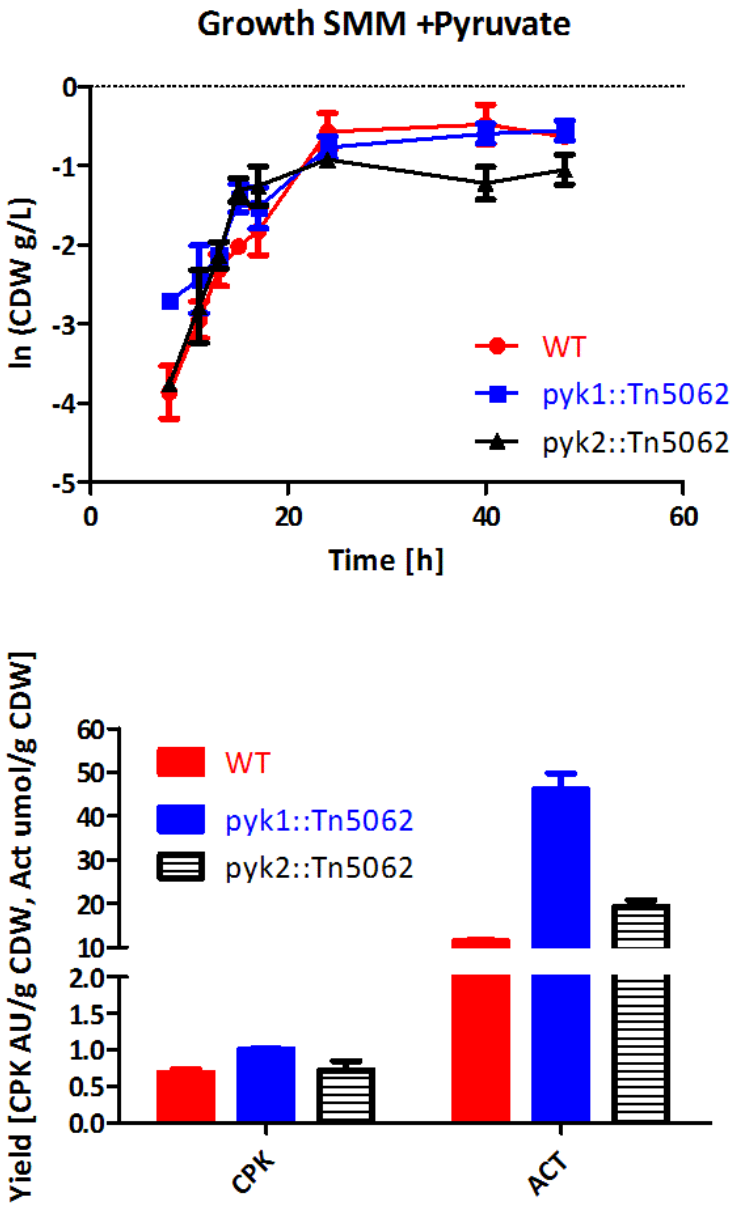



Figure 4.17 Growth curves of pyruvate kinase transposon insertion mutants on supplemented minimal medium (SMM) with 1% pyruvate in 400 ml medium in shaking flasks and specialised metabolite measurement of actinorhodin (ACT) and coelimycin (CPK) at 48 h of growth. Error bars represent standard deviation (σ) calculated by $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$ where N is the sample size, x each of the value and μ the average of all sample values.

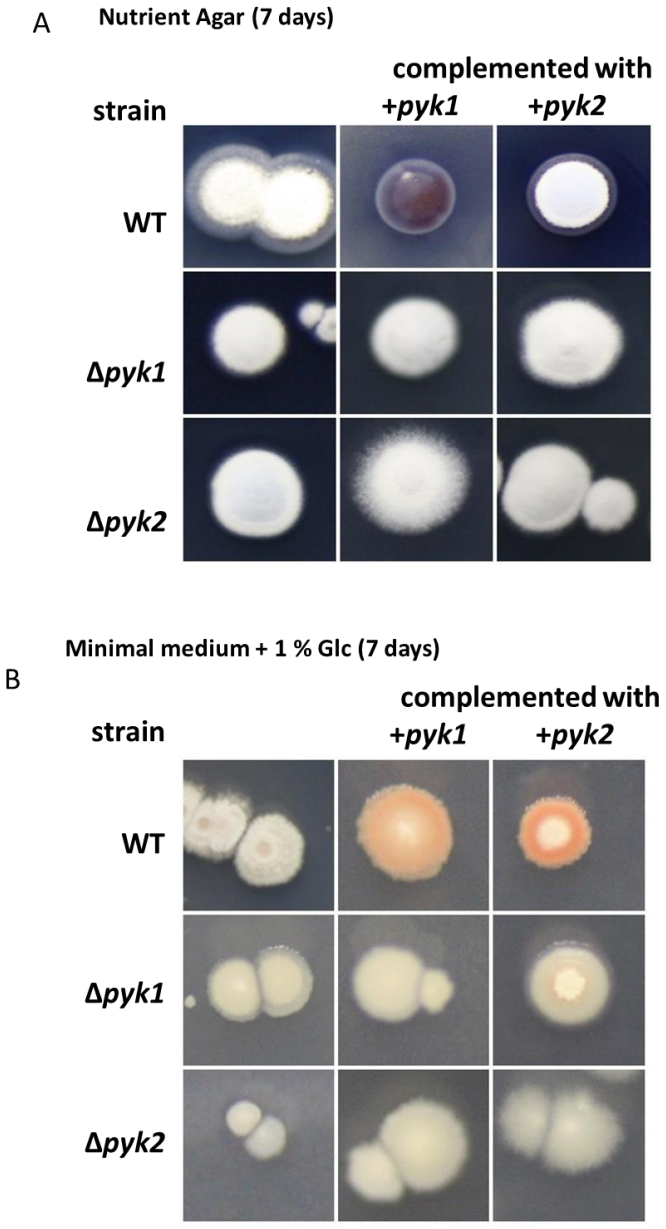



Figure 4.18 Phenotypic characterisation of pyruvate kinase mutants and the complemented strains on two different solid media. Top: Nutrient Agar, Bottom: Minimal medium with glucose

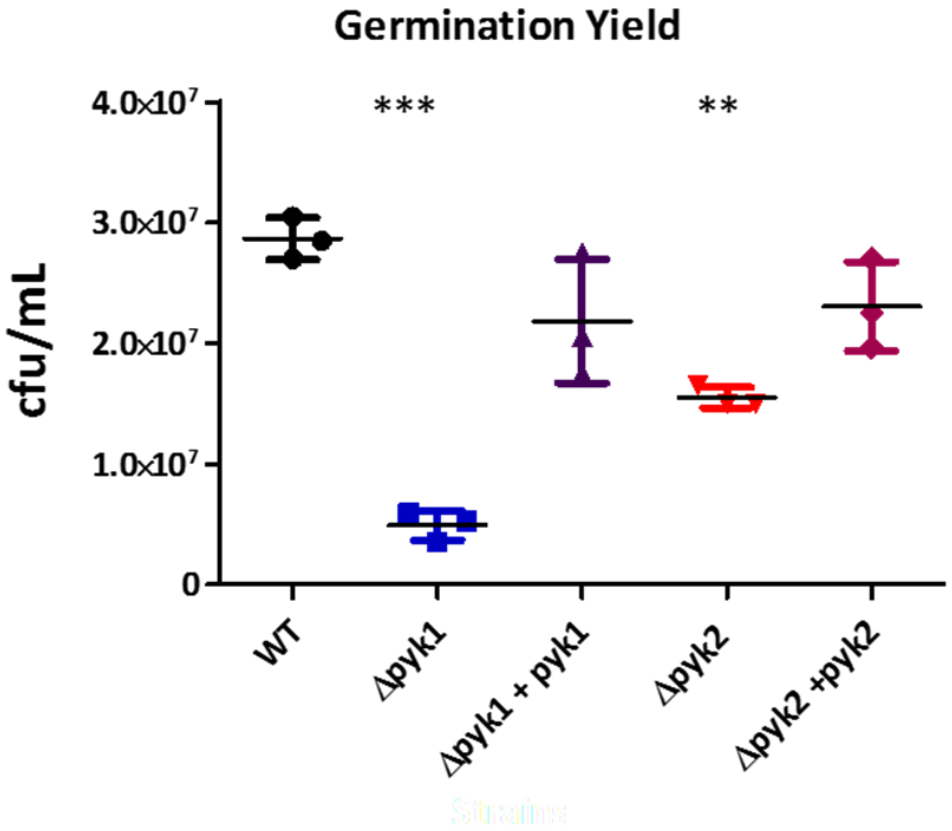



Figure 4.19 Germination yield determination of pyruvate kinase deletion mutants and complemented strains normalised to the WT M145

**Growth comparison of both transposon insertion mutants and deletion mutants in liquid medium at small scale**

To further characterise the growth characteristics of the WT and *pyk* mutants in liquid cultures, strains were grown in 96 well plates in a microplate reader (BioTek) for 48 h to study their growth behaviour. Preliminary experiments showed that growth was reproducible for up to 48 h. Under these conditions antibiotic production was limited. However, growth responses under a range of conditions could be studied in detail. This could be a useful tool in strain improvement to compare them against a reference strain and a much higher throughput. The measurement of $OD_{450}$ is a lot less work and can be carried out at much higher frequency than obtaining samples to determine cell dry weight, whilst the problem of changing the volume of the culture is avoided.

Growth of all strains in liquid medium was followed and compared, semi-log plots of growth were generated and followed by calculation of the specific growth rate. This showed that the Δ*pyk2* strain had an increased specific growth rate (0.2 $h^{-1}$) when compared to WT (0.15 $h^{-1}$) and the Δ*pyk1* strain (0.15 $h^{-1}$) in YEME. The introduction of additional copies of either *pyk1* or *pyk2* to the WT strain negatively impacted on the specific growth rate of the strains, suggesting that there is a fitness cost associated with additional gene dosage (0.07 $h^{-1}$ and 0.10 $h^{-1}$ respectively). The transposon insertion mutants both showed a small decrease compared to the WT with 0.12 $h^{-1}$ and did not seem to be complemented on YEME. Furthermore the complementation of *pyk2*::Tn5062 with *pyk2* lead to a further decrease of the growth rate to 0.07 $h^{-1}$ (Figure 4.20).

Growth of the strains in minimal medium with glucose as the sole carbon source resulted in a doubling of the growth rate in both the Δ*pyk1* and Δ*pyk2* mutants compared to WT with 0.08 $h^{-1}$ compared to 0.04 $h^{-1}$. This phenotype was complemented by the addition of each gene on the integrating plasmid pIJ6902. An additional copy of *pyk1* to the WT strain led to a reduction in the specific growth rate from 0.04 $h^{-1}$ to 0.02 $h^{-1}$. The double *pyk* mutant grew very poorly with a specific growth rate of around 25 % of the WT. For both transposon insertion mutant strains the growth rate was reduced slightly. Complementation with the respective other *pyk* seemed to impair the growth rate in both cases (75% decrease for *pyk1*::Tn5062 + *pyk2* and half for *pyk2*::Tn5062 + *pyk1*). For *pyk2*::Tn5062 + *pyk2* the growth was also worse (about 50 % decrease) than for the mutant and the WT (Figure 4.21).

Tween is a fatty acid (palmitate) based carbon source that delivers TCA cycle intermediates without the requirement for Pyk activity. The growth was very poor on this carbon source with this cultivation system. On minimal medium with Tween40 as the sole carbon source, the Δ*pyk1* mutant had a specific growth rate of 0.015 $h^{-1}$ compared to the WT with 0.03. This approximately 50 % reduction in specific growth rate was only partially complemented by the plasmid pJH08 (pIJ6902 + *pyk1*). The Δ*pyk2* mutant did not have a decreased growth rate. However addition of the plasmid pJH07 negatively affected the growth rate. An additional copy of *pyk1* in the WT strain on this medium also led to a decrease of more than half of the specific growth rate of the WT, whereas an additional copy of *pyk2* led to two-fold increase in specific growth rate. The Δ*pyk1 pyk2::Tn5062* double mutant had an approximate two third decrease in specific growth rate on Tween40. Both transposon mutants showed an increased growth rate compared to the WT - in both cases it was almost doubled. The *pyk1*::Tn5062 + *pyk1* strain showed an even higher growth rate, more than doubled, whereas the complementation with *pyk2* showed very poor growth with approximately 90% decrease in its specific growth rate compared to the WT. The complementation of *pyk2*::Tn5062 worked when complemented with *pyk2*, but not *pyk1*. The latter showed again about 90% decrease in the specific growth rate whereas the former almost identical with the WT (Figure 4.22).

Since a range of different media and strains were tested for the pyruvate kinase the Table 4-1 summarises the results for each condition tested to give a better overview.

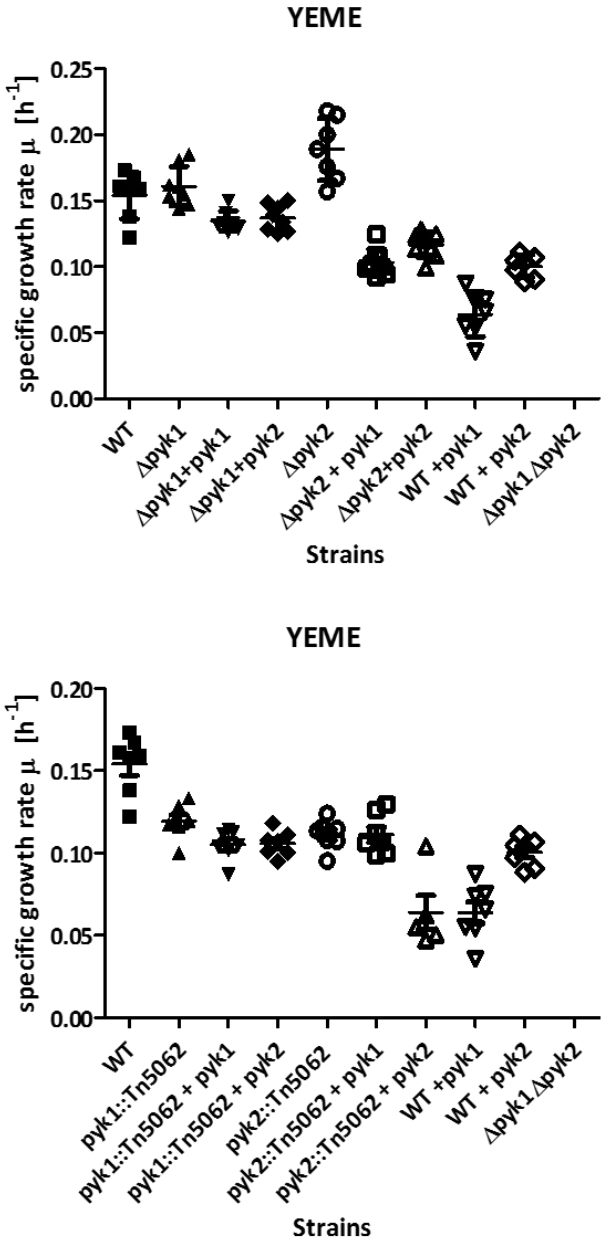



Figure 4.20 Specific growth rate comparison of pyruvate deletion mutants and transposon insertion mutants grown in YEME. Growth was followed by 15 min interval measurements of $OD_{450}$ over 48 h

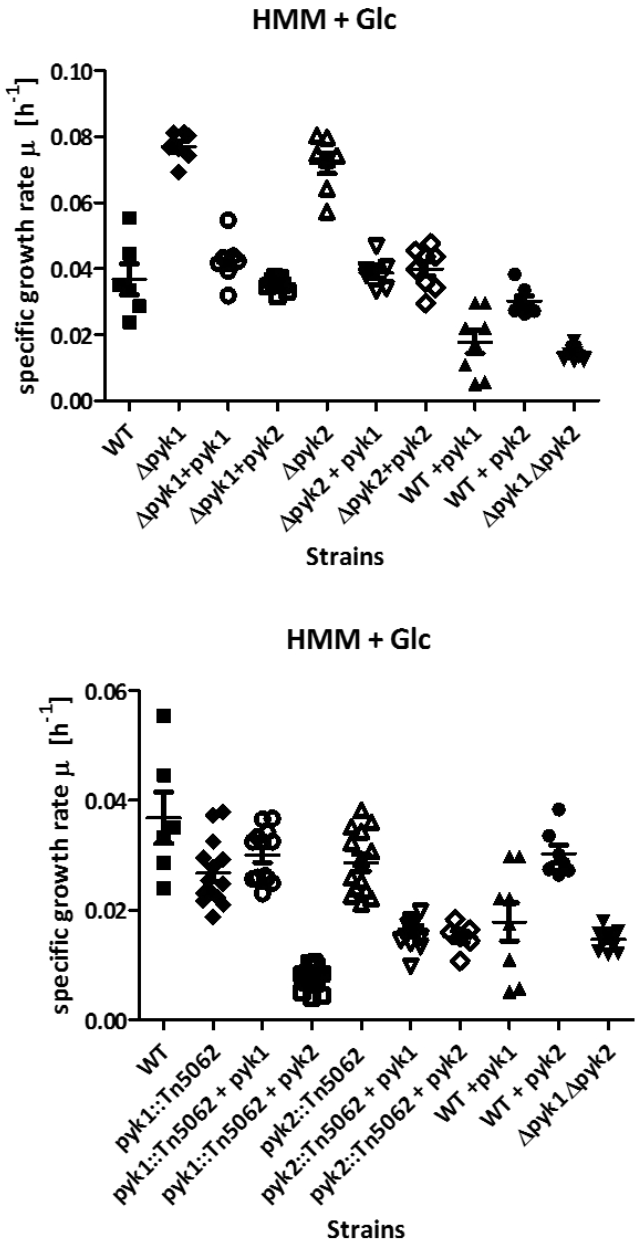



Figure 4.21 Specific growth rate comparison of pyruvate deletion mutants and transposon insertion mutants grown in Hobbs minimal medium supplemented with glucose. Growth was followed by 15 min interval measurements of $OD_{450}$ over 48 h

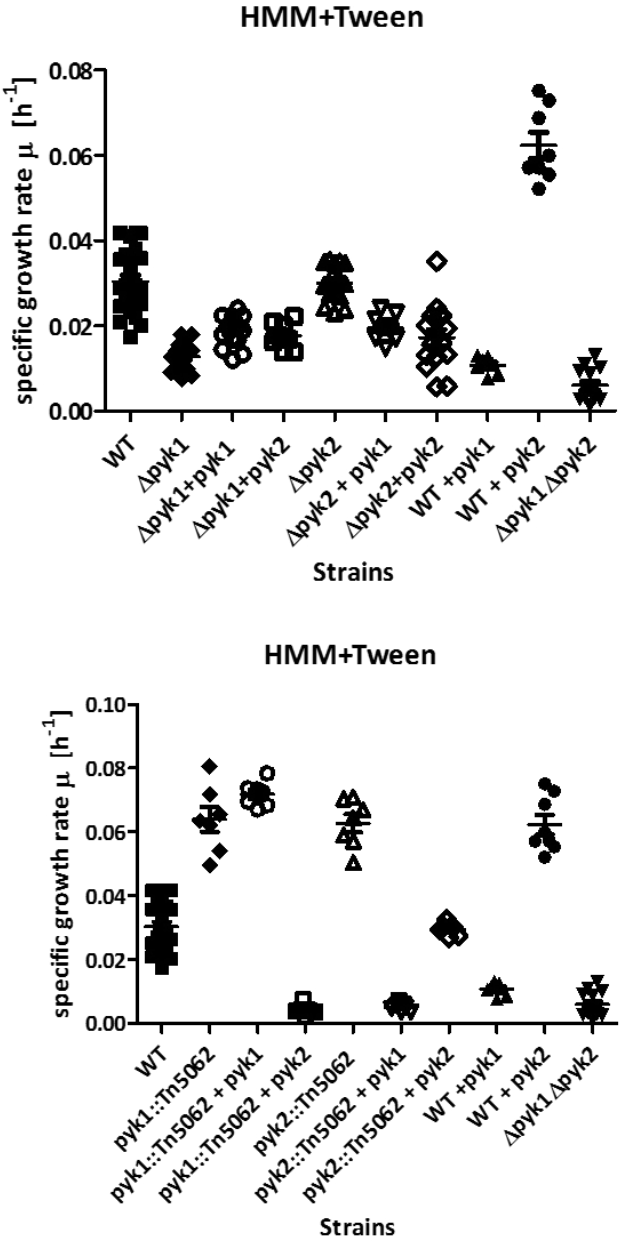



Figure 4.22 Specific growth rate comparison of pyruvate deletion mutants and transposon insertion mutants grown in Hobbs minimal medium supplemented with Tween 40. Growth was followed by 15 min interval measurements of $OD_{450}$ over 48 h

Table 4-1 Summary of phenotypes for pyruvate kinase mutants on different media specifying cultivation conditions, carbon and nitrogen source

| | | | | compared to WT | | |
|---|---|---|---|---|---|---|
| **Medium** | **C-source** | **N-source** | **Conditions** | **Growth** | **Production of spec metabolites** | **Strain** |
| SMM | Glc | casein | 1.5 mL in 24 well plate glass beads ,6 days | not analysed | endpoint: two fold ACT<br>endpoint: two fold ACT | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| YEME | Glc<br>malt | tryptone<br>peptone | 400 ml, metal springs<br>48 hs | Identical | endpoint: +ACT +CPK<br>identical | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| SMM | Pyr | casein | 400 ml, metal springs<br>48 hs | Identical | endpoint: +ACT +CPK<br>identical | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| YEME | Glc<br>Malt | tryptone<br>peptone | 200 µl in 96 well plate 48hs | bit lower/ identical<br>bit lower/ identical<br>bit lower/ identical<br>bit lower/bit higher<br>bit lower/ bit lower<br>lower/ bit lower<br>Lower<br>bit lower | not determined | *pyk1*::Tn5062*/Δpyk1*<br>*pyk1*::Tn5062 + *pyk1/Δpyk1* + *pyk1*<br>*pyk1*::Tn5062 + *pyk2/Δpyk1* + *pyk2*<br>*pyk2*::Tn5062*/Δpyk2*<br>*pyk2*::Tn5062 + *pyk1/Δpyk2* + *pyk1*<br>*pyk2*::Tn5062+ *pyk2/Δpyk2* + *pyk2*<br>WT + *pyk1*<br>WT + *pyk2* |
| HMM | Glc | $NaNO_3$ | 200 µl in 96 well plate 48hs | bit lower or = /higher<br>identical/ identical<br>much lower/ identical<br>bit lower or = /higher<br>lower/ identical<br>lower/identical<br>bit lower<br>Identical<br>25 % of WT | not determined | *pyk1*::Tn5062*/Δpyk1*<br>*pyk1*::Tn5062 + *pyk1/Δpyk1* + *pyk1*<br>*pyk1*::Tn5062 + *pyk2/Δpyk1* + *pyk2*<br>*pyk2*::Tn5062*/Δpyk2*<br>*pyk2*::Tn5062 + *pyk1/Δpyk2* + *pyk1*<br>*pyk2*::Tn5062 + *pyk2/Δpyk2* + *pyk2*<br>WT + *pyk1*<br>WT + *pyk2*<br>*Δpyk1 pyk2*::Tn5062 |

| Medium | C-source | N-source | Conditions | compared to WT<br>Growth | Production of spec metabolites | Strain |
|---|---|---|---|---|---|---|
| HMM | Tween40 | $NaNO_3$ | 200 µl in 96 well plate 48hs | higher/ lower | not determined | *pyk1*::Tn5062*/Δpyk1* |
| | | | | lower / identical or bit lower | | *pyk1*::Tn5062 + *pyk1/Δpyk1* + *pyk1* |
| | | | | higher/ identical or bit lower | | *pyk1*::Tn5062 + *pyk2/Δpyk1* + *pyk2* |
| | | | | higher/ identical | | *pyk2*::Tn5062*/Δpyk2* |
| | | | | lower/ lower | | *pyk2*::Tn5062 + *pyk1/Δpyk2* + *pyk1* |
| | | | | identical/ lower | | *pyk2*::Tn5062 *+ pyk2/Δpyk2* + *pyk2* |
| | | | | Lower | | WT + *pyk1* |
| | | | | Higher | | WT + *pyk2* |
| | | | | 50% of WT | | *Δpyk1 pyk2*::Tn5062 |
| YEME | malt<br>Glc, agar | peptone<br>tryptone | solid, 7 days | increased sporulation<br>Identical | more ACT<br>identical | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| MM | Citrate | L-asparagine | solid , 7 days | Identical<br>Identical | more ACT<br>identical | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| MM | Malate | L-asparagine | solid, 7 days | Identical<br>Identical | more ACT<br>more undecylprodigiosin | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| MM | Pyr | L-asparagine | solid , 7 days | Identical<br>Identical | more ACT<br>more undecylprodigiosin | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| MM | GlcNAc | L-asparagine | solid , 7 days | Identical<br>identical | more ACT<br>more undecylprodigiosin | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |
| MM | Glc | $NaNO_3$ | solid , 7 days | identical<br>identical | more ACT and undecylprodigiosin<br>identical | *pyk1*::Tn5062<br>*pyk2*::Tn5062 |

| Medium | C-source | N-source | Conditions | compared to WT<br>Growth | Production of spec metabolites | Strain |
|---|---|---|---|---|---|---|
| NA | Glc | | solid , 7 days | identical | more ACT | *Δpyk1* |
| | yeast extract | | | identical | identical | *Δpyk1* + *pyk1* |
| | tryptone | | | identical | identical | *Δpyk1* + *pyk2* |
| | peptone | | | identical | a bit more ACT | *Δpyk2* |
| | | | | colonies look hairier | identical | *Δpyk2* + *pyk1* |
| | | | | identical | identical | *Δpyk2* + *pyk2* |
| | | | | colonies are bald | less ACT | WT + *pyk1* |
| | | | | identical | more ACT | WT + *pyk2* |
| MM | Glc | $NH_4^+$ | solid , 7 days | colonies are bald | identical | *Δpyk1* |
| | | | | identical | identical | *Δpyk1* + *pyk1* |
| | | | | identical | identical | *Δpyk1* + *pyk2* |
| | | | | smaller colony size | identical | *Δpyk2* |
| | | | | identical | identical | *Δpyk2* + *pyk1* |
| | | | | identical | identical | *Δpyk2* + *pyk2* |
| | | | | identical | more undecylprodigiosin | WT + *pyk1* |
| | | | | identical | more undecylprodigiosin | WT + *pyk2* |
| MM | Glc | $NH_4^+$ | solid , 7 days | identical | identical | *pyk1*::Tn5062 |
| | | | | identical | less undecylprodigiosin | *pyk1*::Tn5062 + *pyk1* |
| | | | | identical | identical | *pyk1*::Tn5062 + *pyk2* |
| | | | | identical | more undecylprodigiosin | *pyk2::Tn5062* |
| | | | | identical | identical | *pyk2*::Tn5062 + *pyk1* |
| | | | | identical | identical | *pyk2*::Tn5062 + *pyk2* |
| | | | | identical | less undecylprodigiosin | WT + *pyk1* |
| | | | | identical | identical | WT + *pyk2* |

| | | | | compared to WT | | |
|---|---|---|---|---|---|---|
| **Medium** | **C-source** | **N-source** | **Conditions** | **Growth** | **Production of spec metabolites** | **Strain** |
| NA | Glc | | solid , 7 days | more aerial hyphae | more ACT | *pyk1*::Tn5062 |
| | yeast extract | | | more aerial hyphae | more ACT | *pyk1*::Tn5062 + *pyk1* |
| | tryptone | | | pale colony | no specialised metabolite production | *pyk1*::Tn5062 + *pyk2* |
| | peptone | | | identical | identical | *pyk2*::Tn5062 |
| | | | | identical | identical | *pyk2*::Tn5062 + *pyk1* |
| | | | | pale colony | no specialised metabolite production | *pyk2*::Tn5062 + *pyk2* |
| | | | | identical | identical | WT + *pyk1* |
| | | | | more aerial hyphae | more ACT | WT + *pyk2* |
| MM | Pyr | $NH_4^+$ | solid , 7 days | identical | identical | *pyk1*::Tn5062 |
| | | | | identical | identical | *pyk1*::Tn5062 + *pyk1* |
| | | | | identical | identical | *pyk1*::Tn5062 + *pyk2* |
| | | | | identical | less ACT | *pyk2*::Tn5062 |
| | | | | identical | identical | *pyk2*::Tn5062 + *pyk1* |
| | | | | identical | identical | *pyk2*::Tn5062 + *pyk2* |
| | | | | identical | less ACT | WT + *pyk1* |
| | | | | identical | identical | WT + *pyk2* |

## 4.9 SUMMARY

The mutational analysis of genes within the PEP-PYR-OAA node of primary metabolism was undertaken to identify genes that affect polyketide antibiotic production. Primary metabolism needs to be robust, so a gene knockout may not impair the overall cellular function and phenotypes may be hard to observe and define clearly. Additionally there is a high variability in *Streptomyces* growth due to their lifestyle and heterogeneity in the colonies on solid medium or pellets in liquid medium (Nieminen *et al.*, 2013).

However this can be overcome by screening strains in a variety of different media and cultivation systems. Testing the mutants created in this study, it was observed that several members of the PEP-PYR-OAA metabolic node influence specialised metabolite production in *S. coelicolor*. Those were $PDHC_{E1}$ (*aceE3*), both pyruvate kinases (*pyk1* and *pyk2*), pyruvate dehydrogenase (*poxB2*) and one of the malic enzymes (*meNAD*). In combination with the expansion and phylogenetic analysis conducted in Chapter 3 and published data in the literature for the malic enzymes (Rodríguez *et al.*, 2012), the focus of the thesis was on the genes encoding three enzymes; E1 subunit of pyruvate dehydrogenase complex, and the paralogous *pyk1* and *pyk2*.

Studies of the $PDHC_{E1}$ gene disruption indicated a potentially interesting phenotype. However genetic complementation was unsuccessful. Nevertheless this is an interesting enzyme to study, due to the occurrence of multiple genes for this reaction and because it results in extensive actinorhodin overproduction. Due to time and work constraints this work was not pursued further.

Studies of the paralogous pyruvate kinase indicate that the two pyruvate kinases have different effects on the physiology of *S. coelicolor*. The data relating to these genes is robust, given that mutations were created using two different method and both genes can be complemented. The two pyruvate kinases show distinct phenotypes under a range of growth conditions.

The disruption or deletion of *pyk1* led to an increased specialised metabolite production under several conditions, which were mainly gluconeogeneic (YEME, pyruvate, glucose with L-asparagine, citrate, malate, GlcNAc + L-asparagine). However growth on glucose with $NaNO_3$ as nitrogen source showed an increase in actinorhodin and undecylprodigiosin. Disruption or deletion of *pyk2* showed less change in respect to specialised metabolite

production. Increased actinorhodin production was only observed at small scale liquid cultivation on SMM with glucose and on nutrient agar. However undecylprodigiosin production was increased on several solid media (malate, pyruvate and GlcNAc) when using L-asparagine as nitrogen source. The role of paralogues in providing robustness in genetic systems is becoming well-established in bacterial systems (Vitkup *et al.*, 2006; Wagner, 2008; Treangen and Rocha, 2011) and is providing key insights into understanding the ecology and evolution of important industrial organisms that can be exploited for metabolic engineering need, increasing production yields.

# 5 BIOCHEMICAL AND FUNCTIONAL CHARACTERISATION OF THE PARALOGOUS PYRUVATE KINASES OF *S. COELICOLOR*

This chapter is focussed on the biochemical and functional characterisation of the two paralogous pyruvate kinases of *S. coelicolor*. In Chapter 3, it was shown that expansions of specific primary metabolic nodes of metabolism correlate with the capacity of an organism to produce specialised metabolites and that the PEP-PYR-OAA node of central carbon metabolism was particularly rich in expansion events. Further genetic characterisation of this node of metabolism indicated that the duplication event observed for pyruvate kinases in *S. coelicolor* indicates that these isoenzymes play a role in channelling carbon to specialised metabolism as demonstrated by the phenotypes observed for the two mutants; *pyk1* increasing antibiotic production and *pyk2* having little or no effect on antibiotic production. These data suggest that following duplication, the pyruvate kinases have evolved different roles in *Streptomyces*.

## 5.1 CROSS-SPECIES COMPLEMENTATION IN *E. COLI* PYRUVATE KINASE MUTANTS

The two pyruvate kinases from *E. coli*, encoded by the genes *pyk*A and *pyk*F, have been studied in detail - PykF represents the primary pyruvate kinase and is allosterically regulated by fructose 1-6 biphosphate (FBP), whereas PykA plays a minor role in *E. coli* physiology and is regulated by adenosine monophosphate (AMP, Ponce *et al.*, 1995; Valentini *et al.*, 2000). To develop a deeper understanding of the paralogous pyruvate kinases in *S. coelicolor* cross-species complementation experiments were undertaken. The two pyruvate kinase mutants of *E. coli* (Keio collection; Baba *et al.*, 2006) were complemented, in turn with one of the two pyruvate kinase genes from *S. coelicolor* on a plasmid (pJH01 and pJH02, see section 5.2 for plasmid construction). Moreover, a pyruvate kinase double mutant was constructed in *E. coli* and complemented, again with each of the pyruvate kinases from *S. coelicolor*.

The single mutants, Δ*pykA* and Δ*pykF* were obtained from the Keio collection, which is a collection of *E. coli* single gene deletion mutants of all non-essential genes in the strain K-12 BW25113. Each gene was replaced with a kanamycin resistance cassette flanked by FLP sites (Baba *et al.*, 2006)

A double pyruvate kinase mutant Δ*pykA*Δ*pykF* was constructed using the ReDirect protocol for *Streptomyces* with pIJ773 to replace the gene (Gust *et al.*, 2002). Primers were designed

which included 39 nt upstream of *pykF* ending with the start codon of *pykF* plus the first 20 nucleotides of the insertion cassette from pIJ773 as the forward primer and the reverse primer was composed of 36 nt downstream of *pykF* in the genome of *E. coli* and the forward primer last 19 nt of the insertion cassette on pIJ773 and the stop codon of *pykF*. The sequence of pIJ773 was obtained from the ReDirect Manual. Those primers were then utilised in a PCR reaction using Q5 polymerase (NEB) with pIJ773 as template (3 ng/µl) and was carried out as specified in the ReDirect manual, to yield the replacement cassette for *pykF*. The obtained PCR product was then used to transform competent cells of the Δ*pykA* mutant carrying pIJ790 (containing λ red genes) and the colonies were selected on 50 µg/ml apramycin and 50 µg/ml kanamycin (Figure 5.1).

Both single and double *pyk* mutant strains were confirmed by PCR using Taq polymerase (MyTaq, Bioline) with three different PCR reactions, one with primers for the presence of the kanamycin resistance or apramycin cassette. However since this does not verify the correct position of the insertion, two further PCR reactions were carried out with one primer sitting upstream the respective *pyk* gene and one inside the resistance marker as well as one primer downstream the gene and one inside the resistance marker to verify the correct position in the genome. A schematic overview of the different strains and where primers were located can be found in Figure 5.2.

For the presence of the kanamycin resistance a PCR product of 581 bp was expected. For the primer positioned upstream of *pykA* combined with the reverse primer of the kanamycin cassette an amplicon of 969 bp was expected. For the primer pair with the forward primer in the kanamycin cassette and the reverse primer downstream of *pykA,* a size of 934 bp was expected. For the primer located upstream of *pykF* combined with the reverse primer of the kanamycin cassette an amplicon of 989 bp was expected. The primer pair with the forward primer in the kanamycin cassette and the reverse primer downstream of *pykF* would generate an amplicon of 935 bp. These reactions were used for verification of the single pyruvate kinase mutants Δ*pykA* and Δ*pykF* (Figure 5.3).

For the double pyruvate kinase mutant Δ*pykA*Δ*pykF*, the same reactions as for Δ*pykA* were run, but additionally a reaction to verify the presence of apramycin resistance with an anticipated amplicon size of 506 bp. Additionally two more reactions were performed with a forward primer upstream *pykF* in combination with a reverse primer located inside the apramycin cassette with an expected amplicon size of 919 bp. The primer pair with the

forward primer in the apramycin cassette and the reverse primer downstream of *pykF* were expected to generate an amplicon of 1113 bp. All strains could be verified as shown in Figure 5.3.

To establish the physiological role of the *S. coelicolor* pyruvate kinases in an *E. coli* genetic background, all mutants and the complemented strains were grown on LB, M9 plus glucose or M9 plus acetate to obtain detailed information on the effects these genes have under a range of physiological conditions. On LB no significant difference was observed among any of the strains (Figure 5.4).

This is likely due to growth on LB not proceeding through glycolysis. On M9 with acetate as the sole carbon source, the *E. coli ΔpykF* mutant exhibited a decrease in specific growth rate when compared to the isogenic parent strain (*E. coli* BW25113) and this phenotype could be restored with either *pyk1* or *pyk2* from *S. coelicolor* (Figure 5.4).

The greatest differences between strains was when grown on glucose as the sole carbon source; the *E. coli ΔpykAΔpykF* double mutant was unable to grow at all. However when genetically complemented with either *pyk1* or *pyk2* from *S. coelicolor,* growth was partially restored. In *E. coli* the *ΔpykA* and the Δ*pykF* mutants exhibited a specific growth rate of around 50% of the isogenic parental strain (BW25113). Interestingly, Δ*pykF* could be genetically complemented with either the *pyk1* or *pyk*2 gene from *S. coelicolor*. However, the Δ*pykA* mutant of *E. coli* could only be complemented with *pyk1* suggesting a much narrower physiological role for *pykA* in *E. coli* (Figure 5.4). Moreover, complementation of *pykF* by either *pyk1* or *pyk*2 adds additional weight to the phylogeny from Chapter 3. In addition, whilst these data do not provide details on the fine physiological control of *pyk1* or *pyk*2 in *S. coelicolor*, it confirms that both genes encode enzymes with pyruvate kinases, since both were able to restore growth on minimal medium with glucose to an *E. coli* Δ*pykA*Δ*pykF* mutant.

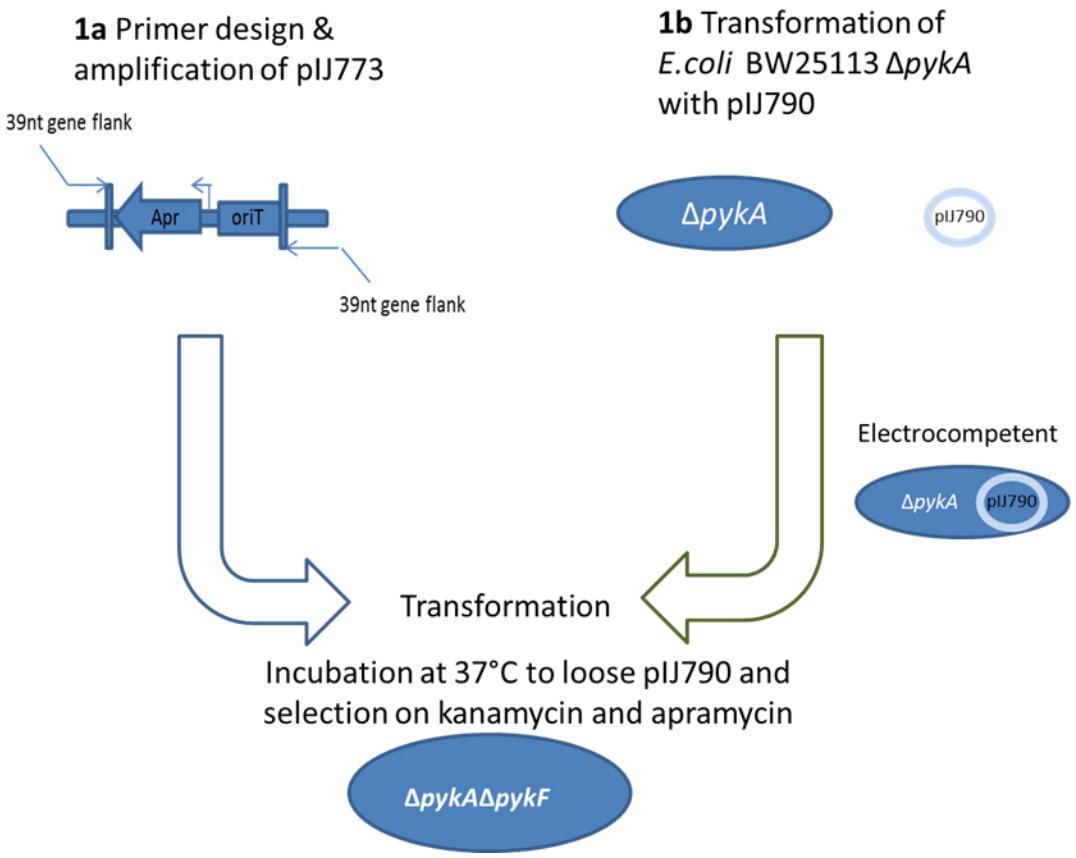



Figure 5.1 Schematic overview of construction of the double Δ*pykA*Δ*pykF* mutant. Primers were designed to amplify the apramycin cassette from pIJ773 with primers composed of gene flanking regions and inside the cassette. The PCR product was then transformed into BW25113 Δ*pykA* carrying pIJ790. Double mutant colonies were selected on apramycin and kanamycin resistance at 37°C to loose pIJ790. Modified from ReDirect Manual (Gust *et al.*, 2002)

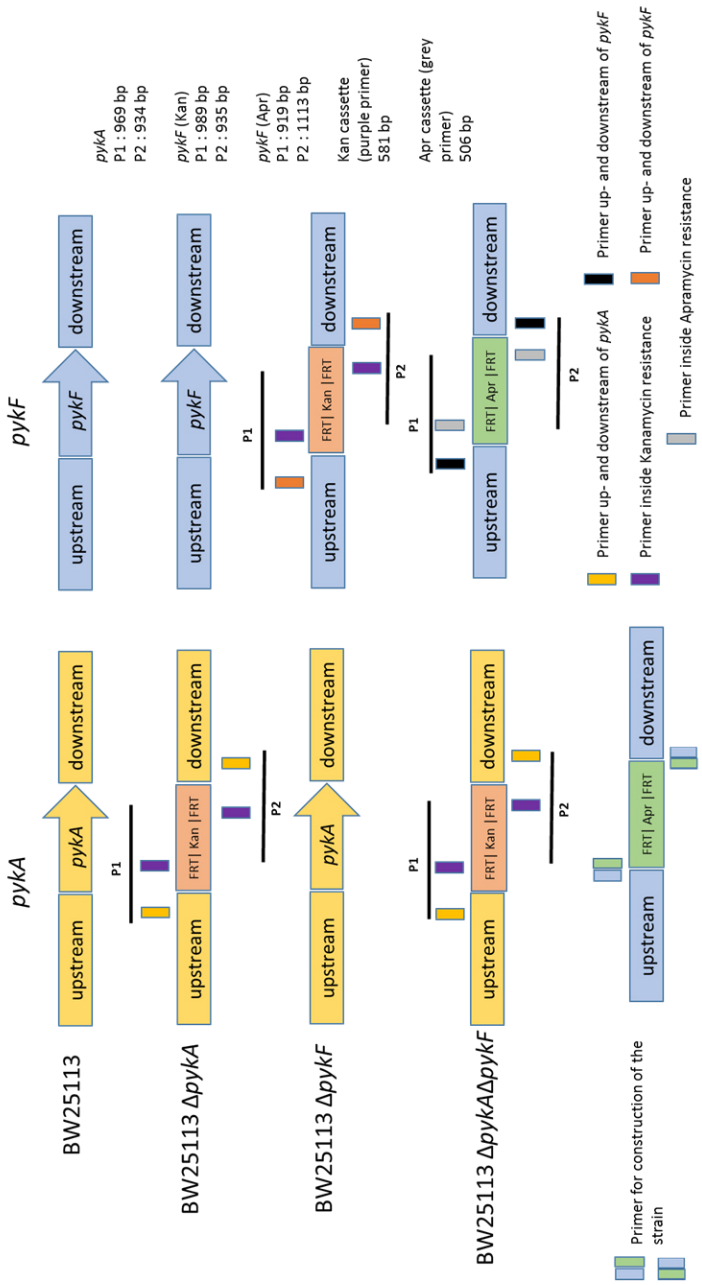



Figure 5.2 Schematic overview of pyruvate kinase mutants in *E. coli* with respective primer binding sites for the verification of each strain

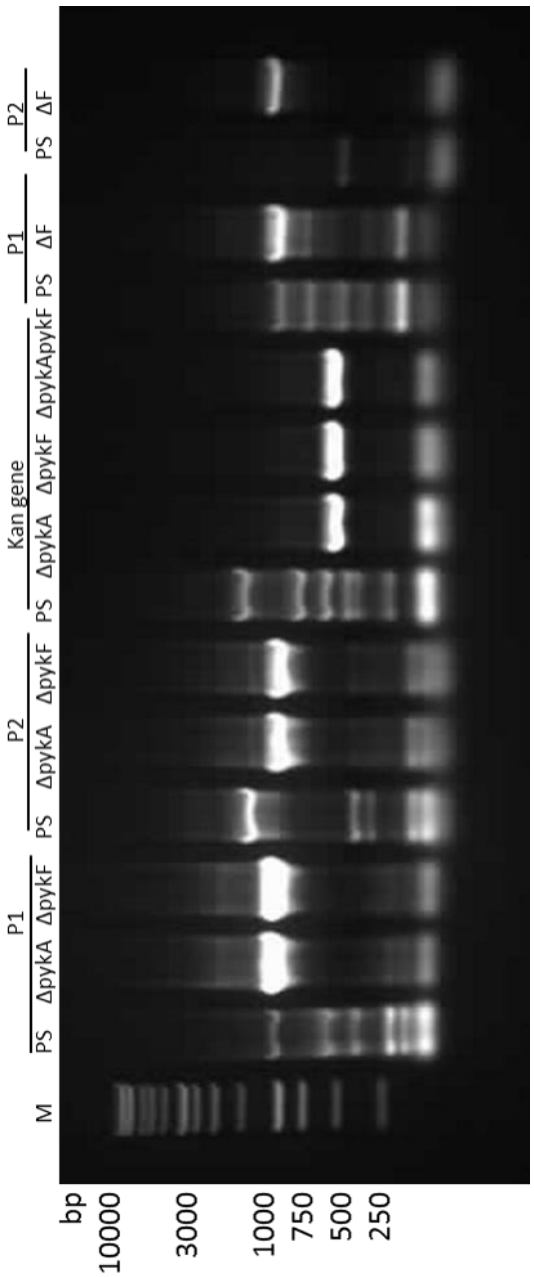



Figure 5.3 Verification of pyruvate kinase mutants in *E. coli* by PCR.Expected amplicon sizes for each mutant: Δ*pykA* P1 969 bp, P2 934 bp, Δ*pykF* (Kan) P1 989 bp, P2 935 bp, Δ*pykF* (Apr) P1 919 bp, P2 1113 bp, Kan cassette: 581 bp, Apr cassette 506 bp

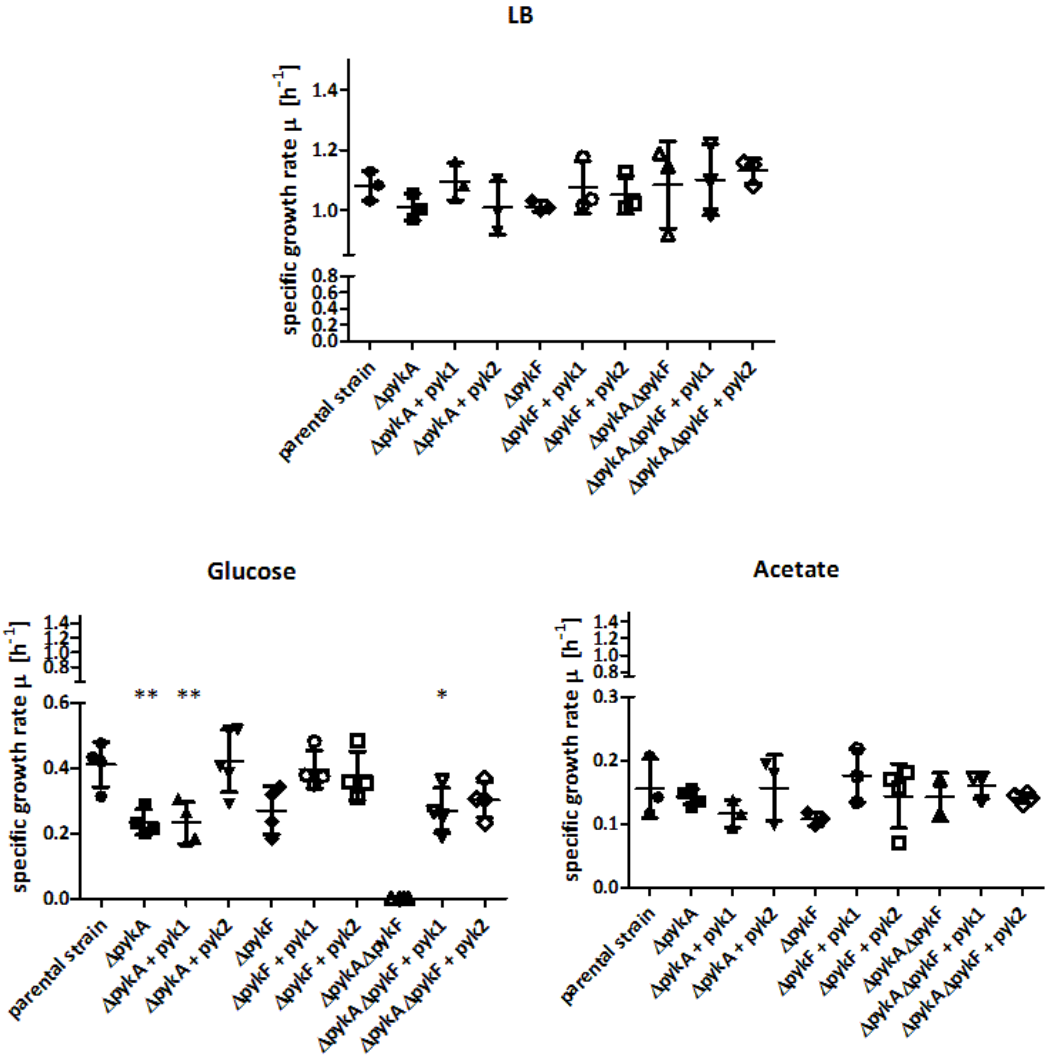



Figure 5.4 Specific growth rates of *E. coli* pyruvate kinase mutants and complemented strains on LB, M9 plus either(1 %) glucose or (0.4 %) acetate. Legend: statistical analysis by 1way ANOVA * p- value ≤ 0.05 ** p-value ≤ 0.01

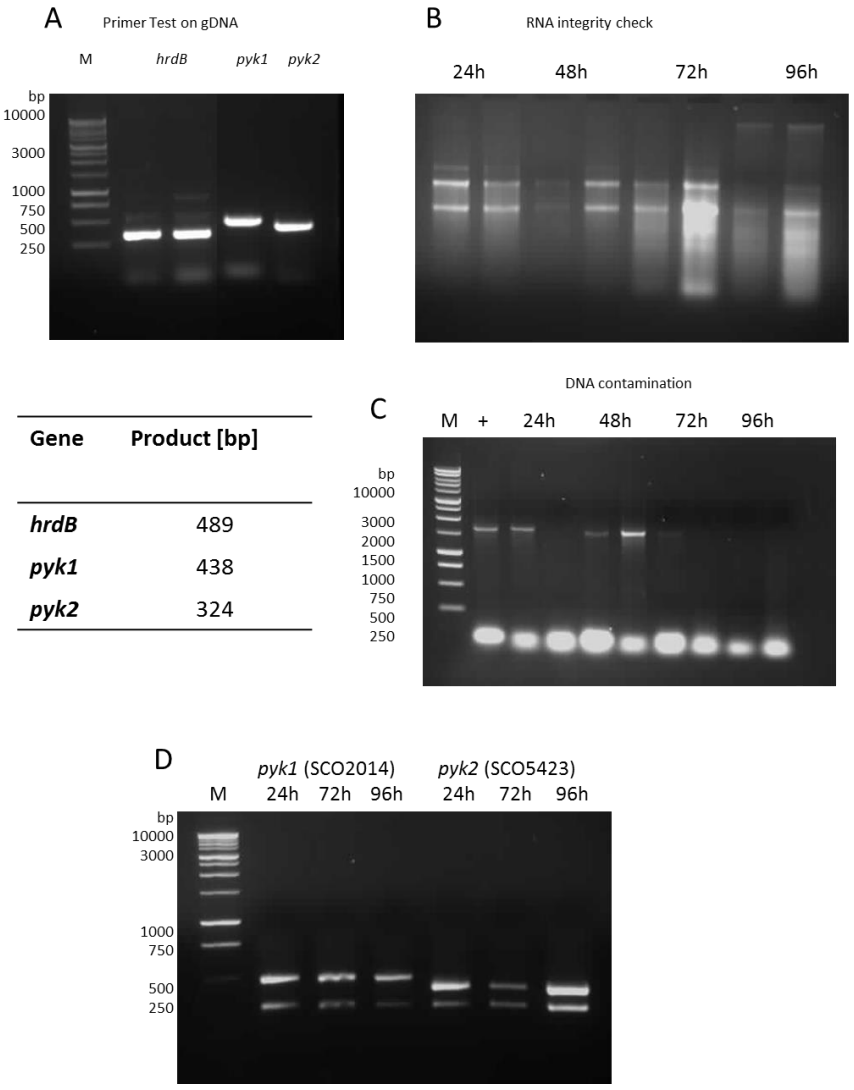
## 5.2 Semi-Quantitative Gene Expression Analysis of Pyruvate Kinase

In addition to horizontal gene transfer, gene duplication is another way of neo-functionalisation. The presence of two pyruvate kinase can be attributed to gene duplication as described in Chapter 3. By the cross-complementation in *E. coli* pyruvate kinase mutants with the pyruvate kinases from *S. coelicolor*, it could be established that both Pyk1 and Pyk2 indeed show pyruvate kinase activity as they both restore growth on glucose in the double mutant Δ*pykA*Δ*pykF.* The two transcriptional regulators *devA* and *devE* in *S. coelicolor* are an example where the duplicated copy diverged after the duplication and now has a different role. This is observed in different expression profile of the two at different stages of the life cycle. The gene duplication of *devA* most likely has led to an increased morphological complexity in streptomycetes (Clark and Hoskisson, 2011).

To determine if the two parologous pyruvate kinases of *S. coelicolor* show a difference in their expression profile, such as at different times during growth, total RNA was isolated from mycelium grown on MS agar plates overlaid with sterile cellophane disks at four different time points of 24, 48, 72 and 96 h. Two samples were taken at each time point.

The primers for RT-PCR of *hrdB*, *pyk1* and *pyk*2 are given in Table 2-7. The conditions for amplification were optimised using genomic DNA as described in 2.7.6. All primer pairs gave a single amplicon when analysed by gel electrophoresis at the expected sizes of 324, 489 and 438 bp respectively (Figure 5.5). Cultures grown on cellophane were harvested and RNA was isolated as described in section 2.10 and was treated with DNaseI (Promega, see section 2.11). The amount of total RNA was quantified by NanoDrop (Thermo Scientific) and its integrity was analysed by gel electrophoresis. This revealed the expected distinct bands for rRNA and a less distinct 'smear' that is mRNA (Figure 5.5). The gel showed that one of the 48 h and one of the 96 h samples were likely degraded due to the lack of clear rRNA bands (Figure 5.5). Each of the samples was tested for DNA contamination by a PCR using Taq polymerase (MyTaq, Bioline). One of the samples for each time points at 24, 48 and 72 h exhibited DNA contamination (Figure 5.5). The DNA contaminated samples were subjected to a second DNaseI digest to remove remaining DNA in the sample to prevent the detection of false positives. However after retreatment with DNase, the sample of 48 h did not show any nucleic acid and this sample was excluded from further analysis. Finally, 50 ng of RNA were used as template in a one-step RT-PCR Kit (Qiagen), multiplexed using the *hrdB* primer pair and each gene of interest. The *hrdB* sequence was used as an expression control as this

gene encodes the vegetative sigma factor, which shows a constitutive expression pattern throughout the different growth stages. Semi-quantative RT-PCR of the two pyruvate kinase genes demonstrated that both are expressed throughout growth of *S. coelicolor* (Figure 5.5). No expression difference could be detected between the two pyruvate kinases using this semi-quantitative approach. However this method might not have been sensitive enough and even though both seem to be expressed at all times, they still might differ in the amount. To further characterise any differences in expression, RNA-Seq was performed to analyse the expression levels across central carbon metabolism on two different carbon sources (Chapter 6).




| Gene | Product [bp] |
|---|---|
| ***hrdB*** | 489 |
| ***pyk1*** | 438 |
| ***pyk2*** | 324 |

Figure 5.5 Semi-quantitative expression analysis of *pyk1* and *pyk2, hrdB* was chosen as reference gene. (A) Primer specificity test on genomic DNA of *S. coelicolor* (B) RNA integrity check on each of the duplicate samples on 2% agarose gel (C) Conventional PCR of each of the samples to check for DNA contamination. In the positive control genomic DNA was used as a template. (D) One step RT-PCR amplicons on 2% agarose gel of each sample.

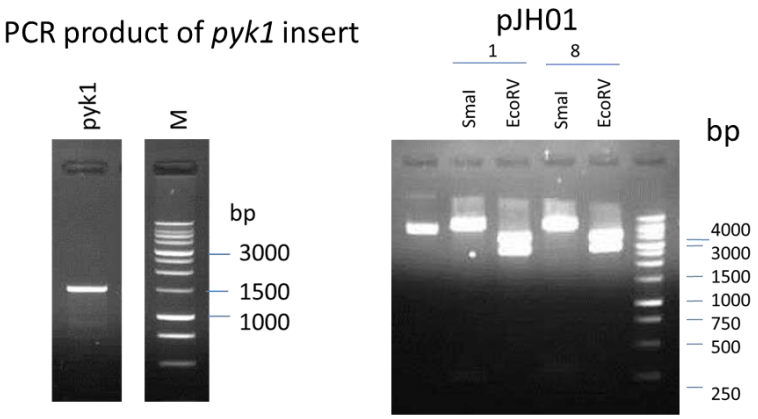
## 5.3 Overexpression of Pyk1 and Pyk2

As neither the cross-complementation nor the expression analysis revealed much insight into the differences between the Pyk1 and Pyk2, the two genes were cloned and the enzymes overexpressed heterologously in *E. coli* in order to perform *in vitro* assays to characterise the kinetics and allosteric regulation of them.

### 5.3.1 Overexpression of Pyk1

Overexpression of Pyk1 (SCO2014) required extensive optimisation of the protocol to obtain functional enzyme. First the plasmid pJH01 was constructed carrying *pyk1* with an N-terminal His-Tag. The gene *pyk1* (SCO2014) was amplified from *S. coelicolor* genomic DNA and was ligated into the commercial plasmid pET-TOPO100 (Invitrogen). Primers were designed with CACC overhang to ensure in frame location to the RBS on the plasmid at the 5' end of the forward primer which was the first 18 nucleotides at the N-terminus of *pyk1.* The reverse primer was from the C-terminus ending with the stop codon of the respective gene (Table 2-7). The PCR product obtained with blunt ends was then ligated into pET-TOPO100 as specified by the supplier's manual. The final construct was then confirmed by restriction digest using *Sma*I and *EcoR*V (Figure 5.6). Furthermore the construct was sent for sequencing (MWG eurofins). Once confirmed the construct was used to transform *E. coli* BL21 or Rosetta to overexpress the protein. The expression of the protein can be induced by the addition of IPTG, a structural analogue of allolactose, which is the natural inducer of lac operon (Ullmann *et al.*, 1967; Sambrook and Russell, 2000).

Protein expression was tested on a range of different temperatures (15, 20 and 30°C) and with different amounts of inducer concentration (0.25, 0.5 or 1 mM). For all conditions no additional band was observed in the soluble fraction, only a band in the cell debris was observed at approximately 55.6 kDa. This is an indicator of misfolded protein forming aggregates called inclusion bodies. An example of SDS gel is shown in Figure 5.7. Since expression levels sometimes can be very low and not visible on the SDS-PAGE, purification was attempted, but no protein was obtained.




| Plasmid | Restriction enzymes (# sites) | Expected band sizes (bp) |
|---|---|---|
| pJH01 | EcoRV (2) | 4230, 2971 |
| (pET_TOPO100_SCO2014) | SmaI (2) | 6895, 306 |

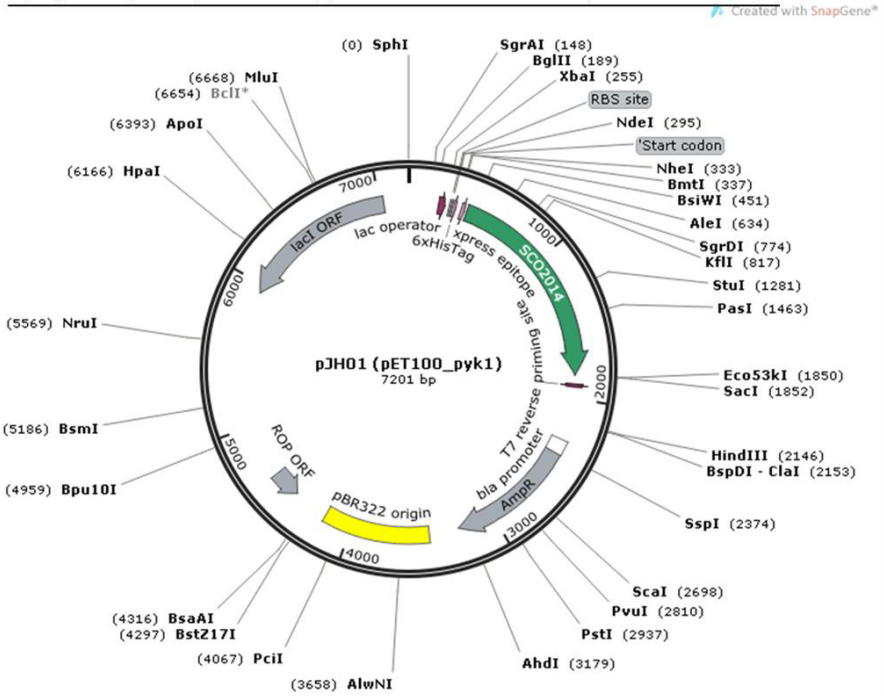



Figure 5.6 Construction and verification of pJH01 using the commercial vector pET_TOPO100 (Invitrogen). The insert was generated by PCR using a forward primer composed of the first 18 nucleotides with a CACC 5' overhang to ensure in frame location to the RBS on the plasmid. The reverse primer was from the C-terminus ending with the stop codon of the respective gene. The blunt end obtained PCR product was then ligated into pET-TOPO100 as specified by the supplier's manual. The final construct was then confirmed by restriction digest using *Sma*I and *Eco*RV

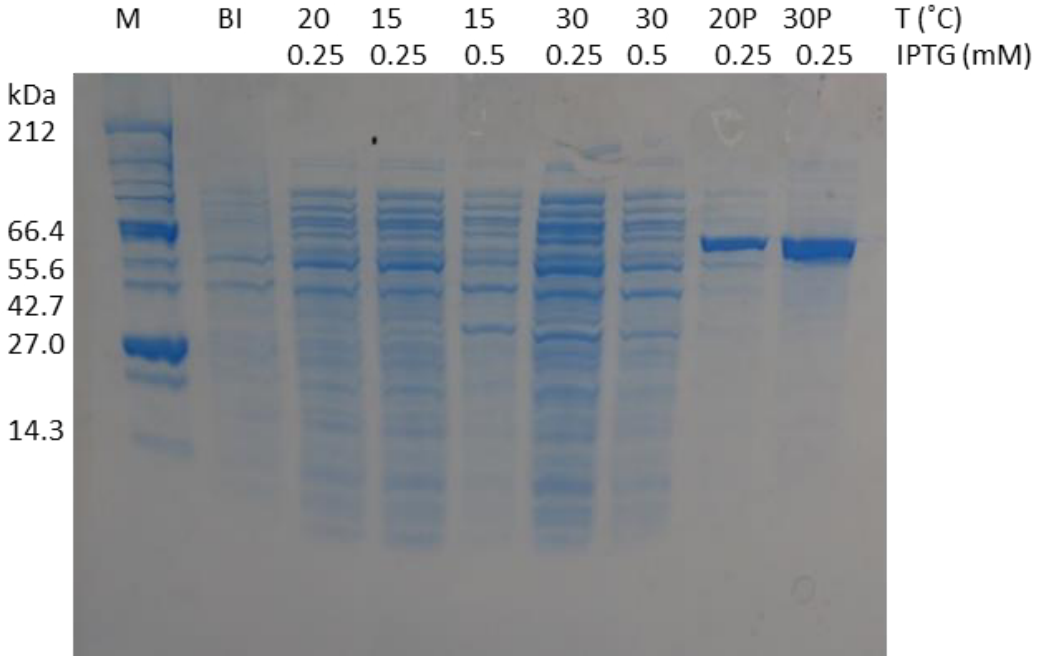
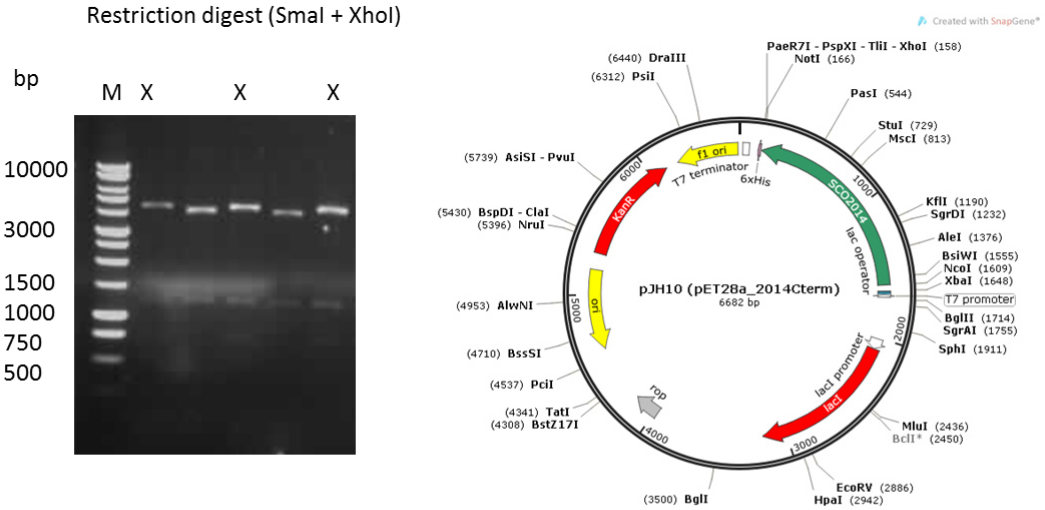



Figure 5.7 SDS-PAGE of protein extracts from *E. coli* Rosetta cultures overexpressing Pyk1 on LB showing the formation of inclusion bodies. Different inducer concentration of IPTG (0.25 mM or 0.5 mM) were tried with different temperatures (15, 20 or 30°C) showing soluble (S) and debri fractions (P). Legend: M = marker, BI= before induction, T = temperature

Because of the formation of inclusion bodies, a second plasmid (pJH10) was constructed with His-Tag at the C-terminus of the protein in case the His-Tag at the N-terminus impeded correct folding. Furthermore it was noticed that the second codon in *pyk1* gene, CGA coding for arginine is one of the least used codons in *E. coli* (0.05 ratio). In case this lead to stalling of the protein synthesis, the codon was exchanged from CGA to CGT on the primer, which is the most used codon for arginine in *E.coli* with a ratio of 0.42 (Maloy *et al.*, 1996).

To construct pJH10 containing *pyk1* with a C-terminal His-Tag, the gene was first cloned into the commercial vector pGEM (Promega). The forward primer was designed including a restriction site for *Nco*I plus two extra bases to ensure in frame location with the His-Tag. Furthermore the second codon after the start codon was changed from CGA to CGT. The reverse primer was designed to include a *Not*I site and the last four codons of the gene and the stop codon were removed as the HisTag needed to be included. The insert was generated by PCR using Taq polymerase (MyTaq, Bioline), cloned into pGEM and colonies were picked with blue white screening using 100 µg/ml carbenicillin, 0.5 mM IPTG and 80 µg/ml X-Gal. The resulting vector (pJH09) was checked by a restriction digest using *Sma*I and the insert was released from pGEM using the inserted *Nco*I and *Not*I sites. The insert was then ligated into pET28a, which was first subjected to a digestion with *Nco*I and *Not*I. The ligation was then used to transform DH5α and resulting clones were selected on kanamycin. The construct was then verified by restriction digestion with *Sma*I and *Xho*I. The construct was then used to transform *E. coli* protein overexpressing strain BL21 (Figure 5.8). The strain was grown in both LB and Auto Induction Medium (AIM) and samples were taken two and three hours after induction. Auto Induction Medium contains both glucose and α – lactose, once the glucose is used up the lactose will induce the expression of the protein from the respective plasmid (Studier, 2005). However the results were similar to the previous attempts- no protein band was observed in the soluble fraction (S), but only in the cell debris fraction (P). Since no improvement was observed, it was concluded that the position of the His-Tag on either terminus was not the cause of the inclusion bodies and the exchange of one codon did not resolve the issue. The same result was observed when using *E. coli* Rosetta for the overexpression. An example of SDS-PAGE is shown in Figure 5.9 when using BL21 at 30°C. Again no protein could be purified from any of those cultures.


Subcloning into pET28a using NcoI and NotI




| Plasmid | Restriction enzymes (# sites) | Expected band sizes (bp) |
|---|---|---|
| pJH09 (pGEM_2014cterm) | SmaI (2) | 4161, 306 |
| pJH10 (pET28a_2014cterm) | SmaI + XhoI (4) | 4425,1227,724,306 |

Figure 5.8 Construction and verification of pJH10. *pyk1* was first cloned into the commercial vector pGEM (Promega). The forward primer was designed including a *Nco*I site plus two extra bases to ensure in frame location with the His-Tag. Furthermore the second codon after the start codon was changed from CGA to CGT. The reverse primer was designed to include a *Not*I site and the last four codons of the gene and the stop codon were removed as the HisTag needed to be included. The insert was generated by PCR and was cloned into pGEM (Promega). The resulting vector (pJH09) was checked by a digestion using *Sma*I and the insert was released from pGEM (Promega) using the inserted *Nco*I and *Not*I sites. The insert was then ligated into pET28a, which was first subjected to a digestion with *Nco*I and *Not*I. The ligation was then used to transform DH5α and resulting clones were selected on kanamycin. The construct was then verified by digestion with *Sma*I and *Xho*I.

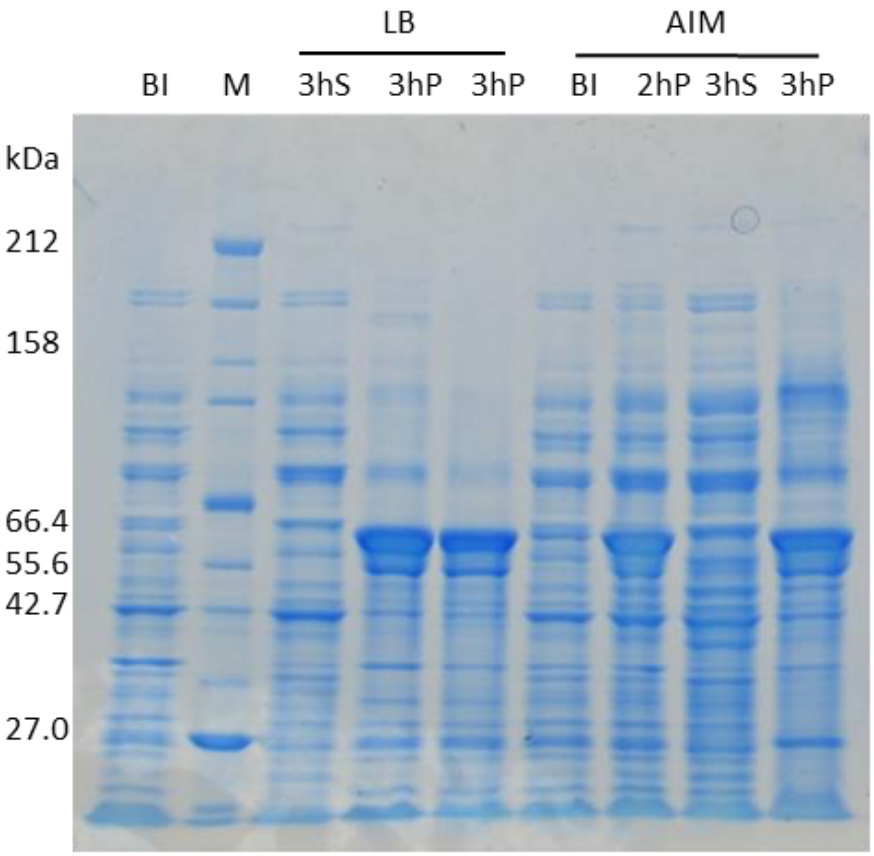
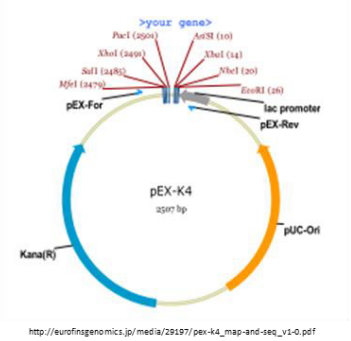



Figure 5.9 SDS-PAGE of protein extracts from *E. coli* BL21 cultures overexpressing Pyk1 on LB and AIM at 30°C showing the formation of inclusion bodies. Two time points after induction showing soluble (S) and debris fractions (P). Legend: M = marker, BI= before induction

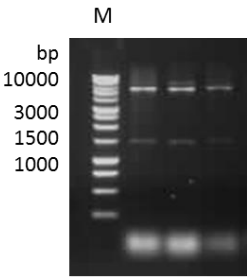
Since previous strategies proved unsuccessful, it was decided to use a synthetic version that has been codon optimised for *E. coli* as well as the addition of restriction sites *Nde*I at the start of the gene and *EcoR*I at the end of the gene. For this purpose the plasmid pJH11 was constructed using a codon-optimised for *E. coli pyk1* sequence. This was synthesized at MWG eurofins and was delivered in the vector pEX-K4. The insert was released using *Nde*I and *EcoR*I sites and was then sub-cloned into pET15b. pET15b was subjected to digestion with *Nde*I and *EcoR*I as well as dephosphorylation using alkaline phosphatase (Promega) prior to ligation. The ligation was then used to transform DH5α and clones were selected on 100 µg/ml carbenicillin. The construct was verified by restriction digestion using *Nde*I and *Eco*RI (Figure 5.10). The construct was then used to transform *E. coli* Rosetta or Origami B. The cells were grown at 30˚C in LB spiked with 1% glucose and were induced at $OD_{600}$ 0.5 and induced with 0.05 mM IPTG and the temperature was decreased to 20 ˚C and left overnight. This approach lead to a successful overexpression of Pyk1. On SDS-PAGE Pyk1 ran higher at approximately 55.6 kDa than its expected size of 51.4 kDa (Figure 5.11).

The combination of low temperature, low inducer concentration and codon optimised gene were the optimal conditions for the overexpression of Pyk1 under the conditions tested.




Codon optimised for *E.coli pyk1* gene was synthesided at MWG eurofins

The insert was subcloned from pEX-K4 using NdeI and EcoRI site into pET15b




| Plasmid | Restriction enzymes (# sites) | Expected band sizes (bp) |
|---|---|---|
| pJH11 (pET15b_SCO2014s) | NdeI + EcoRI | 5375, 1443 |

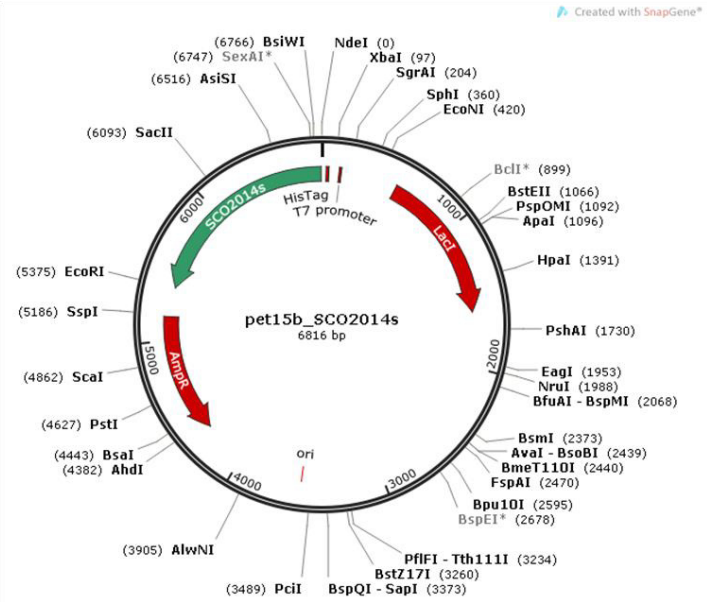



Figure 5.10 Construction and verification of pJH11. It constructed using a codon-optimised for *E. coli pyk1* sequence with the addition of restriction sites *Nde*I at the start of the gene and EcoRI at the end of the gene. This was synthesized at MWG eurofins and was delivered in the vector pEX-K4. The insert was released using *Nde*I and *Eco*RI and was then sub-cloned into pET15b. pET15b was subjected to a digestion with *Nde*I and *Eco*RI as well as dephosphorylation using alkaline phosphatase (Promega). The ligation was then used to transform DH5α and clones were selected on 100 µg/ml carbenicillin. The construct was verified by restriction digest using *Nde*I and *Eco*RI.

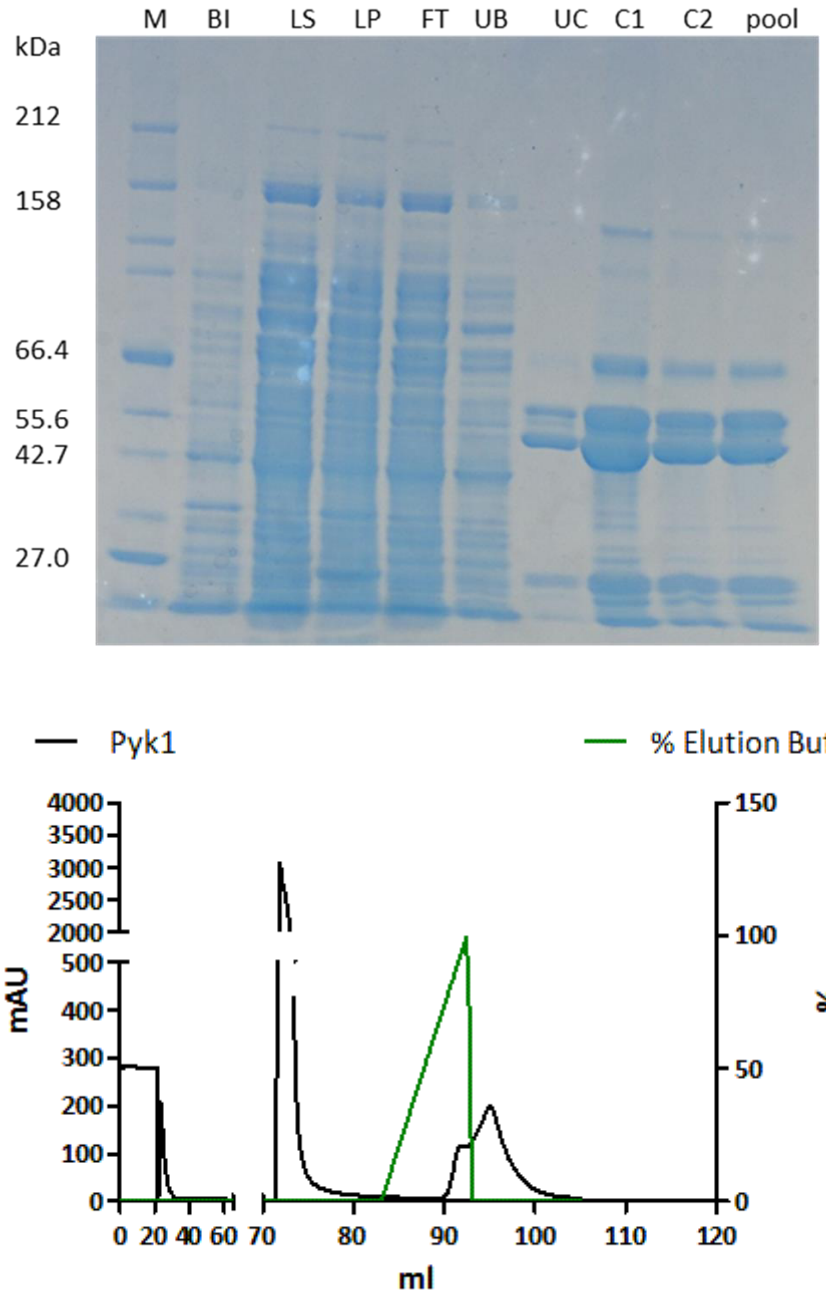



Figure 5.11 Top: SDS Page and purification profile of Pyk1. Overexpression of Pyk1 in *E. coli* Origami B grown at 30˚C in LB with 1% glucose and were induced at $OD_{600}$ 0.5 with 0.05 mM IPTG and the temperature was decreased to 20 ˚C and left overnight. This approach lead to a successful overexpression of Pyk1. In SDS-PAGE the protein ran at approximately 55.6 kDa. Bottom: Typical purification profile of Pyk1 from the cells of 1 L of culture: 70 -80 ml unbound protein; 90 ml Pyk1 followed by Imidazole gradient. Legend: M = marker, BI = before induction, LS= lysate supernatant, LP = lysate pellet, FT = flow through from column loading, UB = unbound protein, UC = unconcentrated eluted protein, C1 = concentration cycle 1, C2 = concentration cycle C2

### 5.3.2 Overexpression of Pyk2

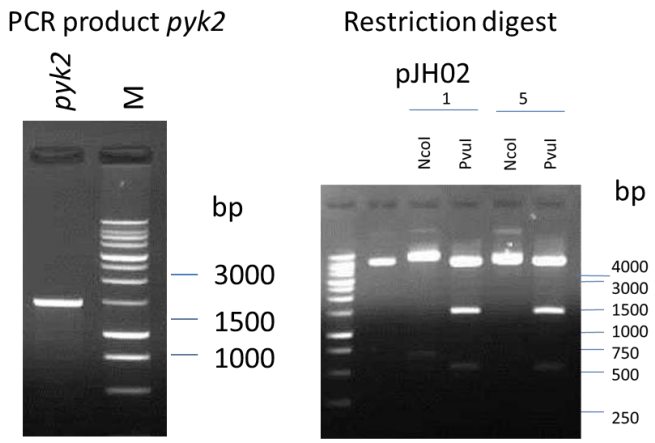
The overexpression of pyk2 was carried out using the plasmid pJH02. To construct this plasmid, the *pyk2* gene was amplified from *S. coelicolor* genomic DNA and was cloned into the commercial plasmid pET-TOPO100 (Invitrogen). Primers were designed with CACC overhang at the 5' end of the forward primer which was the first 18 nucleotides at the N-terminus of *pyk2.* The reverse primer was from the C-terminus ending with the stop codon of *pyk2* (Table 2-7). The blunt-end obtained PCR product was then ligated into pET-TOPO100 as specified by the supplier's manual. The final constructs were then confirmed by restriction digestion using *Nco*I and *Pvu*I (Figure 5.12). Furthermore the constructs were sent for sequencing (MWG eurofins). Pyk2 overexpression was carried out successfully using pJH02 in *E. coli* Rosetta grown on AIM for 2h at 37°C and then switched down to 20˚C and left the culture growing overnight. Pyk2 also shows a higher MW as predicted (51.6 kDa) and ran approximately at 55.6 kDa on SDS-PAGE (Figure 5.13).

## 5.4 Cell lysis

Prior to the purification an experiment was carried out to determine the time needed to break open the cells by sonication. This was necessary to determine maximum protein yield with least damage, as sonication produces significant heat that can lead to denaturation of the protein. Pyk2 overexpressing *E. coli* cultures were utilised for this and by bearing in mind the $OD_{600}$ of the cell harvest 2.5 s per $OD_{600}$ was best. A culture grown on LB plus glucose had an $OD_{600}$ of 40 and 90s sonication was best, which equals ~2.3 sec per OD and AIM had $OD_{600}$ of 80 and 210 s was best, which was ~2.6 sec per $OD_{600}$ (Figure 5.14). The sonication was carried out in 30 s intervals on ice with 60 s cool down on ice in between each pulse.

## 5.5 Purification of Pyk1 and Pyk2

Pyk1 and Pyk2 were purified using the His-tag which was added to the protein by the respective vector (pET15b for Pyk1 and pET_TOPO100 for Pyk2). An imidazole gradient was applied in order to elute the overexpressed protein bound to a Nickel column. Fractions were collected and pooled if necessary. Pyk1 had to be concentrated in order to obtain sufficient amounts to measure activity. Usually 4 L of *E. coli* culture were necessary for about 60 reactions, whereas in case of Pyk2 the purified protein had to be diluted 1:5 or 1:10 depending on the protein titre in order to measure activity in the detectable range.




| Plasmid | Restriction enzymes (# sites) | Expected band sizes (bp) |
|---|---|---|
| pJH02 | PvuI (3) | 5220, 1440, 530 |
| (pET_TOPO100_SCO5423) | NcoI (2) | 6500, 650 |

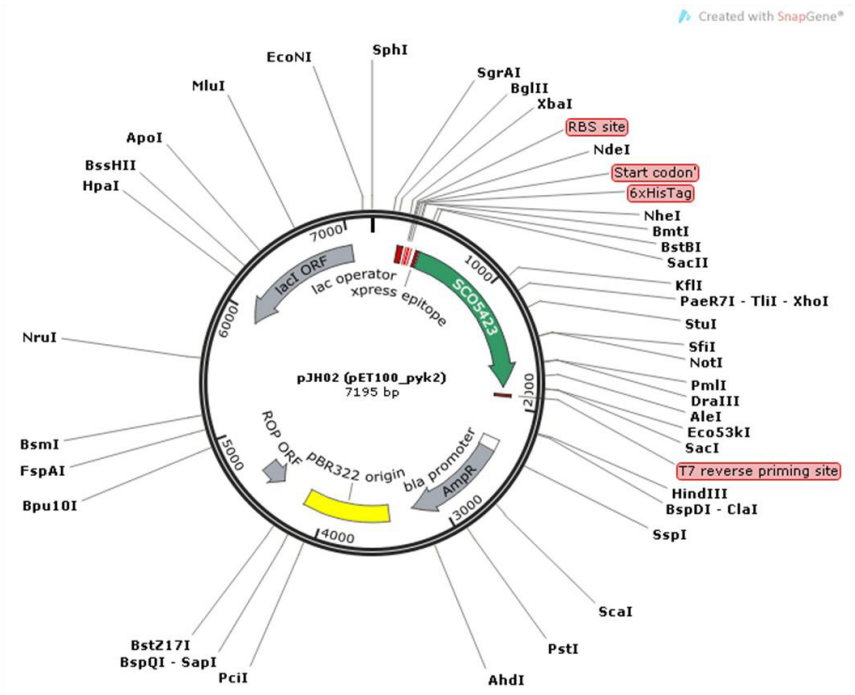



Figure 5.12 Construction and verification of pJH02 using the commercial vector pET_TOPO100. The insert was generated by PCR using a forward primer composed of the first 18 nucleotides with a CACC 5' overhang to ensure in frame location to the RBS on the plasmid. The reverse primer was from the C-terminus ending with the stop codon of the respective gene. The PCR product, which was blunt-ended obtained PCR product was then ligated into pET-TOPO100 as specified by the supplier's manual. The final construct was then confirmed by restriction digestion using *Pvu*I and *Nco*I

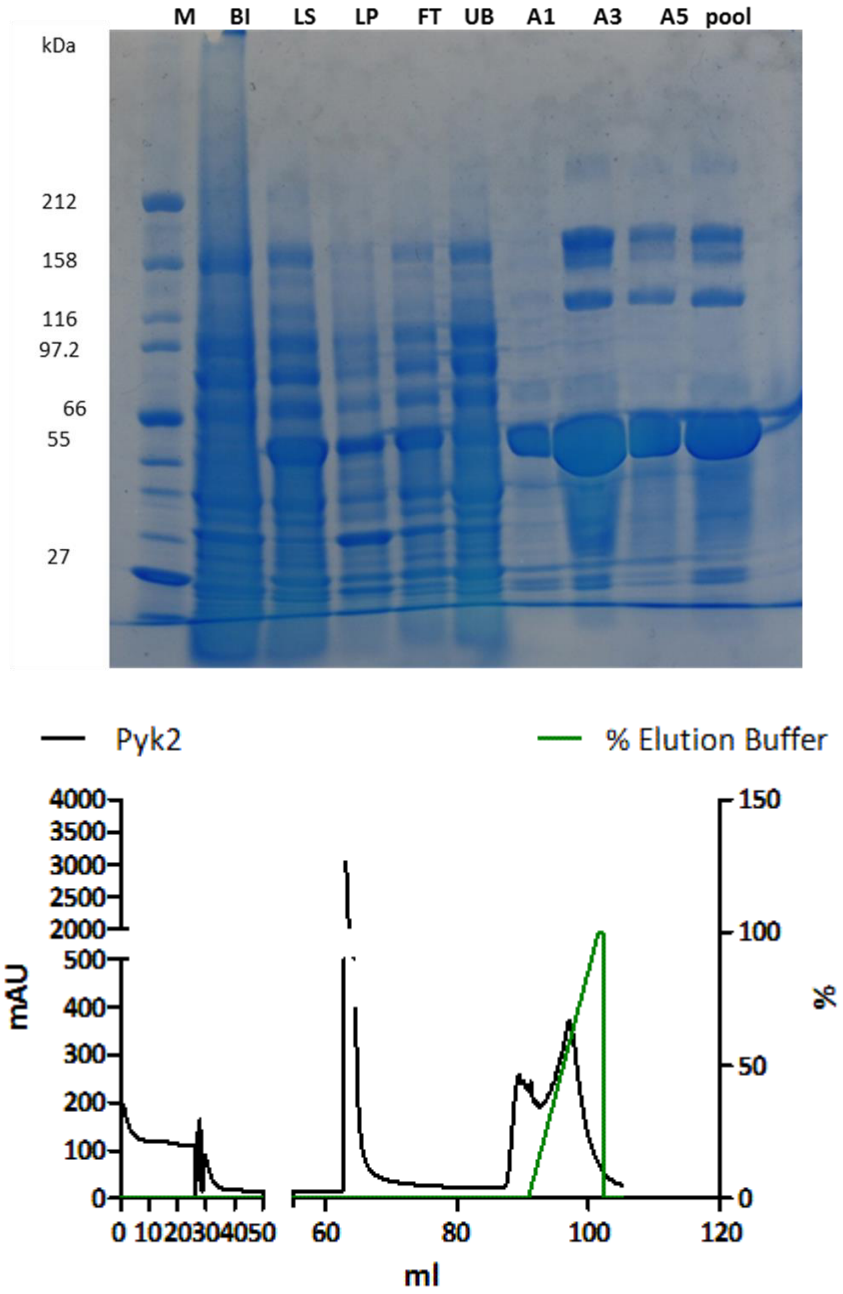



Figure 5.13 SDS Page and purification profile of Pyk2 .Top: Overexpression of Pyk2 in Rosetta grown at 37˚C in AIM for 2 h and then temperature was decreased to 20˚C and left to grow overnight. This approach lead to a successful overexpression of Pyk2. The protein ran higher at 55.6 kDa than its expected size of 51.6 kDa in SDS-PAGE. Bottom: A typical purification profile of Pyk2 from 1 L of culture. 62 ml unbound protein, 88 mL Pyk2 followed by Imidazole gradient. Legend: M = marker, BI = before induction, LS= lysate supernatant, LP = lysate pellet, FT = flow through from column loading, UB = unbound protein, A1/A3 and A5 = different eluted fractions of Pyk2

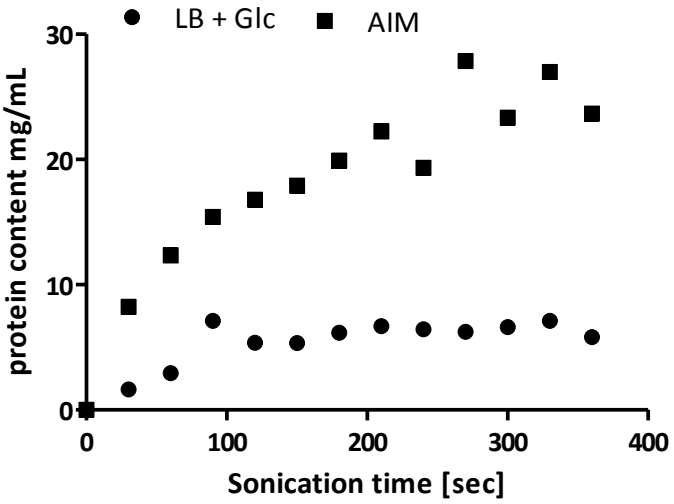



Figure 5.14 Protein content determination by Bradford over time from harvested *E. coli* cells overexpressing Pyk2 in order to optimise the cell lysis to obtain functional protein without damaging the protein by the heat.

## 5.6 Size Determination of Pyk1 and Pyk2 by Gel filtration

In order to calculate $K_{cat}$ from the kinetic data, a determination of the size of each of the pyruvate kinases was necessary as the SDS PAGE only determines the size of a monomer and not of the whole functional protein due to its denaturating conditions. The size determination was performed by gel filtration. The results indicated a tetrameric form for Pyk1 (206 kDa) with and a hexamer for Pyk2 (307 kDa) when using the predicted molecular weight of 51.2 and 51.6 kDa respectively of a single chain (Figure 5.15).

## 5.7 Circular Dichromism (CD)

A circular dichromism measurement was performed to ensure folded protein was present and to gain some insight about the structure of each of the pyruvate kinases. Both spectra look different from the profile of the phosphate buffer. For the interpretation of a spectrum, the range from 200 -240 nm is most important.

A drop in the spectrum at 209-222 nm is an indicator of significant presence of α-helices and β-sheets, which was observed for Pyk2 with the biggest drop at 215 nm (-2.5 µdeg*$cm^{-2}$*$dmol^{-1}$). For Pyk1 the most significant peak was at 227 nm with a value of -8 µdeg*$cm^{-2}$*$dmol^{-1}$. This drop is observed when mainly β-sheets are present (Hu *et al.*, 2011). This indicates that the two proteins differ structurally (Figure 5.15).

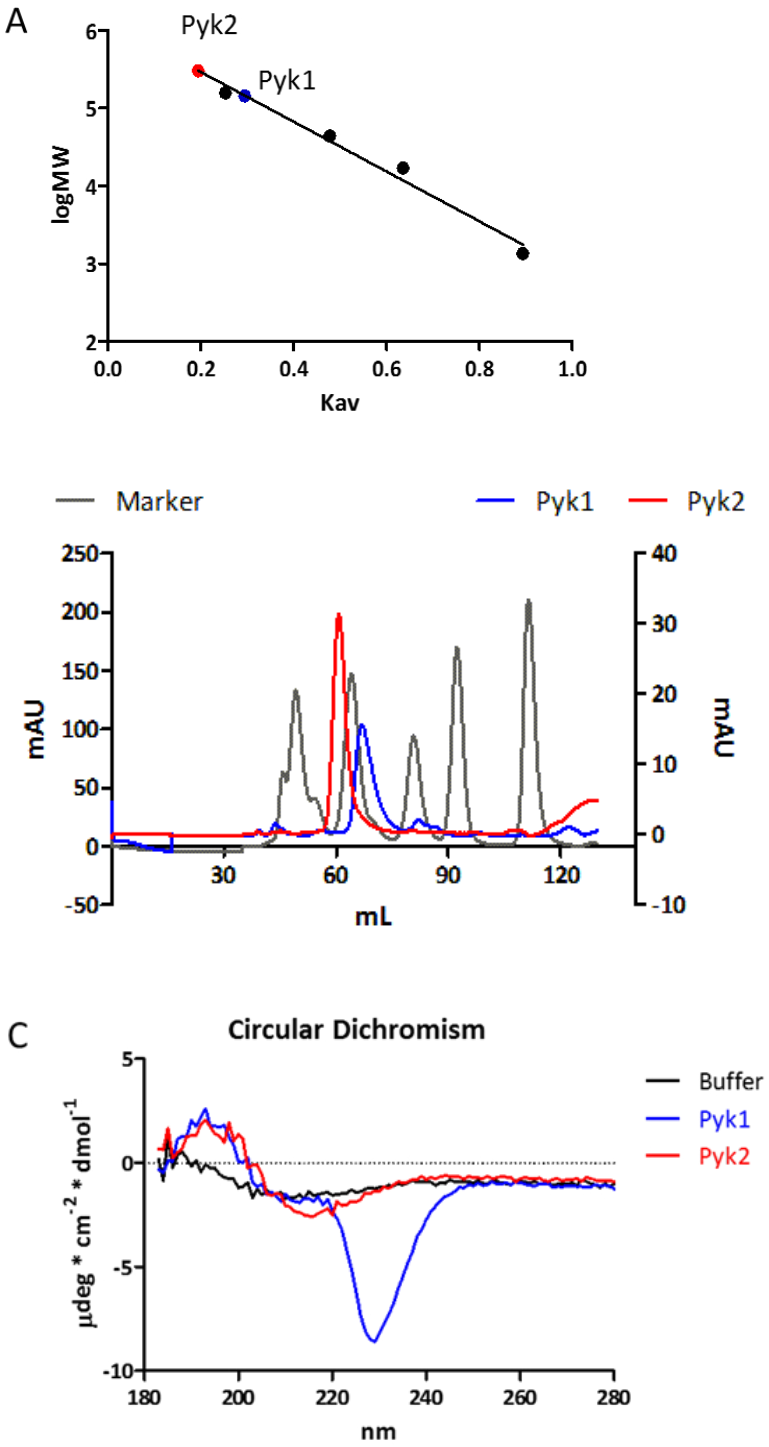



Figure 5.15 (A & B) Size determination of the two pyruvate kinase Pyk1 and Pyk2 by gel filtration in comparison to a molecular weight standard (C) circular dichromism scans of Pyk1 and Pyk2 giving an indication about the folding from 180-280 nm. The range from 200- 240 nm is most important for information on secondary structure.

## 5.8 Functional Characterisation of Pyk1 and Pyk2 – Kinetics and Effectors

A detailed characterisation of the kinetics and effectors for the two pyruvate kinases requires an activity assay. Pyruvate kinase activity is measured in a coupled assay using lactate dehydrogenase (LDH). This assay measures the decrease of β-NADH, which is proportional for the generation of pyruvate. This is possible as pyruvate kinase converts PEP and ADP into pyruvate and ATP, then pyruvate is converted by LDH together with β-NADH into L-lactate and β-NAD.

### 5.8.1 Assay Establishment

To establish the assay conditions, a commercial pyruvate kinase from rabbit muscle was used (Sigma-Aldrich). The conditions specified previously (Bergemeyer *et al.*, 1974) used 5 U of LDH in 1.5 mL reaction volume with 0.58 mM PEP and 1.5 mM ADP. Different LDH concentrations were tested at the specified substrate concentrations. The more LDH was added the quicker the β-NADH was consumed and when no LDH was added, the reaction did not occur. From this it was confirmed that 5 U LDH was the concentration to use as it showed a stable decrease in absorbance over five minutes. This will allow sufficient margin to run the reactions at lower and higher concentrations of substrate in order to obtain kinetic data. Thus the addition of 5 U LDH were used subsequently (Figure 5.16).

In addition to the LDH concentration, the two substrates PEP and ADP concentrations were also varied using rabbit muscle pyruvate kinase.

It is known that most pyruvate kinases show sigmoidal kinetics for PEP, which can be approximated by the Hill equation. The Hill equation describes the reaction velocity as a product of $V_{max}$ and the substrate concentration (s) to the power of the Hill coefficient (n) divided by the sum of $S_{0.5}$ to the power of n, which is the substrate concentration at which ½ $V_{max}$ is reached and the substrate concentration to the power of n. The Hill coefficient (n) is an indicator for cooperativity. When no cooperativity is present it will be 1, if negative cooperativity (inhibition) is present it will be lower than 1 and positive cooperativity (allostery) it will be higher than 1. It is only an indicator and the number does not equal the number of additional sites present apart from the active site (Cornish Bowen, 2012).

$$V = \frac{V_{max}\, s^n}{({S_{0.5}}^n + s^n)}$$

This means that a threshold concentration of PEP needs to be reached before the reaction takes place as PEP acts as its own activator (Figure 5.16).

In contrast to PEP, the substrate ADP follows classic Michaelis-Menten kinetics, which is one of the simplest and best known models to explain kinetics of enzymes in a relation to velocity and substrate concentration. The velocity can be expressed as the ratio of $V_{max}$ times substrate concentration and the sum of $S_{0.5}$ and substrate concentration. $S_{0.5}$ describes the substrate concentration at which ½ $V_{max}$ is reached (Cornish Bowen, 2012).

$$V = \frac{V_{max}\, s}{(S_{0.5} + s)}$$

The reaction rate increased with increasing ADP concentration suggesting Michaelis-Menten kinetics for ADP (Figure 5.16).

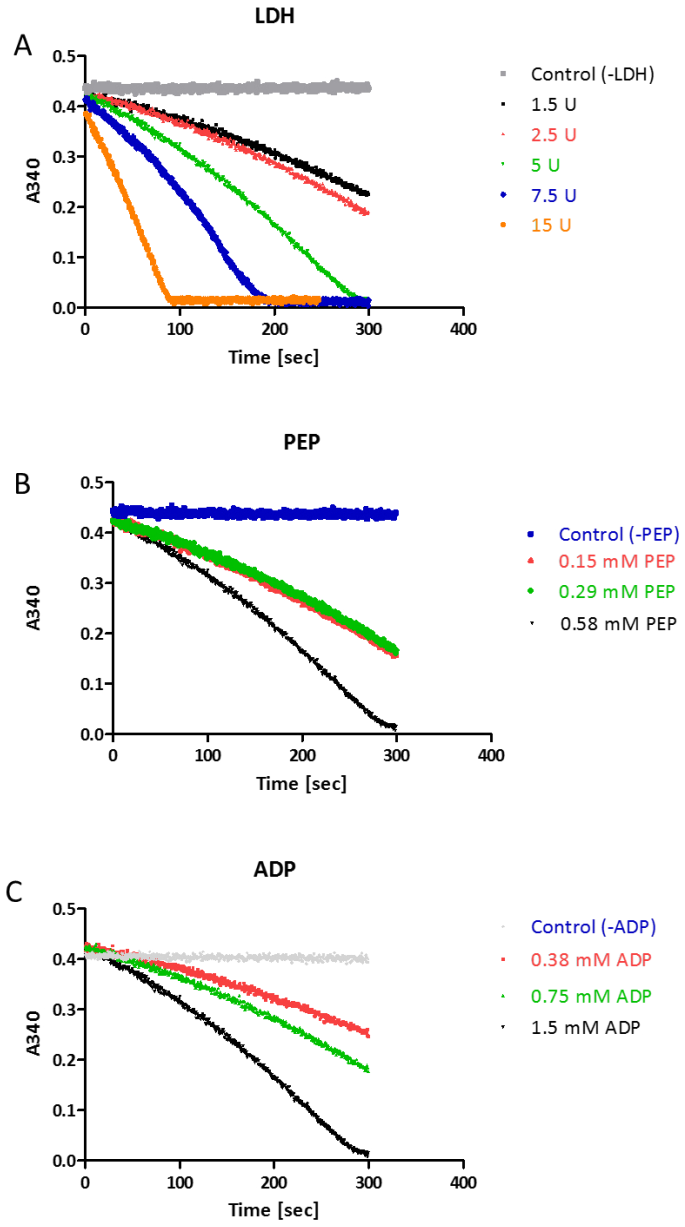



Figure 5.16 Validation of the pyruvate kinase activity assay using commercial pyruvate kinase from rabbit muscle (Sigma) with different amounts of LDH (A), PEP (B) and ADP (C). The concentration of the other chemicals were 39 mM $K_2HPO_4$ (pH 7.2), 0.11 mM β-NADH, 6.8 mM $MgSO_4$ and 0.02 U of rabbit muscle pyruvate kinase (1 unit = 1 µmole of PEP converted to pyruvate per minute at 30C at pH 7.2), for LDH variation (A) 0.58 mM of PEP and 1.5 mM ADP, for PEP (B) 10 U of LDH and 1.5 mM ADP and for ADP (C) 10 U LDH and 0.58 mM of PEP

### 5.8.2 KINETIC CHARACTERISATION OF PYK1 AND PYK2

**Pyk1**

For Pyk1, when utilising a low PEP concentration (0.58 mM) no activity was observed when no AMP was present compared to a low activity when AMP was present ($V_{max}$ of 0.07 U/mg compared to 1.93). The $S_{0.5}$ could not be determined when no AMP was present as it was below detection limit. Overall the presence of 1 mM AMP resulted in almost 28 fold increase of $V_{max}$. However as the concentration was likely not to be saturating for PEP the kinetics for ADP were repeated in the presence of 5 mM PEP (Figure 5.17).

ADP kinetics were consistent with the Michaelis-Menten equation. The $S_{0.5}$ was lowered 4-fold in presence of 1 mM AMP from 0.59 mM down to 0.15 mM. $V_{max}$ was increased 3.5-fold from 21 U/mg to 73.3 U/mg in presence of 1 mM AMP (Figure 5.17).

The PEP kinetics could be described by the Hill equation. In the presence of 1 mM AMP, $V_{max}$ was almost five times increased from 14.05 U/mg to 65.45 U/mg. The substrate concentration at ½ $V_{max}$ ($S_{0.5}$) was lowered more than three-fold (from 3.49 to 1.05 mM) and the Hill coefficient was approximately halved from 3.74 to 1.78, indicating that one effector is AMP, but probably there is another one as the Hill coefficient is still higher than 1 (Figure 5.17).

**Pyk2**

The ADP kinetics at 0.58 mM PEP showed an increase for $V_{max}$ of almost ten-fold from 0.13 U/mg to 1.25 U/mg in presence of 1 mM AMP. Interestingly the $S_{0.5}$ almost doubled in presence of AMP from 0.61 to 1.14 mM (Figure 5.17).

Again, the kinetics were repeated at a PEP concentration of 5 mM. The kinetics could be described using Michaelis-Menten equation and showed a five-fold increase of the $V_{max}$ in presensce of 1 mM AMP from 1.2 U/mg to 6.7 U/mg. The $S_{0.5}$ was lowered three-fold from 0.27 mM down to 0.09 mM (Figure 5.17).

For PEP the kinetics were best described by the Hill equation. In presence of 1 mM AMP $V_{max}$ was 9.09 U/mg compared to 0.48 U/mg without AMP. The $S_{0.5}$ increased from 1.3 mM to 8.6 mM and the Hill coefficient increased from 1.48 to 7. This suggests that, when AMP is present it does increase the $V_{max}$. However the allostery seems to be increased indicating that another allosteric regulator is influencing Pyk2 (Figure 5.17).

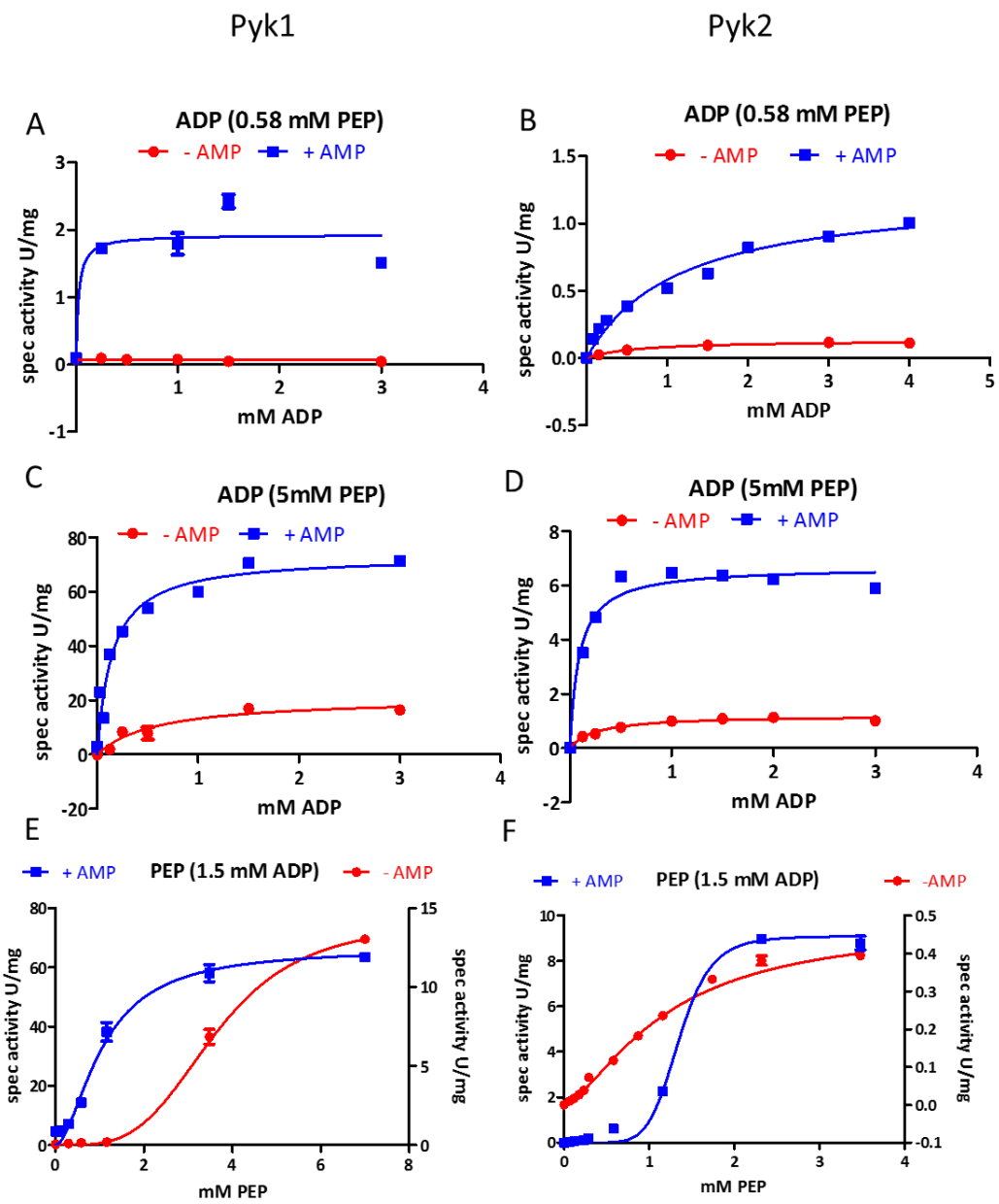



Figure 5.17 Kinetic data of Pyk1 and Pyk2 for both substrates. ADP in presence of low amounts of PEP (0.58 mM,) (A) ADP kinetics at 0.58 mM PEP for Pyk1, fitted with Michaelis Menten kinetics (B) ADP kinetics at 0.58 mM PEP for Pyk2 , fitted with Michaelis Menten kinetics (C) ADP kinetics at 5 mM PEP ADP for Pyk1, fitted with Michaelis Menten kinetics (D) ADP kinetics at 5 mM PEP ADP for Pyk1 , fitted with Michaelis Menten kinetics (E) PEP kinetics at 1.5 mM PEP ADP for Pyk1 , fitted with Hill kinetics (F) PEP kinetics at 1.5 mM PEP ADP for Pyk2 fitted with Hill kinetics

**Comparison of Pyk1 and Pyk2**

During assaying the two enzymes, it was observed that Pyk1 was a lot less stable than Pyk2 with showing activity loss within hours on ice whereas Pyk2 was active for at least two weeks when stored at 4˚C.

Both enzymes responded positively to AMP, so the kinetics were determined in the presence and absence of 1 mM AMP. An overview of the kinetic characteristics is given in Table 5-1 for both enzymes. However the presence of AMP did increase the Hill coefficient for Pyk2, indicating that another allosteric regulator might be involved.

Since both enzymes responded to the presence of AMP, kinetics were measured for varying concentrations of AMP when using $S_{0.5}$ concentration of the substrates, which were 0.6 mM ADP and 3.7 mM PEP for Pyk1 and 0.3 mM ADP and 1.3 mM PEP for Pyk2. The result showed that Pyk1 had a much higher affinity to AMP with a $S_{0.5}$ of 0.01 mM, whereas Pyk2 showed a $S_{0.5}$ of 3.8 mM for AMP. $V_{max}$ was 8.2 U/mg for Pyk1 and 1 U/mg for Pyk2 (Figure 5.18).

Apart from AMP, other effectors have been tested for their influence on the two enzymes. They were selected from the literature as being described effectors for pyruvate kinase in other species: R5P, ATP, G6P, Citrate, FBP (Sakai *et al.*, 1986; Ernest *et al.*, 1994; Jetten *et al.*, 1994; Ernest *et al.*, 1998; Valentini *et al.*, 2000; Zoraghi *et al.*, 2010; Morgan *et al.*, 2014). R5P has a stronger effect on Pyk1 at high concentrations - with an almost eight fold increase of activity compared to 1.8 fold for Pyk2 (Figure 5.18).

ATP has an inhibitory effect at 1 mM for Pyk2, but not Pyk1. At high concentration (5 mM) both enzymes were inhibited similarly and the activity decreased by about 10%. G6P showed an inhibitory effect on Pyk1 but not Pyk2. Both citrate and FBP showed no effect on either enzyme (Figure 5.18).

Additionally to the kinetic parameters, the $K_{cat}$ was calculated, which is given by the ratio of $V_{max}$ and enzyme concentration (µmol of active sites). Another indicator is the specificity constant $K_{cat}/ S_{0.5}$ which reveals information about the specificity of the enzyme.

For ADP in absence of AMP, the $K_{cat}$ was of 941 $sec^{-1}$ and a $K_{cat}/ S_{0.5}$ of 1594 $mM^{-1}$ $s^{-1}$ for Pyk1 compared to 39 $sec^{-1}$ and a $K_{cat}/ S_{0.5}$ of 145 $mM^{-1}$ $s^{-1}$ in Pyk2. In the presence of AMP, the $K_{cat}$ was 4703 $sec^{-1}$ and $K_{cat}/ S_{0.5}$ was 31,359 $mM^{-1}$ $s^{-1}$ for Pyk1 compared to 215 $sec^{-1}$ and $K_{cat}/ S_{0.5}$ was 2388 $mM^{-1}$ $s^{-1}$ for Pyk2 (Table 5-1).

The substrate PEP showed a $K_{cat}$ of 350 $sec^{-1}$ and a $K_{cat}/ S_{0.5}$ of 100 mM-1 s-1 for Pyk1 compared to 18 $sec^{-1}$ and $K_{cat}/ S_{0.5}$ of 14 $mM^{-1}$ $s^{-1}$ in Pyk2. In the presence of AMP, the $K_{cat}$ was increased to 4200 $sec^{-1}$ and a $K_{cat}/ S_{0.5}$ was 4000 $mM^{-1}$ $s^{-1}$ for Pyk1 compared to 336 $sec^{-1}$ and a $K_{cat}/ S_{0.5}$ of 39 $mM^{-1}$ $s^{-1}$ for Pyk2 (Table 5-1).

For AMP Pyk1 had an $K_{cat}$ of 424 $sec^{-1}$ and a $K_{cat}/ S_{0.5}$ of 42,373 $mM^{-1}$ $s^{-1}$ $K_{cat}$ and Pyk2 of 42 $sec^{-1}$ and a $K_{cat}/ S_{0.5}$ of 11 $mM^{-1}$ $s^{-1}$ respectively (Table 5-1).

Table 5-1 Kinetic parameters of Pyk1 and Pyk2, nd = not determinable, LOD = limit of detection

| Substrate | Parameter | no AMP | | 1 mM AMP | |
|---|---|---|---|---|---|
| | | Pyk1 | Pyk2 | Pyk1 | Pyk2 |
| 0.58 mM PEP ADP | $V_{max}$ (U/mg) | 0.07 | 0.13 | 1.93 | 1.25 |
| | $S_{0.5}$ (mM) | < LOD | 0.61 | 0.02 | 1.14 |
| | $K_{cat}$ ($sec^{-1}$) | 1.7 | 4.8 | 48 | 46 |
| | $K_{cat}/ S_{0.5}$ ($mM^{-1}$ $s^{-1}$) | Nd | 4 | 2405 | 76 |
| 5 mM PEP ADP | $V_{max}$ (U/mg) | 21.0 | 1.22 | 73.3 | 6.68 |
| | $S_{0.5}$ (mM) | 0.59 | 0.27 | 0.15 | 0.09 |
| | $K_{cat}$ ($sec^{-1}$) | 941 | 39 | 4703 | 215 |
| | $K_{cat}/Km$ ($mM^{-1}$ $s^{-1}$) | 1594 | 145 | 31359 | 2388 |
| 1.5 mM ADP PEP | $V_{max}$ (U/mg) | 14.05 | 0.48 | 65.45 | 9.09 |
| | $S_{0.5}$ (mM) | 3.49 | 1.31 | 1.05 | 8.60 |
| | n | 3.74 | 1.49 | 1.78 | 7.14 |
| | $K_{cat}$ ($sec^{-1}$) | 350 | 18 | 4200 | 336 |
| | $K_{cat}/ S_{0.5}$ ($mM^{-1}$ $s^{-1}$) | 100 | 14 | 4000 | 39 |
| AMP | $V_{max}$ (U/mg) | 8.18 | 0.99 | | |
| | $S_{0.5}$ (mM) | 0.01 | 3.8 | | |
| | $K_{cat}$ ($sec^{-1}$) | 423.7 | 42.0 | | |
| | $K_{cat}/ S_{0.5}$ ($mM^{-1}$ $s^{-1}$) | 42373 | 11 | | |

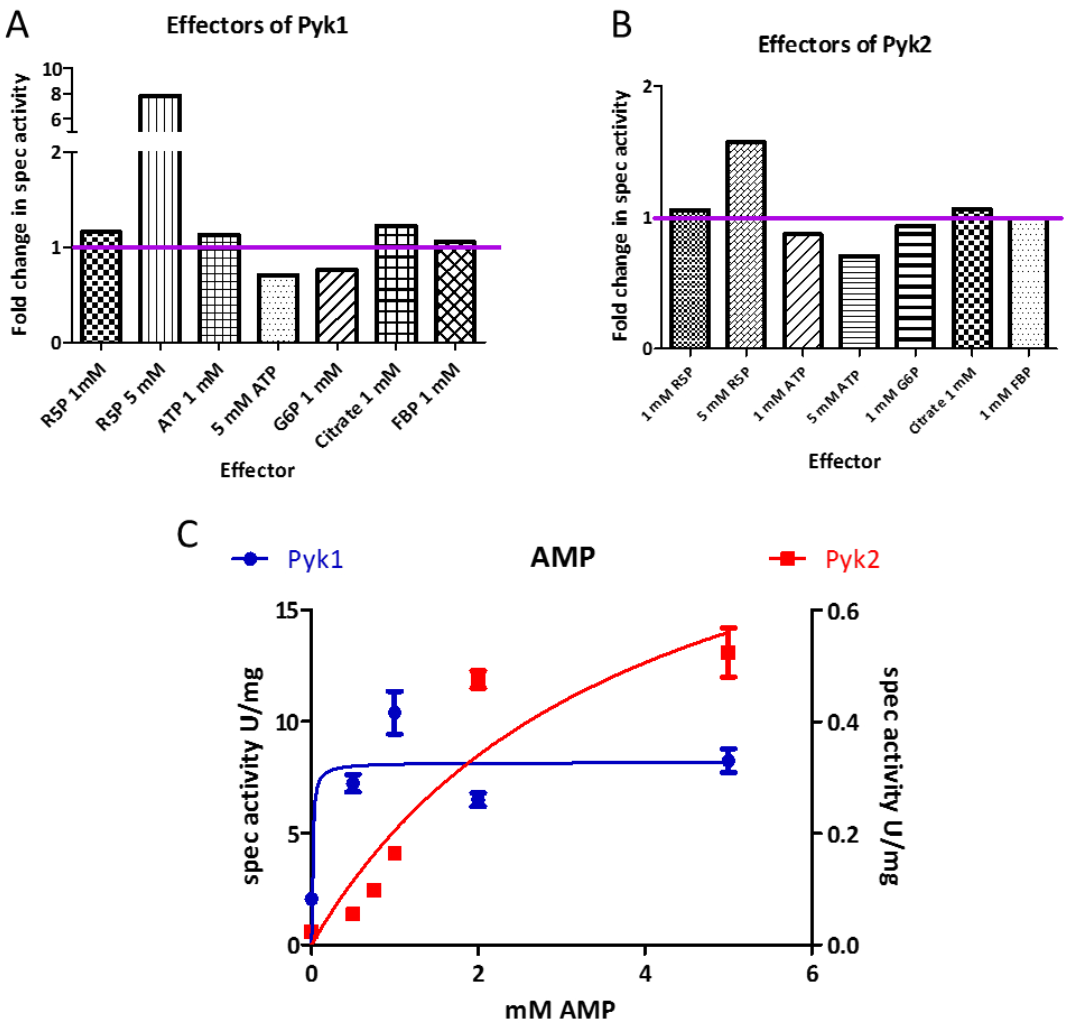



Figure 5.18 Fold change of activity in the presence of different effectors for Pyk1 (A) and Pyk2 (B) as well as detailed kinetics for AMP for both Pyk1 and Pyk2 (C). Legend: R5P = ribose-5-phosphate, ATP = adenosinetriphosphate, G6P = glucose-6-phosphate, FBP = fructose-bisphosphate

## 5.9 Analysis of Active Centre and Effector Sites of Pyk1 and Pyk2

The biochemical characterisation showed that Pyk1 and Pyk2 have distinct kinetic properties and furthermore indicated structural difference between them. Therefore the two pyruvate kinase genes from *S. coelicolor* were aligned to pyruvate kinases which have their structures solved (*E. coli*, *S. aureus*, *B. stearothermophilus*, *H. sapiens* (M2PYK), *L. mexicana*, *T. brucei* and *T. cruzi)* in order to analyse the effector site, active site, metal binding sites in Pyk1 and Pyk2 (Ernest *et al.*, 1994; Mattevi *et al.*, 1995; Ernest *et al.*, 1998; Suzuki *et al.*, 2008; Zoraghi *et al.*, 2011; Morgan *et al.*, 2013).

The residues involved in the active centre are marked with orange arrows and are in position 32, 242/243 and 275 being argininie, glycine, aspartic acid and threonine (R – GD – T). A difference observed here is the presence of a phenylalanine (F) at position 33 in both Pyk1 and Pyk2 compared to either leucine (L) or methionine (M) in the other bacterial or eukaryotic pyruvate kinases.

The positions for potassium binding are 34, 36 and 67 which normally are asparagine, serine and threonine respectively (N- S -T). However in Pyk1 and Pyk2 threonine, is replaced by leucine. All other sequences share the motif “LDTK” (leucine, aspartic acid, threonine and lysine) around the threonine at position 67 and Pyk1 and Pyk2 have “ADLQ” (alanine, aspartic acid, leucine and glutamine) instead.

The motif ‘KIEN’ (lysine, isoleucine, glutamic acid, asparagine) at position 222-226 is replaced by ‘KVEK/R’ (lysine, valine, glutamic acid and either lysine or arginine, position 228-232/229-233) in Pyk1 and Pyk2, as well as in *E. coli* PykA. This is the site where $Mg^{2+}$ binds and which acts as transition stabilizer.

The effector site is well described for F1,6BP and F2,6BP and is located at position 400-407 in the protozoan sequences. The sequence in Pyk1 and Pyk2 is different from the motif in human (LTKSGRS = leucine, threonine, lysine, serine, glycine, arginine and serine) and protozoic sequences (LSNT(S)GRS = leucine, serine, asparagine, threonine (or serine), glycine, arginine and serine) with FTQSGDT (=phenylalanine, threonine, glutamine, serine, glycine, aspartic acid and threonine). There is also no significant C’-domain present in the sequences of Pyk1 and Pyk2 in contrast to the sequences of *S. aureus* or *G. stearothermophilus* (Morgan *et al.*, 2014).

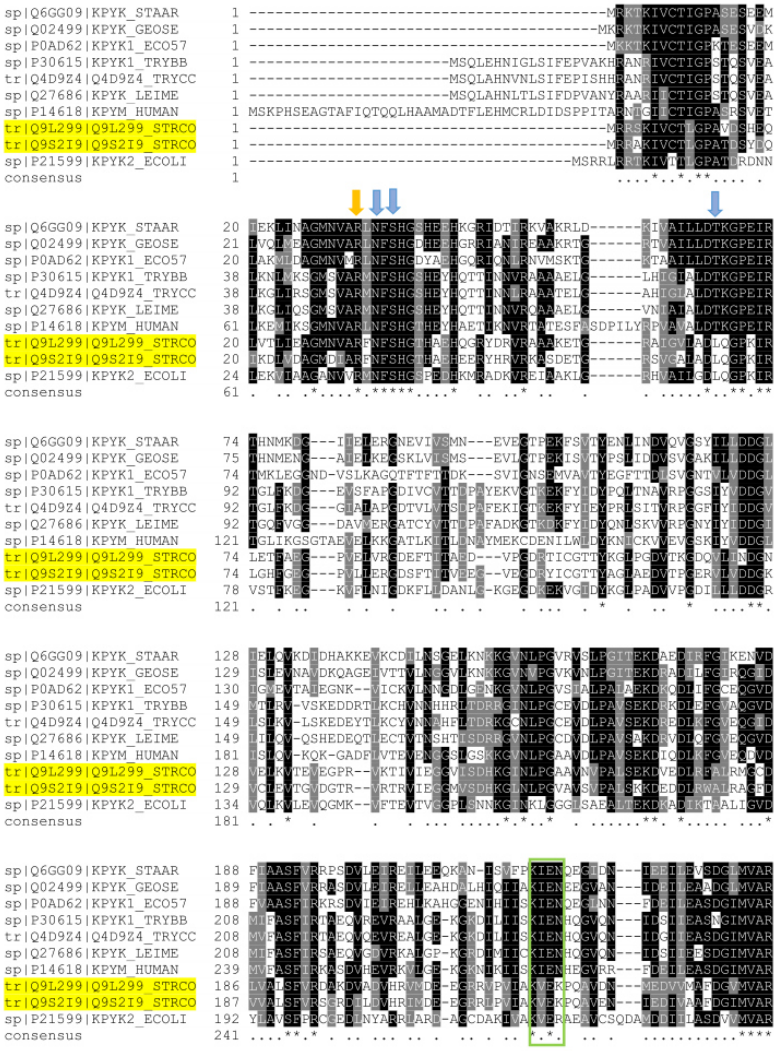


Figure 5.19 Alignment of pyruvate kinase sequences which have their 3D structure resolved. Alignment performed in ClustalW and the shading in Boxshade.

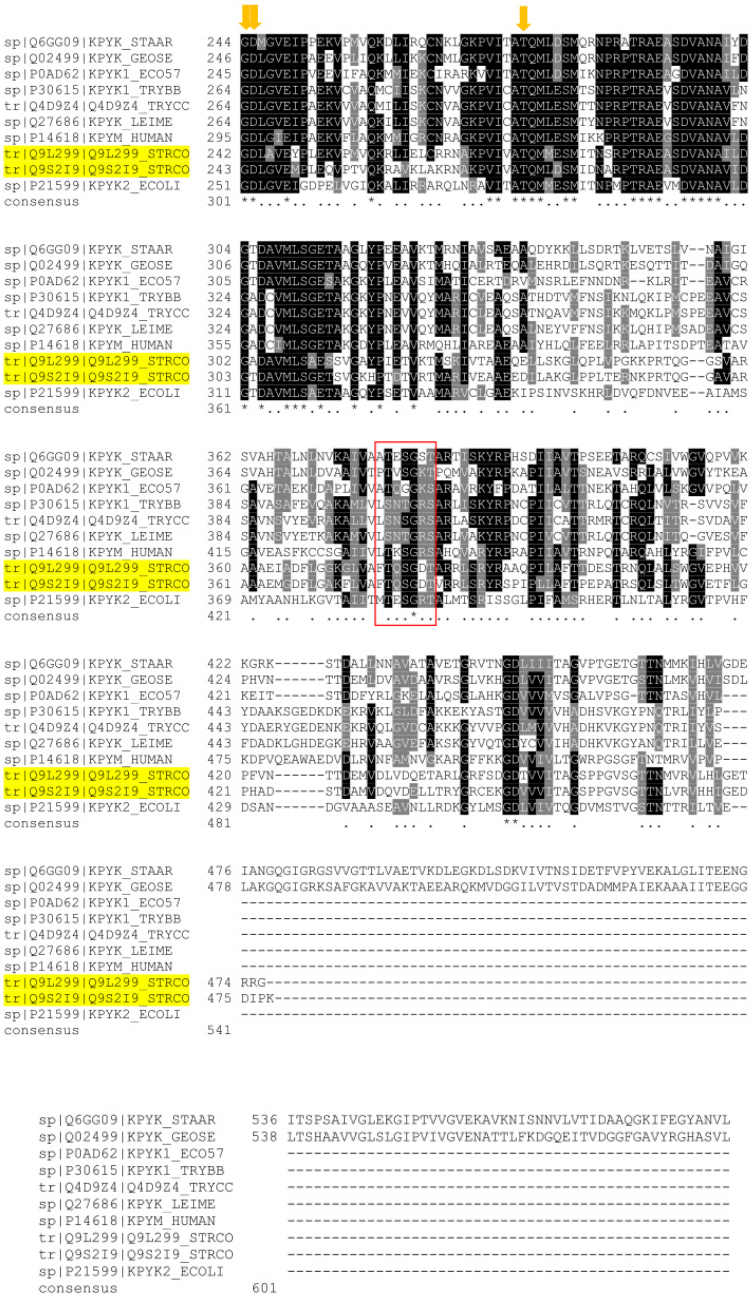


Figure 5.20 Pyruvate kinase alignment of pyruvate kinase sequences which have their 3D structure resolved. Alignment performed in ClustalW and the shading in Boxshade. Legend: orange arrows = active site residues, green box: $Mg^{2+}$ binding site/transition state stabilizer, blue arrows: $K^{+}$ binding site and red box – effector site in mammalian Pyk

In an attempt to gain further insight into the functionality of both Pyk1 and Pyk2, both were modelled using a homology modeller with the default settings (SWISS modeller, http://swissmodel.expasy.org/). Based on the alignment with pyruvate kinases whose structures are known, *E. coli* PykF was used as the underlying structure to model both Pyk1 and Pyk2 for the important sites such as active site, potassium, magnesium binding and effector sites, which were most similar among these than compared to the other prokaryotes, protozoan or mammalian pyruvate kinases.

The model created here is only hypothetical and actually disagrees with the experimental data at least for Pyk2, which suggests that the protein is a hexamer rather than a tetramer. This is the reason why only the sites that are involved in the actual reaction, and binding potassium, magnesium and the effector sites were analysed. Currently there is no structure available of a pyruvate kinase in presence of AMP and thus the allosteric binding site and exact position for binding AMP remains unknown and perhaps allosteric binding may affect the multimeric state of the protein.

The active centre has been conserved among the solved structures from *E. coli*, *S. aureus* and *B. stearothermophilus*. In the model for Pyk1 and Pyk2, the predicted active site as well as the binding sites for potassium, magnesium and the effector site seem similar, however as the model has not been tested it is only hypothetical. The presence of a phenylalanine next to one of the catalytic residues at position 33 might change the conformation of the active center as the side chain is much bigger than most other residues and depending where it faces it might cause a change in the conformation of the structure (Figure 5.21).

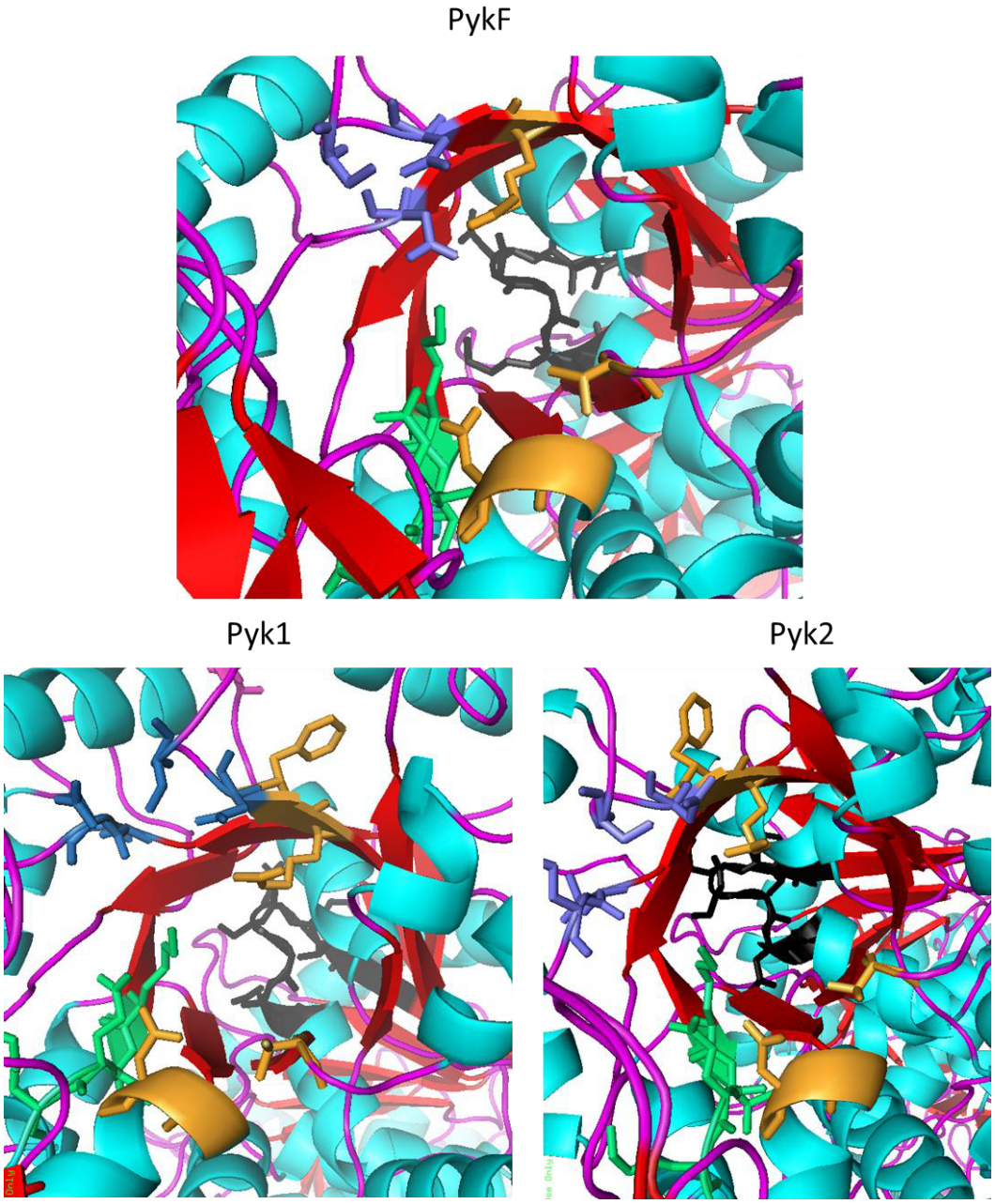



Figure 5.21 Figures depicting the active site (orange), potassium binding site (blue), magnesium binding site (green), effector site for FBP (black) for PykF from *E. coli*, Pyk1 and Pyk2 from *S. coelicolor*, generated by the SWISS modeller engine using PykF as template (Mattevi *et al.*, 1995)

## 5.10 SUMMARY

It can be concluded that both enzymes act primarily as pyruvate kinase since both could complement the pyruvate kinase double mutant in *E. coli*, which was unable to grow on glucose as sole carbon source. The semi quantitative expression analysis gave information about how each gene was expressed during different stages of the life cycle and no obvious differences could be detected using this method.

Overexpression of the enzymes was challenging for Pyk1, especially in order to obtain sufficient amounts. However by using a codon optimised copy of the gene and decreasing the temperature, sufficient active amounts were obtained to study the kinetics. Another observation was that Pyk1 seemed much less stable than Pyk2, as it lost its activity within hours after the purification. From the kinetics, it can be said that they differ in their allosteric regulation. Pyk1 had a much higher affinity to AMP and was also inhibited by glucose-6-phosphate. ATP had a more drastic effect on Pyk2 at lower concentrations. R5P was an activator for both; however Pyk1 activity increased by 8 fold whereas Pyk2 showed 1.5 fold increase. Furthermore the $K_{cat}$ for Pyk1 was about one order of magnitude higher than Pyk2. Another difference was the Hill coefficient determined for PEP kinetics increased for Pyk2 when 1 mM AMP was present and Pyk1 showed a decrease. The enzymes seemed to oppose each other in this particular condition. However in the absence of 1 mM AMP, Pyk2 has almost three fold lower $S_{0.5}$ for PEP than Pyk1 with 1.31 mM compared to 3.74 mM respectively.

Gel filtration indicated that the quartenary structures of the proteins are likely different; Pyk2 is most likely a hexamer whereas Pyk1 is a tetramer. This was also supported by the circular dichromism spectra which showed different profiles suggesting a mix of α-helices and β-sheets for Pyk2 and predominantly β-sheets for Pyk1. The most common structure reported for pyruvate kinase is a homotetramer (Muñoz and Ponce, 2003; Morgan *et al.*, 2014).

In general it can be said, that even though both are pyruvate kinases and are both expressed throughout the life cycle, their biochemical and structural properties are different. This fits with the different phenotypes observed for the mutants in Chapter 4.

# 6 DIFFERENTIAL GENE EXPRESSION OF CENTRAL CARBON METABOLISM ON GLUCOSE AND TWEEN

The growth curves and the RNA Sequencing were performed in collaboration with Richard Reumerman, a PhD student at the time with Dr Paul Herron at the University of Strathclyde.

The aim of this Chapter was to analyse differences and similarities in the transcription profile of redundant genes under conditions that drive growth via glycolysis or gluconeogenesis with a view to establishing the physiological roles of different isoenzymes.

## 6.1 RNA-SEQUENCING

Total RNA was isolated at log and stationary phase from cultures grown in minimal medium with either glucose or Tween40 as sole carbon source (Figure 6.1). In this minimal medium, the sole nitrogen source was $NaNO_3$. The log phase of growth was found to be shorter on glucose, approximately 10 h compared to approximately 25 h on Tween40. To account for these differences, sampling was based on the overall growth curve profile rather than based on time. The glucose sample was taken at 19 h and Tween sample at 36 h. The stationary phase sample was taken at 36 h from glucose and at 48 h from Tween (Figure 6.1). Only log phase RNA samples were used for the RNA sequencing experiment, while both log and stationary samples were analysed by qPCR (6.4).

The initial concentration of RNA extracted from the biomass sample was greater than 300 ng/uL for the glucose samples and more than 100 ng/uL for the Tween40 samples (Table 6-1). The RNA integrity number (RIN) is a software tool from Agilent taking the whole electrophoretic trace into account. To enable the calculation the software needs to detect the marker trace. It ranges on a scale from 10 for intact to 1 for completely degraded (Imbeaud *et al.*, 2005). The marker trace was only detected for two of the Tween samples and gave 8.5 and 8.7 as RIN value. As all the samples looked alike on the gel electrophoresis, all were processed for RNA-Seq (Figure 6.2) as a value of over 7.5 is suggested for sequencing from the supplier.

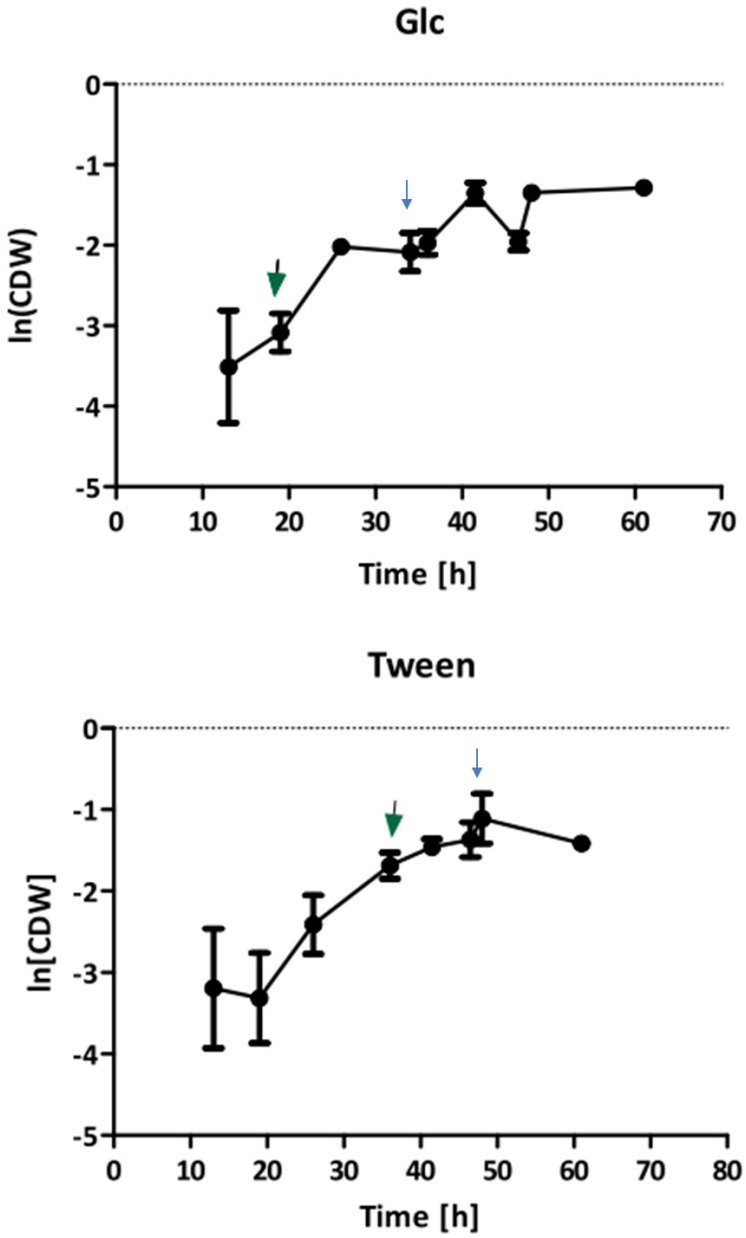



Figure 6.1 Growth curves of *S. coelicolor* M145 on HMM with either glucose or Tween as sole carbon source. The average of three biological replicates with standard deviation was plotted as ln of CDW versus cultivation time. 400 mL working volume in 2L flasks containing a metal spring were used. Samples for log phase are indicated by arrows. Additional samples at 36 h for glucose and 42 h for Tween were taking for RNA isolation and subsequent analysis with qPCR (6.4) Error bars represent standard deviation (σ) calculated by $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$ where N is the sample size, x each of the value and μ the average of all sample values.

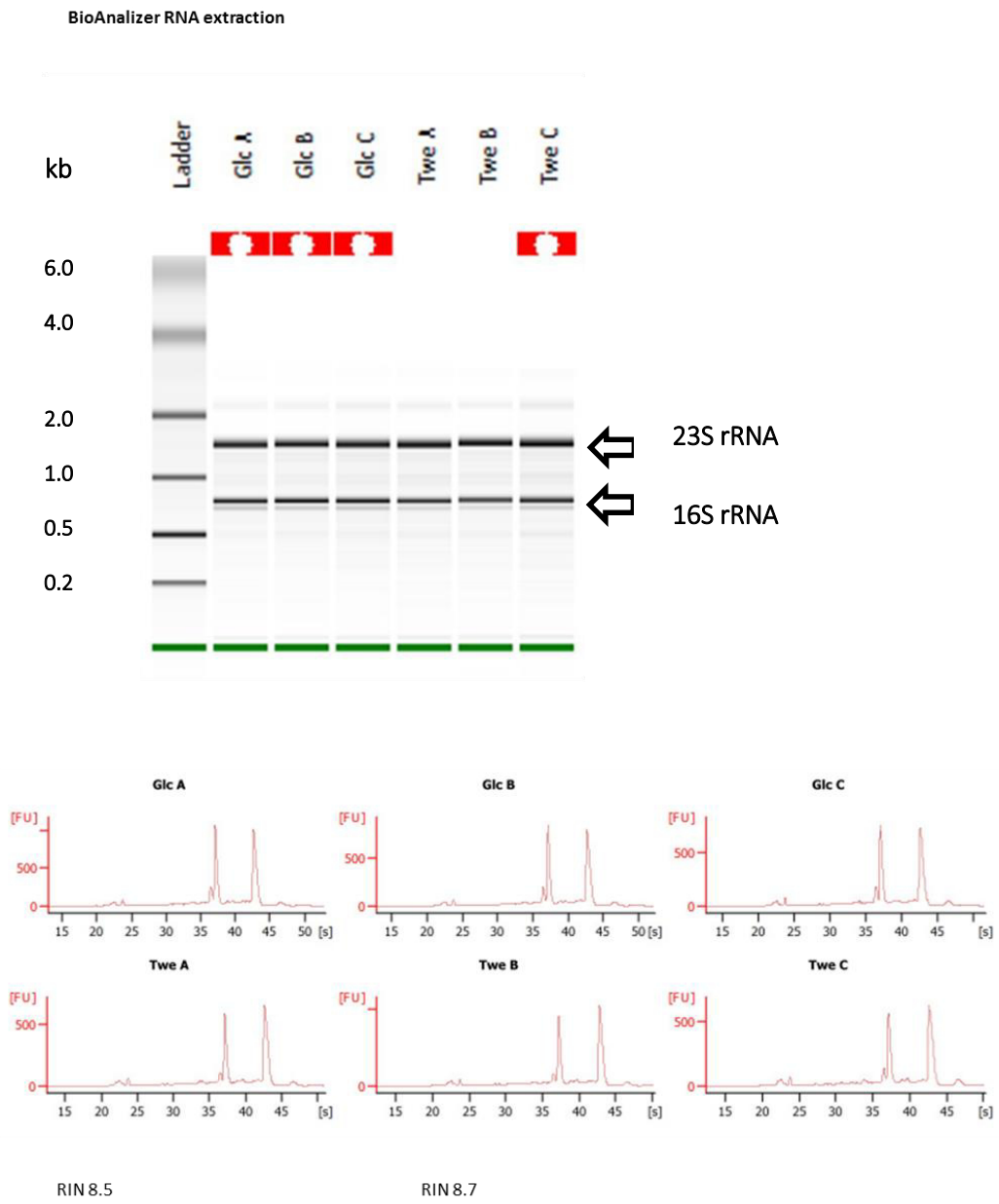



Figure 6.2 BioAnalizer traces using RNA 6000 Pico Kit (Agilent) for the isolation of RNA from samples obtained from the growth curves. Top: Gel view of the run, Bottom: Chromatograms of each sample. Legend: GlcA–C = three samples from glucose cultures, TweA-C = three samples from Tween cultures

Before proceeding with the ribosomal depletion, the isolated RNA samples were used as templates for a standard PCR using *hrdB* primers (RTPCR primer, Table 2-7) to check for genomic DNA contamination. None of the samples showed a PCR product apart from the positive control with genomic DNA as template (Figure 6.3).

To deplete ribosomal RNA, an aliquot of total RNA (2.7 µg) was used per sample and quantified by Qubit following the depletion (Table *6-1*). All samples were subsequently analysed by Bioanalizer gel chip (RNA 6000 Pico Kit, Agilent) for quality control. The two distinct bands for ribosomal RNA (16s rRNA and 23s rRNA) were found to have disappeared, as expected. However another peak appeared after 20-25 sec which was unexpected, but because it was present in all samples, it was concluded that it was likely to be a medium specific peak (Figure 6.4).

The samples were fragmented and ligated to barcoding oligonucleotides and then converted into cDNA using SuperScript III Enzyme Mix (Life Technologies, part of Ion Total RNA-Seq Kit v2). The cDNA library was again analysed by using the BioAnalizer (DNA High Sensitivity Kit, Agilent). The distribution of the fragments was found to be in the range of 50- 1000 bp fragments. Each sample was diluted to 500 pM (Figure 6.5). The three biological replicates were pooled at a concentration of 20 pM for each. The two samples, each from a pool of three biological replicates were emulsified, amplified and enriched to obtain the template for the sequencing. The percentage of template ion spheres was 29.2 % for glucose and 23.5% for Tween, which was in the range of 10 – 30% as recommended by the manufacturer.

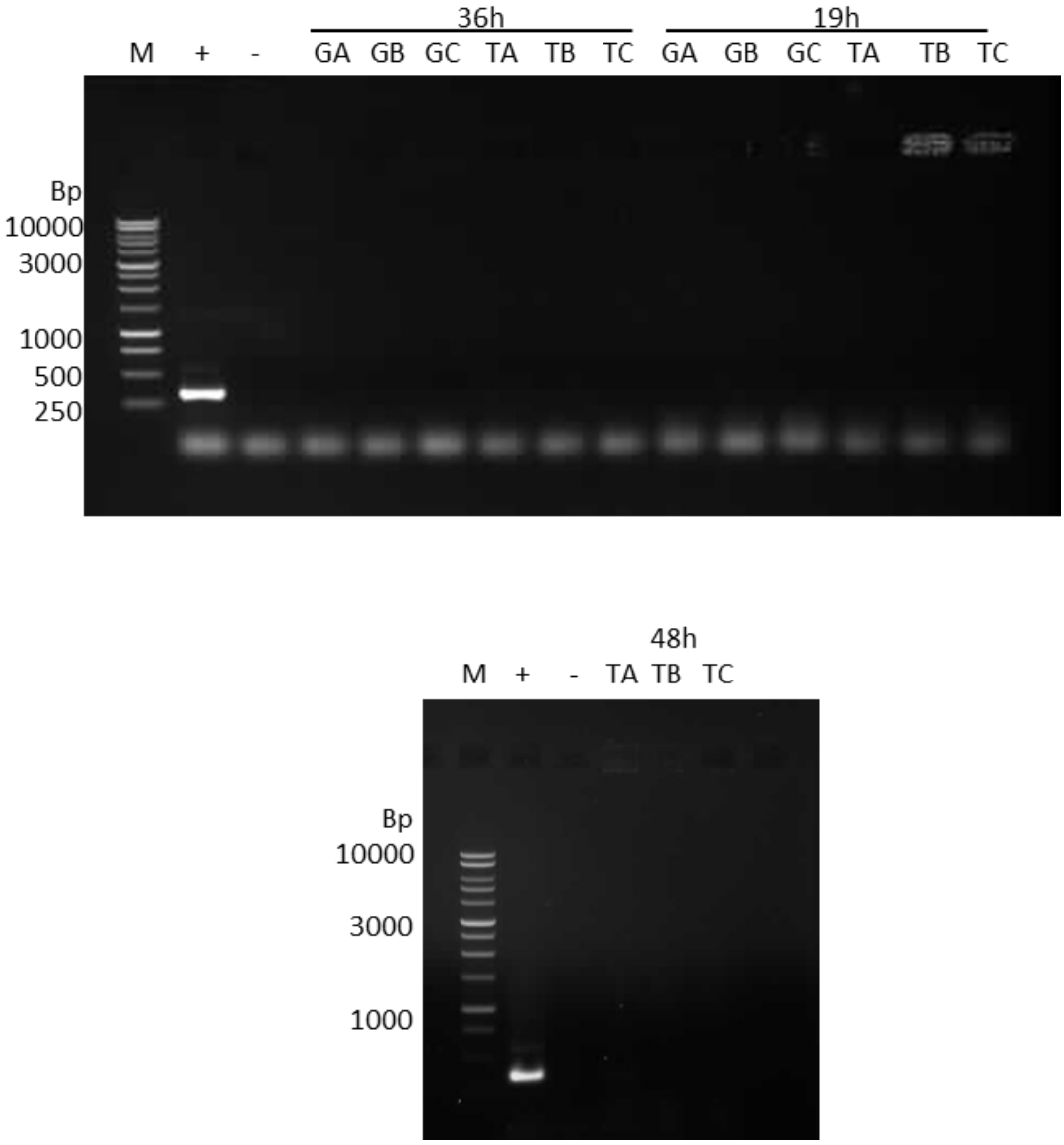



Figure 6.3 Gel image of conventional PCR using RT-PCR primers for *hrdB* (Table 2-7) with MyTaq to check for DNA contamination in RNA samples, positive control was genomic DNA and negative control RNAse free $dH_2O$, which was used to elute the samples during the isolation. Legend GA, GB, GC = biological triplicates of glucose cultures;TA, TB, TC = biological triplicates of Tween cultures, 19h = log phase of glucose samples, 36h = log phase of Tween samples and stationary phase samples of glucose, 48 h = stationary phase of Tween samples

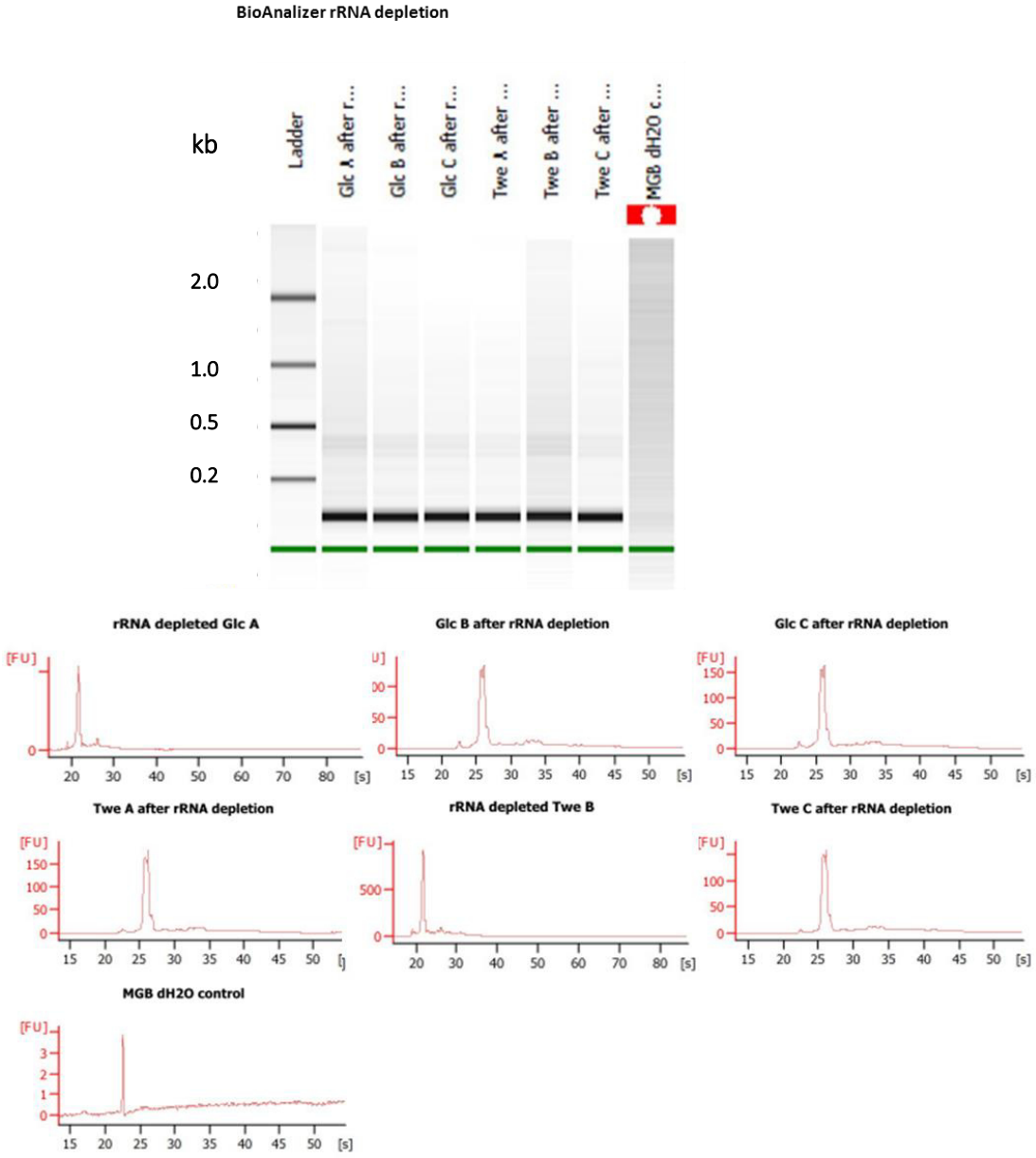



Figure 6.4 BioAnalizer traces using RNA 6000 Pico Kit (Agilent) after depletion of ribosomal RNA isolated from the growth curves. Top: Gel view of the run, Bottom: Chromatograms of each sample. Legend: GlcA–C = three samples from glucose cultures, TweA-C = three samples from Tween cultures

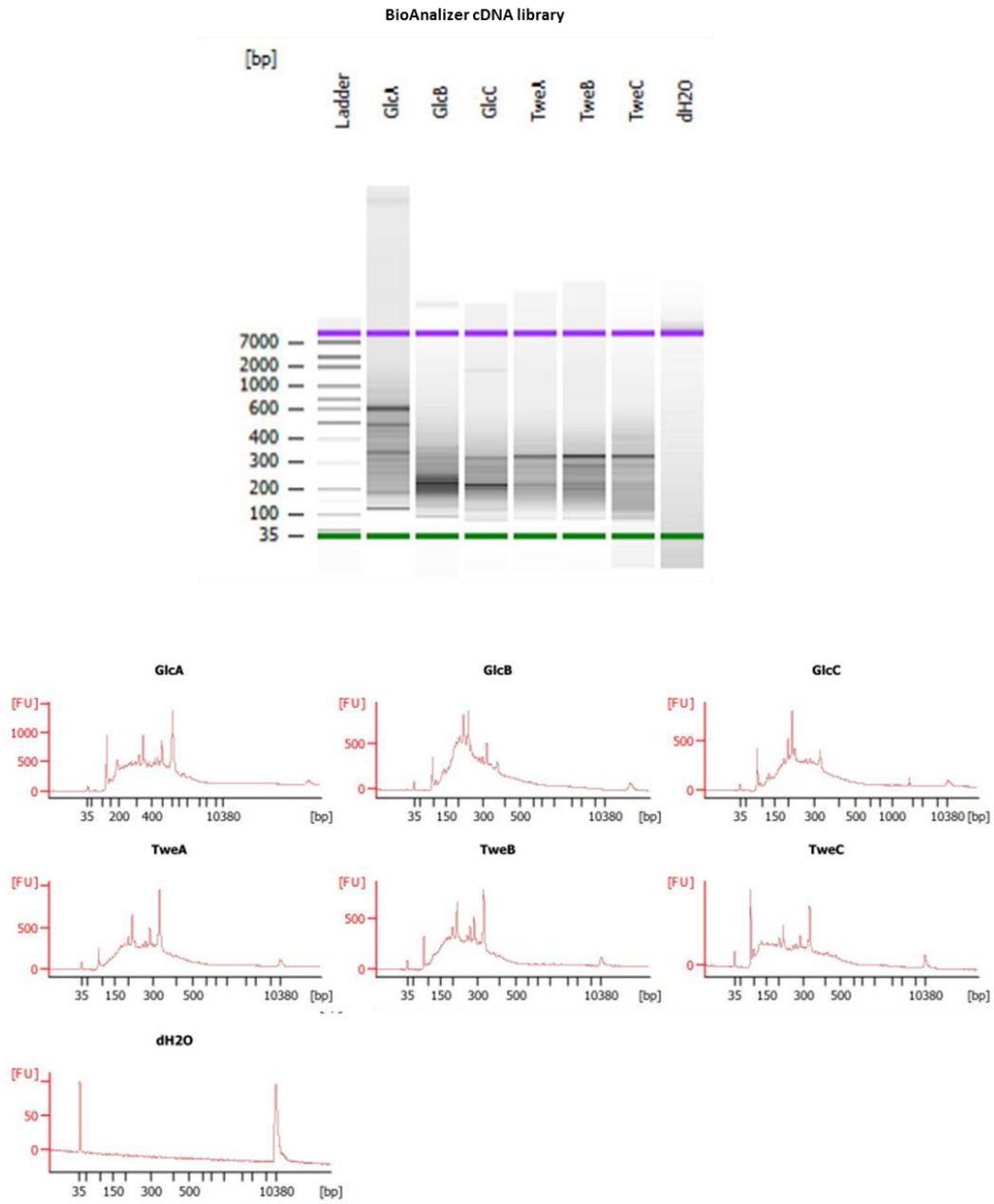



Figure 6.5 BioAnalizer traces using DNA High Sensitivity Kit (Agilent) after reverse transcription into cDNA and creation of a library. Top: Gel view of the run, Bottom: Chromatograms of each sample. Legend: GlcA–C = three samples from glucose cultures, TweA-C = three samples from Tween cultures

Each sample was loaded on to a 316v2 chip and sequenced using the Ion Torrent PGM (Life Technologies). Overall a 30x coverage was achieved for the sequencing experiment. The chip loading was of 86% for both runs. The sequencing resulted in 2.51 (glucose) and 2.53 (Tween) million total reads, however 46% and 47% of the reads were usable for glucose and Tween respectively. The number of polyclonal reads was high for both reactions at 42 and 43 %. The mean read length was of 120 and 126 bp and 90% and 91% of reads could be aligned (Figure 6.6 and Figure 6.7).

The FASTQC is a quality control software which assesses the base calling accuracy. It reports the Q score, which is a logarithmic relation to the base calling error probabilities (Ewing *et al.*, 1998; Ewing and Green, 1998).

$$Q = -10 \log_{10} P$$

A Q score of 30 (Q30) equals a probability of a base calling error of 1 in every 1000 bases (accuracy of 99.9 %), whereas a Q score of 20 (Q20) already equals a probability of 1 error in every 100 bases, which is an accuracy of 99 %. Ideally one expects the Q score to be above 30. Usually the longer a read then the base calling accuracy starts to drop more and more. In the two sequencing runs, the Q scores were just below Q30, ranging from Q29 to Q27 and started to drop further down to mid twenty Q scores from position 80 onwards (Figure 6.6 and Figure 6.7).

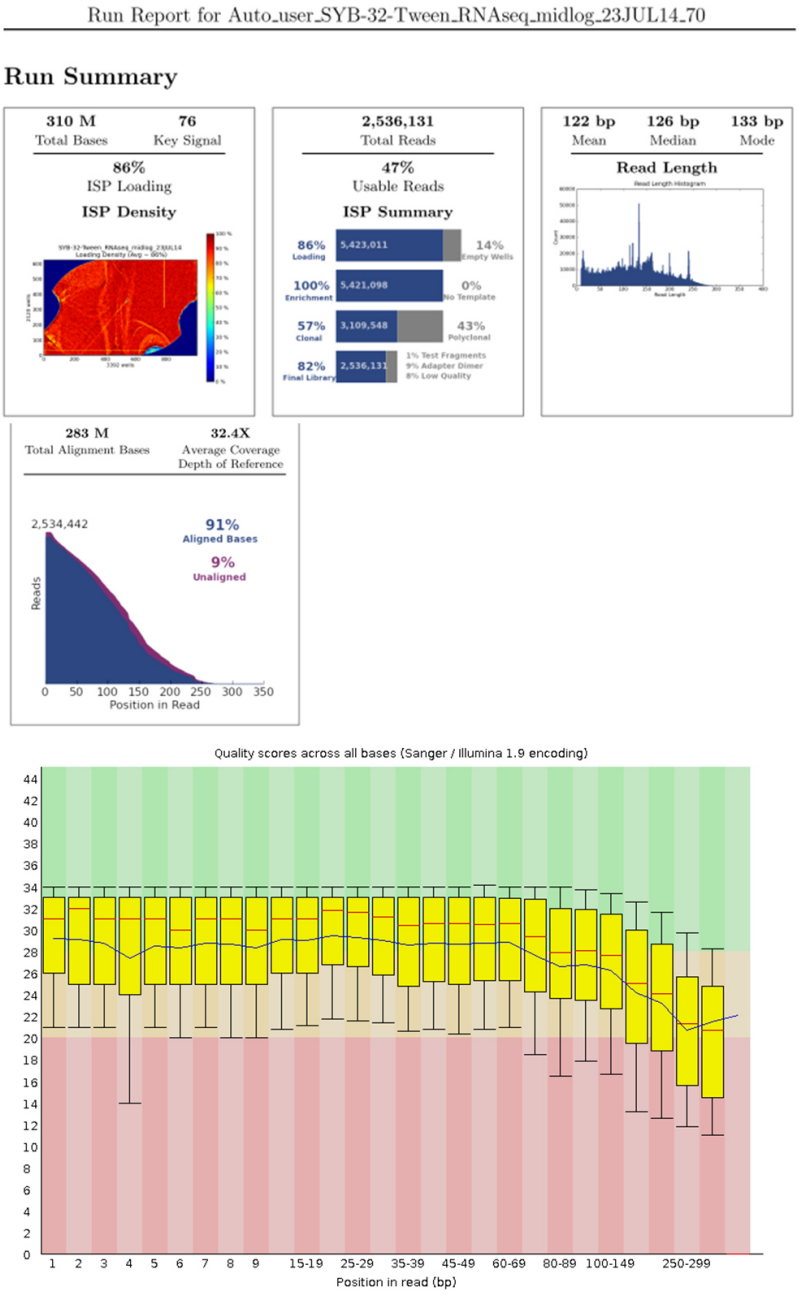



Figure 6.6 Summary of characteristics of the sequencing run for pooled glucose samples: Chip loading, number of usable reads, mean read length and alignment as well as FastQC plot for base calling accuracy obtained from the Ion Torrent Server Suite V4.0.2

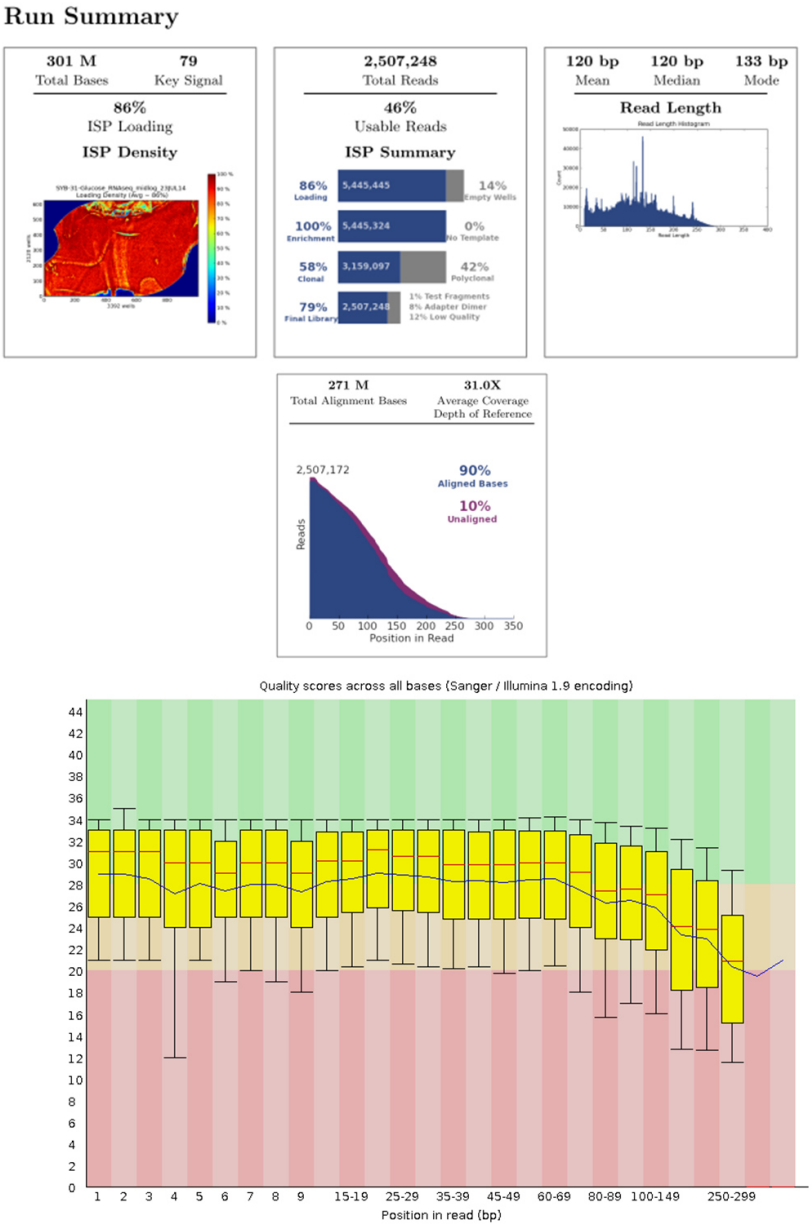



Figure 6.7 Summary of the characteristics of sequencing run for pooled Tween samples: Chip loading, number of usable reads, mean read length and alignment as well as FastQC plot for base calling accuracy obtained from the Ion Torrent Server Suite V4.0.2

Additionally, an internal quality control was used, as recommended by the manufacturer, which consisted of a set of 92 known polyadenylated transcripts that were added to the samples. They mimic a set of eukaryotic transcripts of 250-2000 nts that enable the assessment of performance of the experimental procedure. Each of the transcripts was blasted against the genome of *S. coelicolor* and the identity was always less than 10% except for SCO4139 with ERCC0095 which had a 186 bp stretch with 35% identity when using dissimilar sequences blast algorhithm (discontigous megablast, Cai *et al.*, 2009). However when using the PlugIn Analysis tool from Life Technologies it became obvious that the analysis was not reliable. When starting the analysis, there is the option whether or not to include a reference to which it should align. Depending on whether a reference was chosen or not, the ERCC PlugIn gave different results. When no reference was selected in the glucose samples, 34 ERCC transcripts were detected and 70 when selecting a reference genome. When the same analysis was performed on the Tween samples, the opposite was observed - when no reference was selected 68 of the ERCC transcripts were detected, but only 38 transcripts when selecting with reference (Figure 6.8). Life Technologies Support was contacted. However no satisfactory answer was received as they were unable to provide an answer for the observed results. Therefore inclusion of this analysis was discontinued for the remaining data analysis, as no conclusion could be drawn.

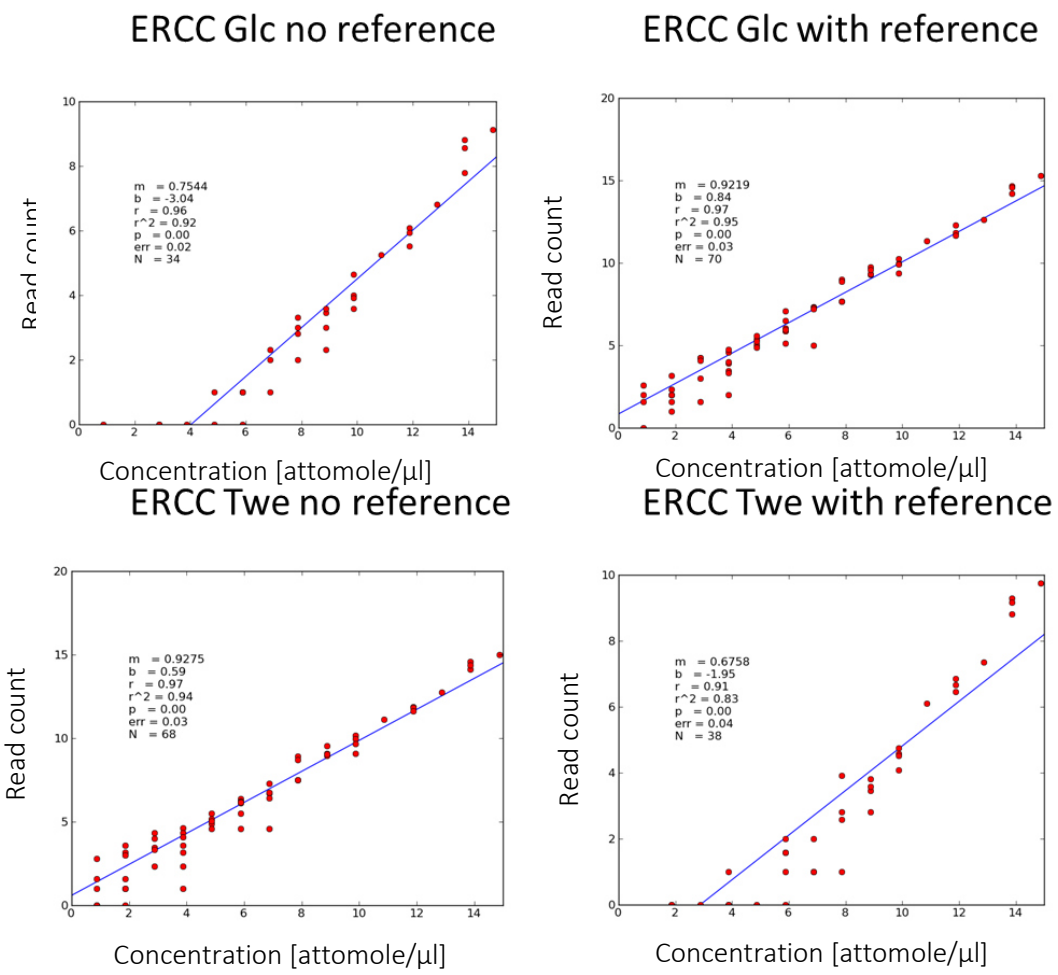



Figure 6.8 ERCC transcript detection for both runs comparing inclusion using a reference sequence or not

Genome sequence of *S. coelicolor* genome deposited in Genbank was used as reference (NC_003888.3). The data were analysed using three different bioinformatics tool pipelines: Rockhopper (McClure *et al.*, 2013), RNA Rocket (Goecks *et al.*, 2010; Warren *et al.*, 2015) and CLC Genomics Workbench (Qiagen). The lack of prokaryotic specific RNA-Seq tailored bioinformatics pipelines suggested that multiple analysis methods, followed by consensus analysis would yield the most accurate results. At the level of alignment of the reads to the genome, the different pipelines had different results. Rockhopper showed on average 67% aligned reads for glucose and 62.3% of the reads for Tween aligned compared to 75.6% and 72.6% compared in RNA Rocket respectively and 99.8% and 99.7% in CLC. The intrinsic alignment on the Ion Torrent server was of 97.3 and 96.3% respectively for glucose and Tween (Table 6-2) .Due to the poor alignment results of Rockhopper, it was decided to proceed only with RNA Rocket and CLC for the remaining data analysis.

Aligning all reads to the genome of *S. coelicolor* showed a dominant peak, which was found in all of the samples, representing about 25 % of all the reads in each sample. This peak was identified as the gene *ssrA*, a transfer-messenger RNA (tmRNA), which plays an important role in 70S ribosome recycling during stress (Mikulík *et al.*, 2008, Figure 6.9). However as the same was observed in all samples, it was decided to proceed with the analysis. One hypothesis is that the additional peak observed in the BioAnalizer trace following ribosomal RNA depletion, may equate to this highly abundant transcript (Figure 6.4).

The six samples were subjected to principal component analysis (PCA). A PCA is a statistical method, which analyses the most important influencing factors while retaining most of the variability on the dataset and plotting them accordingly (Ringnér, 2008). Basically, when analysing complex data sets, PCA filters those variables out that most influence the data, the so called principle components (PC) and can be used to plot the data on PC1 against PC2. It is a quick way to determine whether samples can be grouped. As expected the three replicates of each condition were clustered (Figure 6.10).

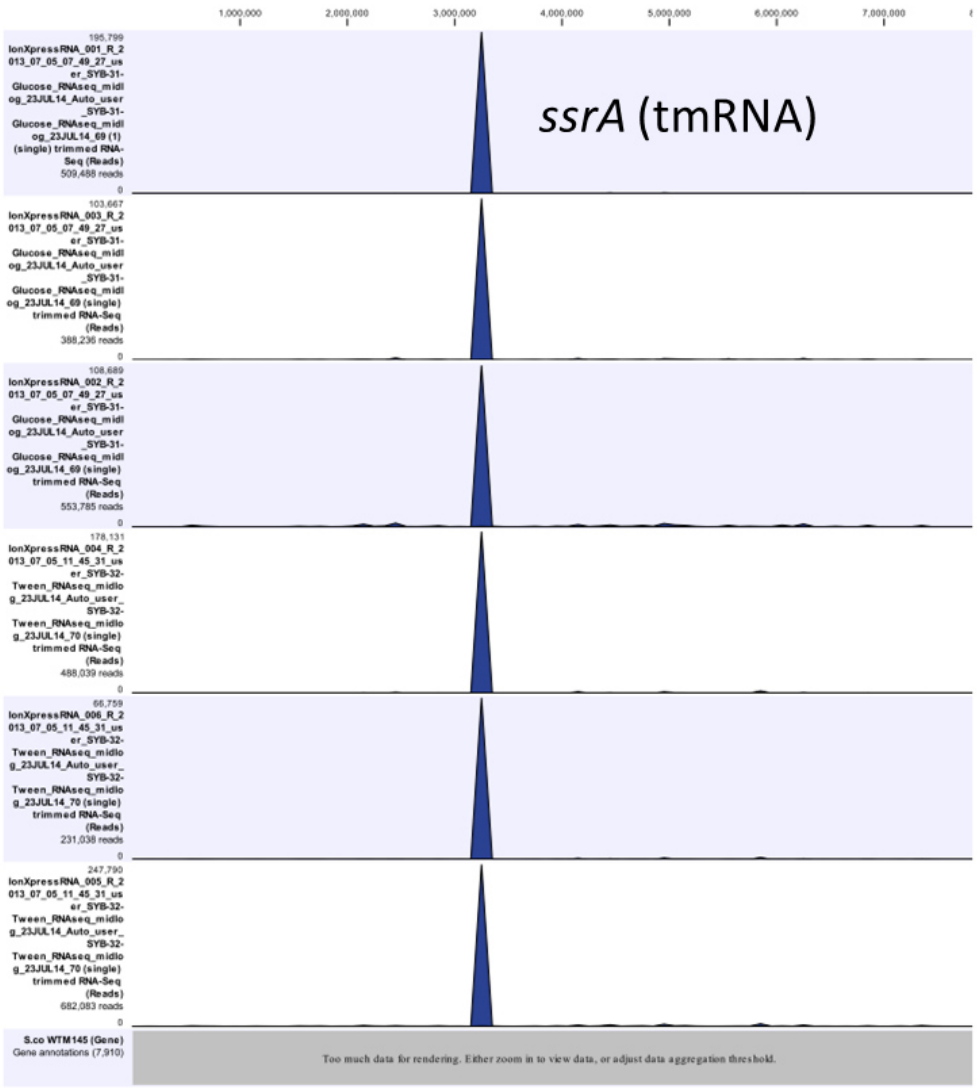



Figure 6.9 Overall alignment of reads for all six samples showing one big peak which is a transfer messenger RNA *ssrA* (Mikulík *et al.*, 2008)

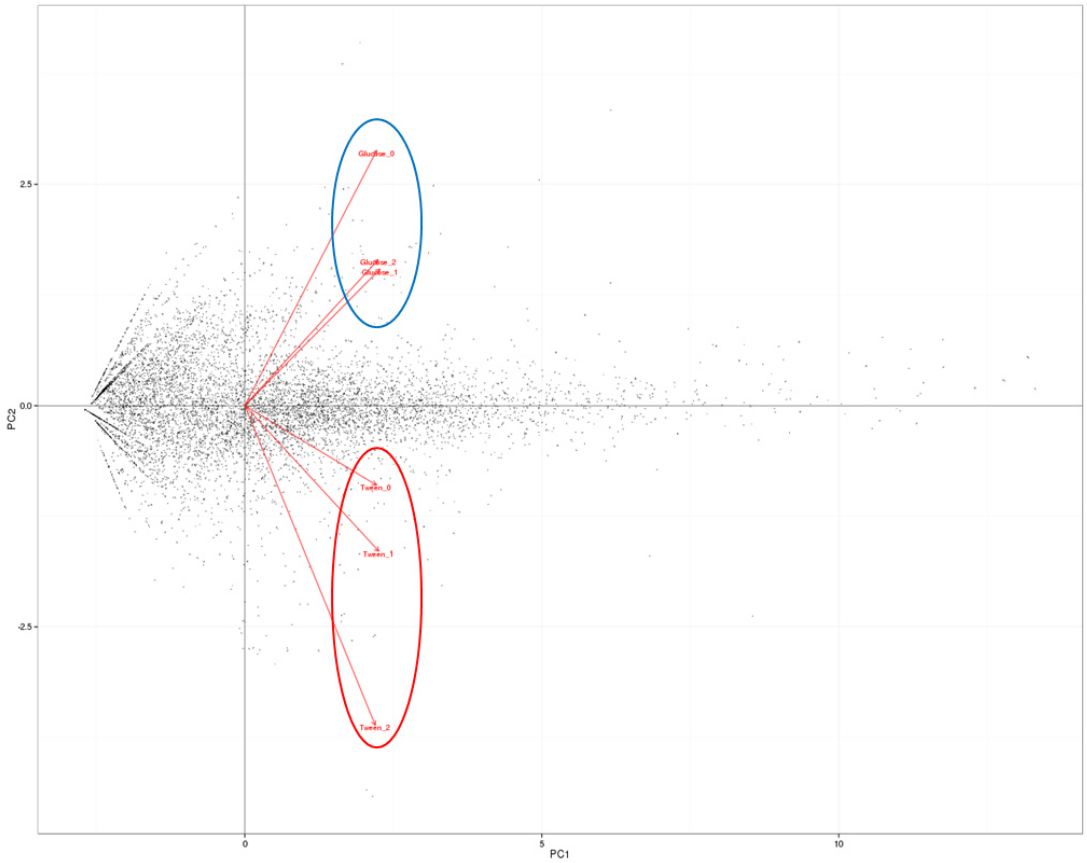



Figure 6.10 Principal component analysis of samples from RNA-Rocket analysis pipeline (Warren *et al.*, 2015) showing the six samples, blue circle is for glucose samples and the red for Tween

To examine differential gene expression analysis within the dataset, the cutoff was set at a p-value of 0.05. A p-value of 0.05 means that either the observation is real or it happened as coincidence as one in twenty attempts (Fisher, 1934). The CLC differential gene expression tool is based on the edgeR package (Robinson *et al.*, 2010) with was specifically designed to determine differential gene expression. Genwise dispersions are estimated by conditional maximum likelihood using the total count for the gene of interest followed by empirical Bayes to obtain a consensus value (Smyth and Verbyla, 1996; Robinson and Smyth, 2007). The differential expression is then assessed using Fisher's exact test adjusted to overdispersed data (Robinson and Smyth, 2008).

The alignment of reads for *pyk1* and *pyk2* for each is shown in Figure 6.11. The top row is glucose and the bottom Tween. Both *pyk1* and *pyk2* were more highly expressed during glucose growth than on Tween, in all samples *pyk2* had higher expression levels than *pyk1* (Figure 6.11).

RNA Rocket showed 526 differentially expressed genes on Tween versus glucose and CLC detected 644. Both shared 352 differentially expressed genes (Figure 6.14). At this point qPCR was performed on the samples for five different genes in order to validate the results obtained by the RNA-Seq.

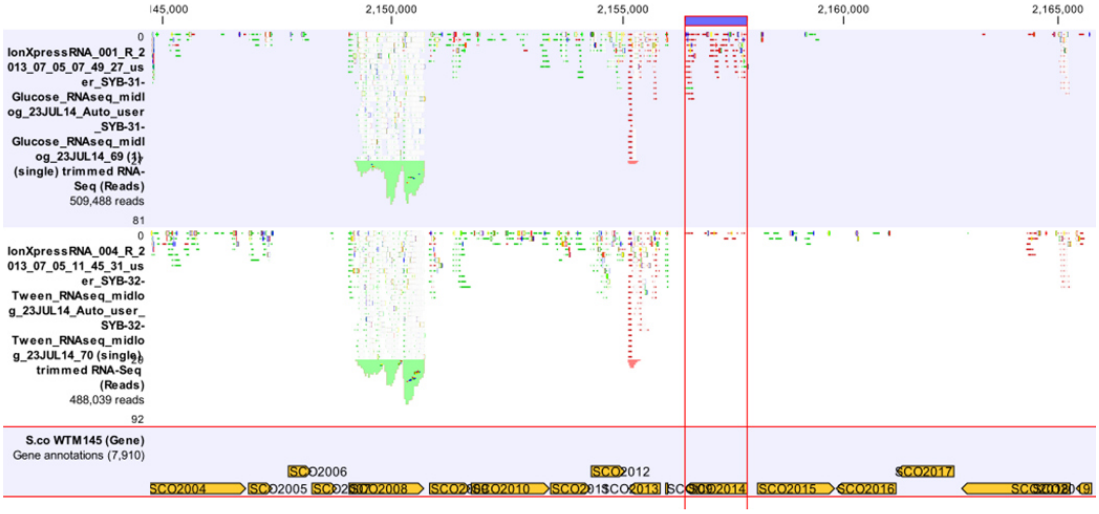
Table 6-1 RNA concentration determination during the sample preparation for the sequencing

| Sample | RNA isolation [ng/uL] | RIN | rRNA depletion [ng/uL/ total ng] | cDNA library (pM/L; 50 – 1000 bp) |
|---|---|---|---|---|
| **GlcA** | 336 | Nd | 10.00/ 70.0 | 246005.6 |
| **GlcB** | 324 | Nd | 9.93/ 69.5 | 138979.3 |
| **GlcC** | 372 | Nd | 6.12/42.8 | 102882.9 |
| **TweA** | 103 | 8.5 | 7.62/53.3 | 101099.6 |
| **TweB** | 124 | 8.7 | 6.63/46.4 | 84646.1 |
| **TweC** | 148 | Nd | 8.75/61.3 | 71439.0 |

Table 6-2 Alignment percentage of samples to reference genome in different analysis tools

| Sample | Rockhopper [%] | RNA Rocket [%] | CLC [%] | Ion Torrent [%] |
|---|---|---|---|---|
| **Glc A** | 65 | 74 | 99.8 | |
| **Glc B** | 70 | 77 | 99.9 | |
| **Glc C** | 67 | 76 | 99.7 | |
| **Mean Glc** | 67.3 | 75.6 | 99.8 | 97.3 |
| **Twe A** | 66 | 75 | 99.8 | |
| **Twe B** | 63 | 72 | 99.9 | |
| **Twe C** | 58 | 71 | 99.5 | |
| **Mean Twe** | 62.3 | 72.6 | 99.7 | 96.3 |

A *pyk1*




B *pyk2*

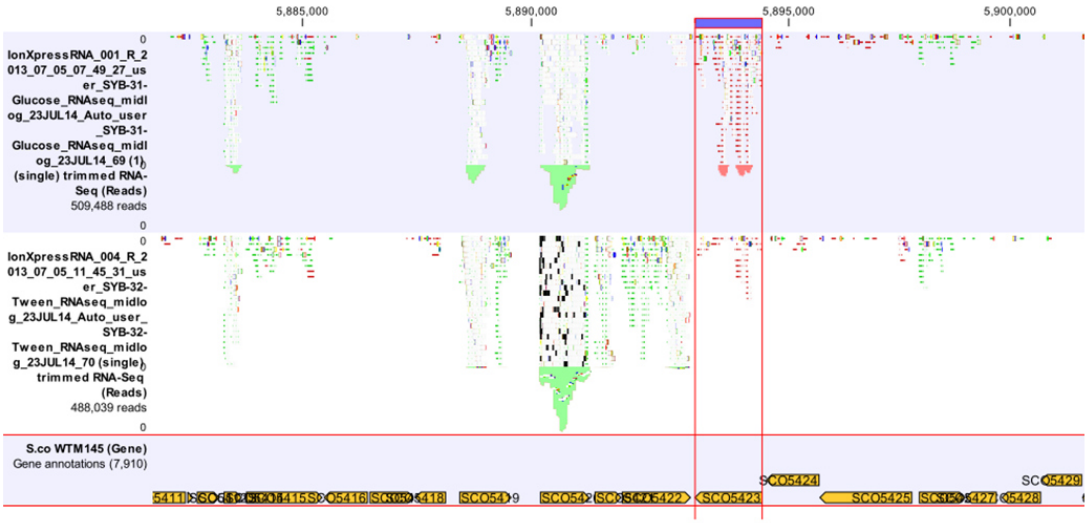



Figure 6.11 Example screenshot from CLC Genomics Workbench (Qiagen) showing the alignment of reads to reference genome for *pyk1* (A) and *pyk2* (B) for one of the three samples from glucose or Tween

## 6.2 CONFIRMATION OF RNA-SEQ EXPERIMENT DATA BY Q-PCR

To validate the data obtained from the RNA sequencing independently, qPCR was performed on five genes of *S. coelicolor* on the same samples and compared. As the interest was on comparing glycolysis and gluconeogeneic conditions, two genes from glycolysis were chosen (*pyk1* and *pyk2)* and two genes from gluconeogenesis (*ppdK1* and *ppdK2*), as well as *hrdB* as reference gene, which should be expressed at all times.

Samples from stationary phase were also analysed for the expression of the same genes. For qPCR it is important to have optimal PCR conditions in order to measure the actual copy number. To achieve this, the MIQE guidelines (**M**inimum **I**nformation for Publication of **Q**uantitative Real-Time PCR **E**xperiments) were followed for this (Huggett *et al.*, 2013). To ensure the assay was working optimally the primer concentration had to be optimised and the PCR efficiency needed to be determined.

The primer concentrations were optimised to decrease primer dimer formation. This was necessary as each primer pair would have different reaction kinetics, even between the forward and reverse primer for the same target template, as a result from their different sequences and affinities for their specific target sequences (Huggett *et al.*, 2013). Therefore five different concentrations of each primer were tested in all combinations: 1000 nM, 500 nM, 250 nM, 125 nM and 62.5 nM (Table 6-3).

For optimisation runs, purified PCR product from genomic DNA was used as a template (2.4 ng per reaction) for each target, as well as a no template control (RNase free $dH_2O$) for each of the concentrations. Each reaction was run in duplicates. To select the optimal concentrations, the amplification curves and the melt curves were analysed. The amplification was measured as cycle threshold (Ct) and is defined as number of cycle required for the fluorescent signal to cross the threshold and is inversely proportional to amount of DNA in the sample (Morrison *et al.*, 1998). The melt curve analyses the dissociation of a double-stranded DNA, generated by plotting fluorescence of a DNA specific dye as function of temperature. The melt curve is influenced by GC/AT ratio, length and sequence and can be used to distinguish undesired products (Wittwer *et al.*, 1997). For each of the primer concentrations both Ct and melt curves were analysed and primers were selected based on a steep signal in the amplification at a low Ct plus a high confined peak in the melt curve.

It was found that all *hrdB* primer concentrations tested were very similar in their Ct values, with all of them between 7 -9 cycles. The melt curve peaks were also similar and showed their maximum values at 85˚C. In this case the primer concentration did not seem to interfere with the reaction. The no template control (NTC) started to show a signal at cycle 25 and exhibited a similar high signal in the melt curve compared to the sample melt curves. Following the Ct and melt curve analysis, a concentration of 1000 mM for the forward primer and 250 mM for the reverse primer was chosen (Figure 6.12, Table 6-4).

The different *pyk1* primer pairs showed a similar Ct for all the primer concentrations tested and the melt curves were similar with their maximum at 87.5˚C. The no template control (NTC) Ct value was at 24/26 cycles and showed a peak at the same melting T as the samples. Primer concentrations of 1000 mM for the forward primer and 250 mM for the reverse primer were chosen (Figure 6.12, Table 6-4).

The *pyk2* primer pairs showed two clusters at six cycles or eight cycles as Ct. Thus a concentration pair was chosen at six cycles with a melt curve showing a confined maximum. This was a concentration of 1000mM for the forward and 250 mM for reverse primer. The no template control (NTC) started to give a signal after 20 cycles. However the melt curve showed a different peak maximum than the samples- 80˚C compared to 90˚C (Figure 6.12, Table 6-4).

The primers for *ppdK1* (SCO0208) showed one outlier in the Ct analysis. All other concentrations were at 6 cycles and the no template control (NTC) started to give a signal at 28 to 30 cycles. The melt curves peak maximum was at 94˚C for the samples and 80˚C for the no template control (NTC). The latter also showed a broad and small peak in the melt curve. The concentrations of forward and reverse primer were chosen at 1000 and 250 mM respectively (Figure 6.12, Table 6-4).

The primer for *ppdK2* (SCO2494) showed a range of 7 to 10 cycles for the Ct for all different primer concentrations while no template control (NTC) showed a signal at 26/30 cycles. The melt curve of no template control (NTC) looked different to the samples, the peak was much lower and at 77˚C compared to the one of the samples: which showed a maximum at 90 ˚C. Concentrations of 250 mM and 1000 mM for the forward and reverse primer were chosen (Figure 6.12, Table 6-4).

The optimisation of primer concentration it was necessary to determine the PCR efficiency of each of the reactions, which were also used as standard curves in order to determine the copy number of each gene.

Each gene was amplified by PCR and quantified and then diluted to copy number concentrations of $1x10^1$ to $10^7$. Each standard curve was performed in triplicates and log concentration versus Ct was plotted. The slopes were used to quantify the PCR efficiency which should be as close as possible to 100%. The plots for each gene are shown in Figure 6.13. *hrdB* had an efficiency of 89.2%, *pyk1* of 90.3%, *pyk2* of 100.5%, *ppdK1* of 84.8% and *ppdK2* of 87.9% (Table 6-5).

The experimental samples were subsequently analysed in duplicate with 10 ng of cDNA as template. The results confirmed the observations of the RNA-Seq experiment. Changes in the housekeeping sigma factor, *hrdB* were not statistically significant in any of the analyses and were found to be -1.18 fold in CLC, 0.94 fold in RNA Rocket and -2.3 in the qPCR when Tween grown cultures were compared with glucose grown cultures. This was expected as *hrdB* is known to be constitutively expressed (Kelemen *et al.*, 1996; Clark and Hoskisson, 2011). Analysis of *pyk1* was -7.9/-7.16 in CLC/RNA Rocket compared to -8.1 fold in the qPCR. *pyk2* showed a -6.5/6.8 and -4.5 fold decrease in expression respectively. The two genes *ppdK1* and *ppdK2* encoding PPDK also showed a similar increase in expression 11.3/13.6 fold on Tween compared to glucose when analysed by CLC and RNA-Rocket and a 12.3 fold in the qPCR for *ppdK1*. *ppdK2* showed 30.8 fold increase in CLC or 33.98 in RNA-Rocket and 25.7 in the qPCR. For all genes of interest the same overall trends were observed by qPCR as for the RNA-Seq analysis. Remarkably, the only exception was *hrdB*, where RNA Rocket showed an increase whereas the other two analysis methods showed a decrease in expression, but these were found not to be statistically significant (Table 6-10, Figure 6.14). However since the analysis of the qPCR was more similar to the results of the analysis from CLC, the analysis method using CLC was used for the detailed global analysis gene expression (Figure 6.14).

Table 6-3 Primer combinations to find optimal concentrations for each primer pair

| Reverse Primer (nM) | Forward Primer (nM) 1000 | 500 | 250 | 125 | 62.5 |
|---|---|---|---|---|---|
| **1000** | 1000/1000 | 1000/500 | 1000/250 | 1000/125 | 1000/62.5 |
| **500** | 500/1000 | 500/500 | 500/250 | 500/125 | 500/62.5 |
| **250** | 250/1000 | 250/500 | 250/250 | 250/125 | 250/62.5 |
| **125** | 125/1000 | 125/500 | 125/250 | 125/125 | 125/62.5 |
| **62.5** | 62.5/1000 | 62.5/500 | 62.5/250 | 62.5/125 | 62.5/62.5 |

Table 6-4 Primer concentrations that were used for the sample runs selected from the optimisation runs

| Gene | Forward primer [mM] | Reverse primer [mM] |
|---|---|---|
| ***hrdB* (SO5820)** | 125 | 500 |
| ***pyk1* (SCO2014)** | 1000 | 250 |
| ***pyk2* (SCO5423)** | 1000 | 250 |
| ***ppdK1* (SCO0208)** | 1000 | 250 |
| ***ppdK2* (SCO2494)** | 250 | 1000 |

Table 6-5 Efficiency of PCR reactions of the genes of interest

| Gene | Slope | Mean Slope | Efficiency (Mean) |
|---|---|---|---|
| ***hrdB* (SO5820)** | -3.47/-3.48/-4.06 | -3.58 | 89.2 % |
| ***pyk1* (SCO2014)** | -3.70/-3.07/3.41 | -3.58 | 90.3 % |
| ***pyk2* (SCO5423)** | -3.69/-2.95/-3.22 | -3.35 | 100.5 % |
| ***ppdK1* (SCO0208)** | -4.30/-2.80/-3.80 | -3.75 | 84.8 % |
| ***ppdK2* (SCO2494)** | -3.68/-3.50/-3.09 | -3.65 | 87.9 % |

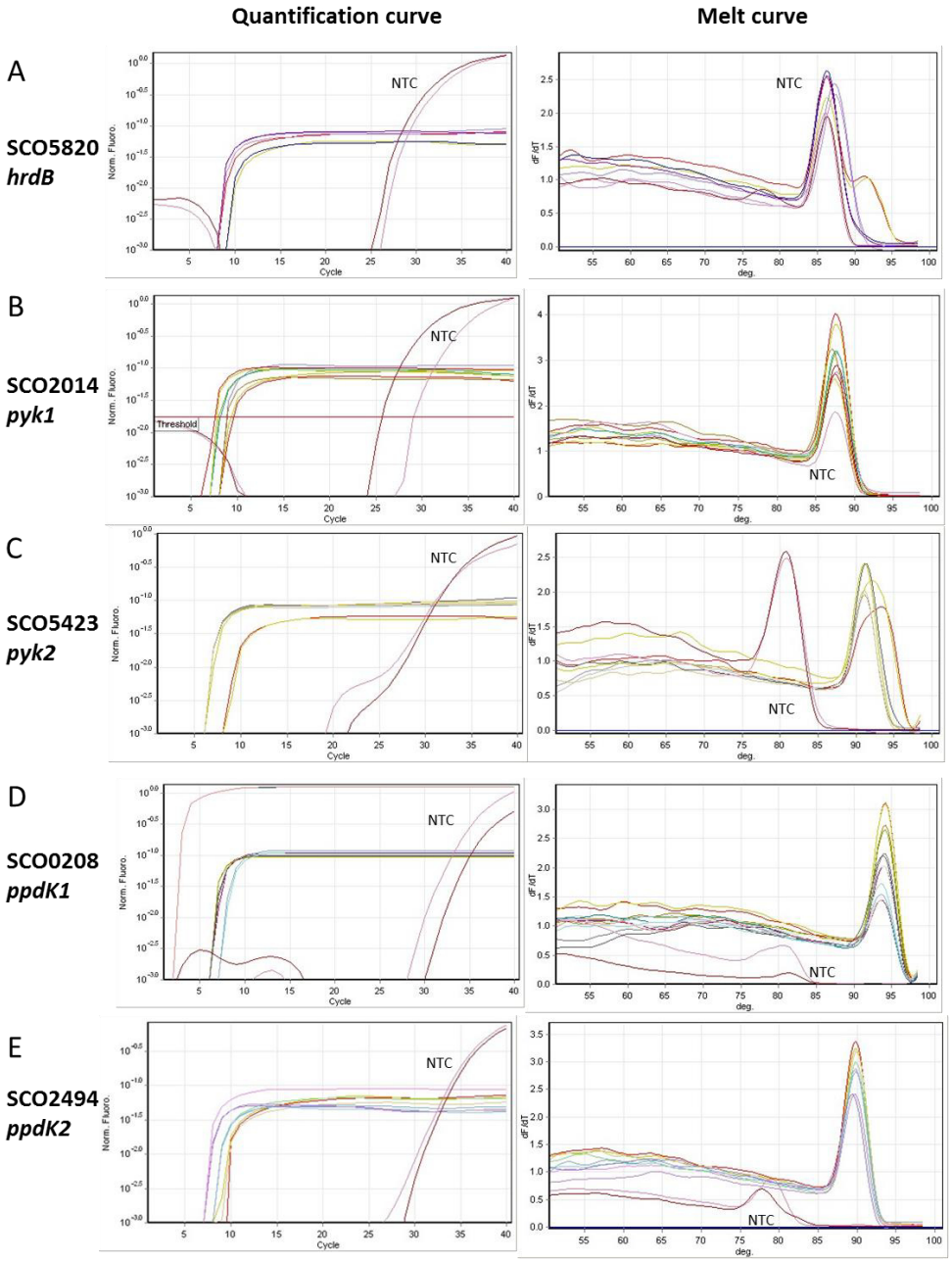



Figure 6.12 Quantification of Ct and melt curve for different primer concentrations for primer pairs of the five genes analysed in order to find the optimal PCR conditions. Legend: NTC = no template control

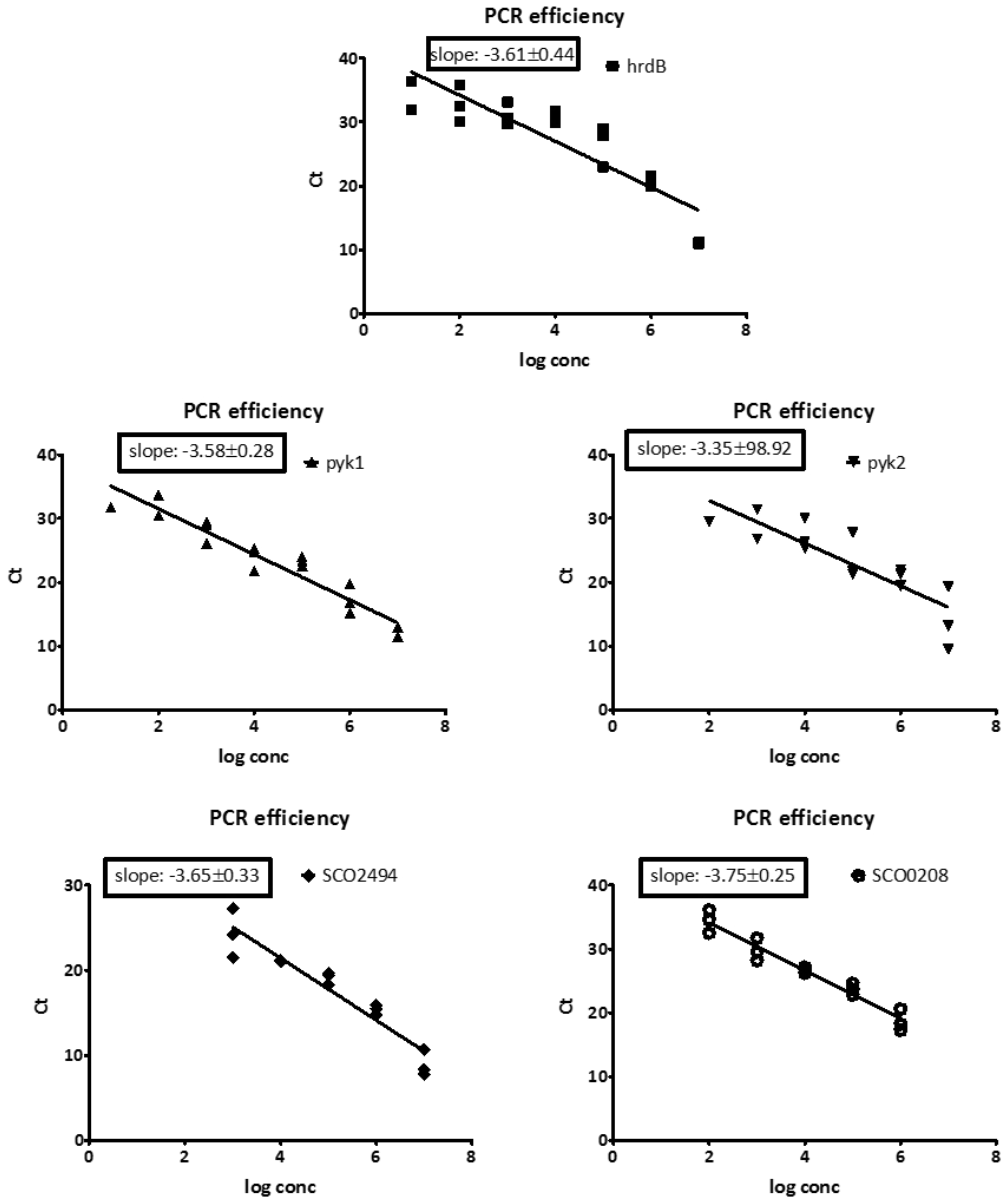
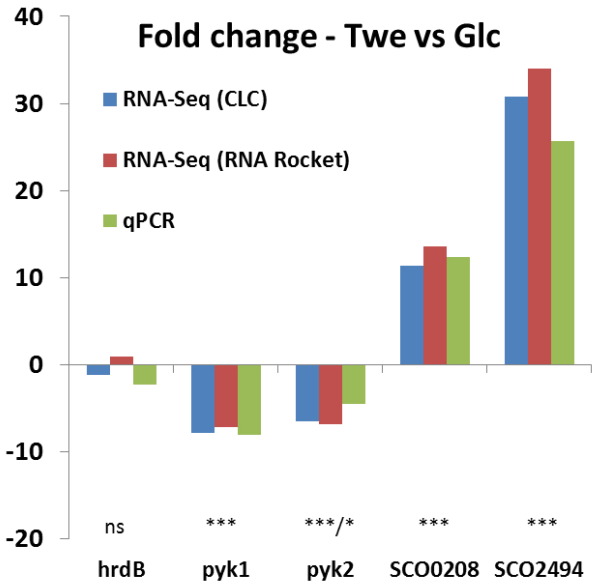



Figure 6.13 PCR efficiency for each of the five genes using a standard curve of known concentrations of the amplicon



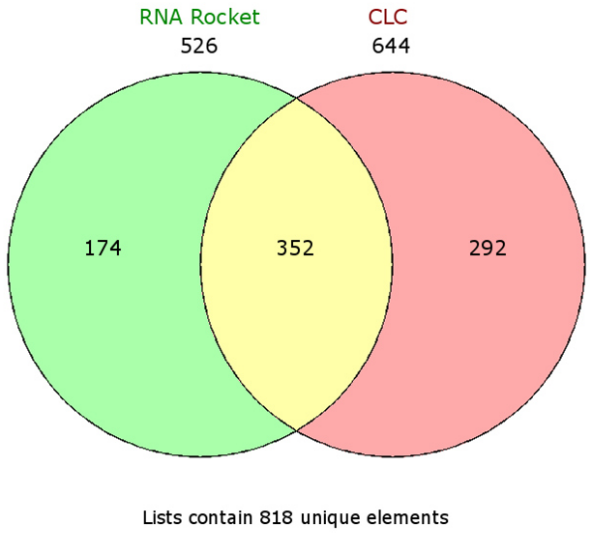




Figure 6.14 Top: Comparison of fold changes result from RNA Rocket, CLC genomics and qPCR for the genes *hrdB, pyk1, pyk2, ppdK1* and *ppdK2*. Bottom: Venn Diagram of number of differentially expressed genes found in RNA Rocket and CLC Genomics Workbench

## 6.3 Global Differential Gene Expression Analysis on Tween and Glucose as Sole Carbon Sources

To characterise the global gene expression response to cells growing under glycolytic or gluconeogenic conditions, CLC Genome Workbench approach described above was used, which showed most congruence with the qPCR validation.

In total 644 genes showed a significant change in expression on Tween compared to glucose, of which 61 were regulatory genes (9.3%), 251 were hypothetical genes (39%) and some metal and nitrogen metabolism genes were expressed differentially. Generally it was observed that genes for gluconeogenesis, fatty acid degradation, isorenieratane biosynthesis were up-regulated, whereas genes for glycolysis, fatty acid biosynthesis and actinorhodin were down-regulated.

The differential expression analysis of the metabolism of *S. coelicolor* on different carbon sources was used to analyse the expression of expanded central carbon metabolism genes (Table 6-6). In the expansion analysis (Chapter 3) 14 genus specific expansion were detected in *Streptomyces* and a further three species specific in *S. coelicolor* (Table 3-2), those were studied with particular interest in order to establish whether they were expressed differently on glucose and Tween. Again, the genes were organised by pathway, in total 34 functional roles were included of which 21 showed two or more genes annotated for this function.

**Glycolysis**

In glycolysis the genes encoding glucose kinase, enolase, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), phosphofructokinase (Pfk), phosphoglucose isomerase, pyruvate kinase (Pyk) and triosephosphate isomerase showed differential gene expression profiles in addition to several genes annotated with the same function. Four of these enzymes were found to be expanded in *Streptomyces*: enolase, GAPDH, Pfk and Pyk.

Of the two potential glucose kinase encoding genes, only glucose kinase (SCO0063) showed a change in expression, a 5-fold increase on Tween whereas the *glkA* (SCO2126) did not show any change (Table 6-6).

The two genes that encode enolase (SCO3096 and SCO7638) showed a decrease in expression on Tween, however SCO3096 exhibited a 2-fold decrease in expression and the other 11 fold (SCO7638, Table 6-6).

For GAPDH, where there are three genes annotated with that putative function, two showed a decrease in expression: SCO1947 almost 3-fold and SCO7511 a 19-fold decrease. The third putative GAPDH (SCO7511) showed a ten-fold increase in gene expression (Table 6-6).

Only one of the three phosphofructokinase genes (SCO1214) exhibited a change in expression, with a three-fold decrease in expression (Table 6-6).

For the two genes encoding phosphoglucose isomerase, SCO6659 and SCO1942, only SCO6659 showed almost seven fold decrease in expression, whereas SCO1942 showed no significant change among the two carbon sources (Table 6-6).

In case of the pyruvate kinases, *pyk1* and *pyk2*, both were expressed lower on Tween compared to glucose - almost eight fold for *pyk1* and 6.5-fold for *pyk2* (Table 6-6).

The two genes encoding triosephosphate isomerase were also differentially expressed on Tween compared to glucose and actually opposed each other: SCO1945 is two-fold decreased in Tween and SCO5366 is two-fold increased on Tween (Table 6-6).

**Gluconeogenesis and Pentose-Phosphate-Pathway**

The expression of genes required for gluconeogenesis was increased on Tween, as would be expected, since gluconeogenesis is required to obtain glycolytic intermediates for biosynthesis from TCA cycle intermediates. The gene encoding FBP bisphosphatase exhibited a 2.8- fold increase in expression and *ppck* showed a 27-fold increase in gene expression. The genes encoding putative PpdK activity were also found to increase in expression, with *ppdK1* showing an 11-fold increase and *ppdK2* showing a 31 fold increase in expression. This enzyme was also found to be expanded in the genus of *Streptomyces* (Table 3-2). Moreover, the gene encoding pyruvate carboxylase (SCO0546) exhibited a 3.5-fold increase in expression. Only PEP synthase (SCO5896) showed no significant change in expression (Table 6-6).

The genes required for pentose phosphate pathway activity showed a change in expression in three out of the eight reactions. There was a three-fold decrease in expression in *zwf1* gene (SCO6661), whereas SCO1937 (*zwf2*) showed no significant change (Table 6-6).

Two of the three 6-phosphogluconate dehydrogenase genes showed two-fold (SCO3877) and 7-fold decrease (SCO6658) in expression, whereas SCO0975 showed no change in expression.

Only one of the four genes annotated for transketolase activity (SCO6768) showed differential expression with a two-fold increased on Tween (Table 6-6).

**TCA cycle**

The TCA cycle also exhibted differential expression of key genes when *S. coelicolor* was grown on either glucose or Tween. The two genes annotated as succinate dehydrogenase (SCO0922 and SCO0924) were both expressed around two-fold higher expressed on Tween than on glucose. One of the putative fumarase genes (SCO5042) showed a two-fold expression decrease. The malic enzyme SCO2951 showed a two-fold decrease in expression. Two of the six $PDHC_{E1}$ showed differential expression: SCO2183 ($PDHC_{E1}$) had a 3.5-fold decrease in expression whereas SCO1270 ($PDHC_{E1\alpha}$) showed six-fold increase of expression on Tween. This was another enzyme which had be found to be expanded in *Streptomyces* (Table 3-2). The other genes showed no significant changes on the two different media (Table 6-6).

Table 6-6 Differential gene expression (DE) on Tween versus glucose in central carbon metabolism showing all genes annotated for the function

| **Glycolysis** | | **# genes** | **DE** | **fold change (T vs G)** | | | **SCO** |
|---|---|---|---|---|---|---|---|
| **Glk** | glucose kinase | 2 | 1 | 5.1 | - | | 0063, 2126 |
| **Eno** | enolase | 2 | 2 | -2.2 | -11.1 | | 3096 , 7638 |
| **Fba** | FBP aldolase | 1 | 0 | - | | | 3649 |
| **GAPDH** | G3P dh | 3 | 3 | -2.8 | 10.6 | -19.3 | 1947, 7040, 7511 |
| **Pfk** | phosphofructokinase | 3 | 1 | -3.1 | - | - | 1214, 2119, 5426 |
| **Pgi** | phosphoglucose isomerase | 2 | 1 | -6.7 | | | 6659, 1942 |
| **Pgk** | Phosphoglycerate kinase | 1 | 0 | - | | | 1946 |
| **Pgm** | phosphoglycerate mutase | 6 | 0 | - | | | 4209, 2299, 2576, 4470, 6818, 7219 |
| **Pyk** | pyruvate kinase | 2 | 2 | -7.9 | -6.5 | | 2014, 5423 |
| **Tpi** | triosephosphate isomerase | 2 | 2 | -1.9 | 2.2 | | 1945, 5366 |
| **Gluconeogenesis** | | **# genes** | **DE** | **fold change (T vs G)** | | | **SCO** |
| **FBPase** | FBP bisphosphatase | 1 | 1 | 2.8 | - | | 5047 |
| **PEPCk** | PEP carboxykinase | 1 | 1 | 26.7 | - | | 4979 |
| **PEPS** | Pep synthase | 1 | 0 | - | | | 5896 |
| **PPDK** | pyruvate phosphate dikinase | 2 | 2 | 11.4 | 30.8 | - | 0208, 2494 |
| **Pyc** | pyr carboxylase | 1 | 1 | 3.5 | - | | 0546 |
| **PPP** | | **# genes** | **DE** | **fold change (T vs G)** | | | **SCO** |
| **zwf** | glucose-6-phosphate dh | 2 | 1 | -3 | - | | 6661, 1937 |
| **Pgl** | 6-phosphoglucono-lactone lactonase | 1 | 0 | - | | | 1939 |
| **GhD** | 6-phosphogluconate dh | 3 | 2 | -1.9 | -6.5 | | 3877, 6658, 0975 |
| **Rpe** | ribose-5-phosphate epimerase | 1 | 0 | - | | | 1464 |
| **Rpi** | ribose-5-phosphate isomerase | 3 | 0 | - | | | 0579, 1224, 2627 |
| **Tal** | Transaldolase | 2 | 0 | - | | | 1936, 6662 |
| **Tkt** | Transketolase | 4 | 1 | 2.2 | - | | 6768, 6663, 6497, 1935 |
| **TCA** | | **# genes** | **DE** | **fold change (T vs G)** | | | **SCO** |
| **Cs** | Citrate synthase | 4 | 0 | - | | | 2736, 4388, 5831, 5832 |
| **Aco** | Aconitase | 1 | 0 | - | | | 5999 |
| **Idh** | Isocitrate dh | 1 | 0 | - | | | 7000 |
| **AKGdh** | Alpha KG dh (only E1) | 1 | 0 | - | | | 1268, 2181, 7123 |
| **Suc** | succinyl-CoA synthetase | 4 | 0 | - | | | 4808-9, 6568-9 |
| **Sdh** | succinate dh | 4 | 1 | 2.1 | 1.8 | | 0922, 0924, 0923, 4855-8, 5106-7, 7109 |
| **Fum** | Fumarase | 2 | 1 | -2.3 | - | | 5042, 5044 |
| **Mdh** | malate dh | 1 | 0 | - | | | 4827 |
| **me** | malic enzyme | 2 | 1 | -1.8 | - | | 2951, 5261 |
| **Icl** | Isocitrate lyase | 1 | 0 | - | | | 982 |
| **Ms** | malate synthase | 2 | 0 | - | | | 983, 6243 |
| **$PDH_{E1}C$** | E1 of pyruvate dh complex | 6 | 2 | -3.5 | 6.2 | - | 2183, 1270, 1269, 7124, 2371, 3816/17, 3830/31 |

**PEP-PYR-OAA-Node**

When analysing the metabolic node of interest in detail, there was no change in expression on different carbon sources for malate dehydrogenase, one of the malic enzymes (SCO5261), pyruvate dehydrogenase and PEP synthase. The genes involved in gluconeogenesis such as *ppck*, *pyc* and *ppdK1/2* all showed an increase in gene expression along with $PDHC_{E1\alpha}$. A decrease in gene expression was observed for *pyk1/2, ppc, me-NAD* and $PDHC_{E1}$ (SCO2183, Figure 6.15). Since no general trend could be observed for the node it seemed likely that the node is involved in many different pathways and reinforces the idea that the PEP-PYR-OAA node serves as a flux junction. In addition, it is clear from the work here on pyruvate kinase and the work of others that subtle kinetic differences play a role in regulation pathway activity, rather than transcriptional control. These data do not reveal any insight into the redundancy of certain genes, for both genes encoding Pyk and PpdK, as genes behaved similarly at the level of transcription. These data indicate that both paralogues are either both down regulated (*pyk*) or both up regulated (*ppdK*). The PDHC only showed differential expression for two of the 14 genes.

**Differential expression analysis of other parts of metabolism**

Pathways other than central carbon metabolism were identified as being differentially expressed - as expected the fatty acid utilisation genes were expressed higher on Tween, especially cholesterol esterase (SCO5420) which had an increase of almost 37 fold. The enzyme catalyses the hydrolysis of the head group of Tween and the fatty acid palmitate, which are then utilised as carbon source (Plou *et al.*, 1998; Pratt *et al.*, 2000; Sakai *et al.*, 2002). The other genes showed between 3-12-fold increase of expression.

Isorenieratene biosynthesis (SCO0185-0191) was another pathway that was expressed higher on Tween than on glucose, showing between 3- 8 fold increase in expression. Isorenieratene is a carotenoid normally produced under blue light induction, the pathway is being found in green photosynthetic bacteria and a few actinobacteria (Krügel *et al.*, 1999; Takano *et al.*, 2005). Additionally a NRPS pathway (SCO6429-6438) and the actinorhodin cluster (SCO5071-5092) were both decreased in their expression on Tween compared to glucose by 2-5-fold and 3- 11-fold respectively (Table 6-7).

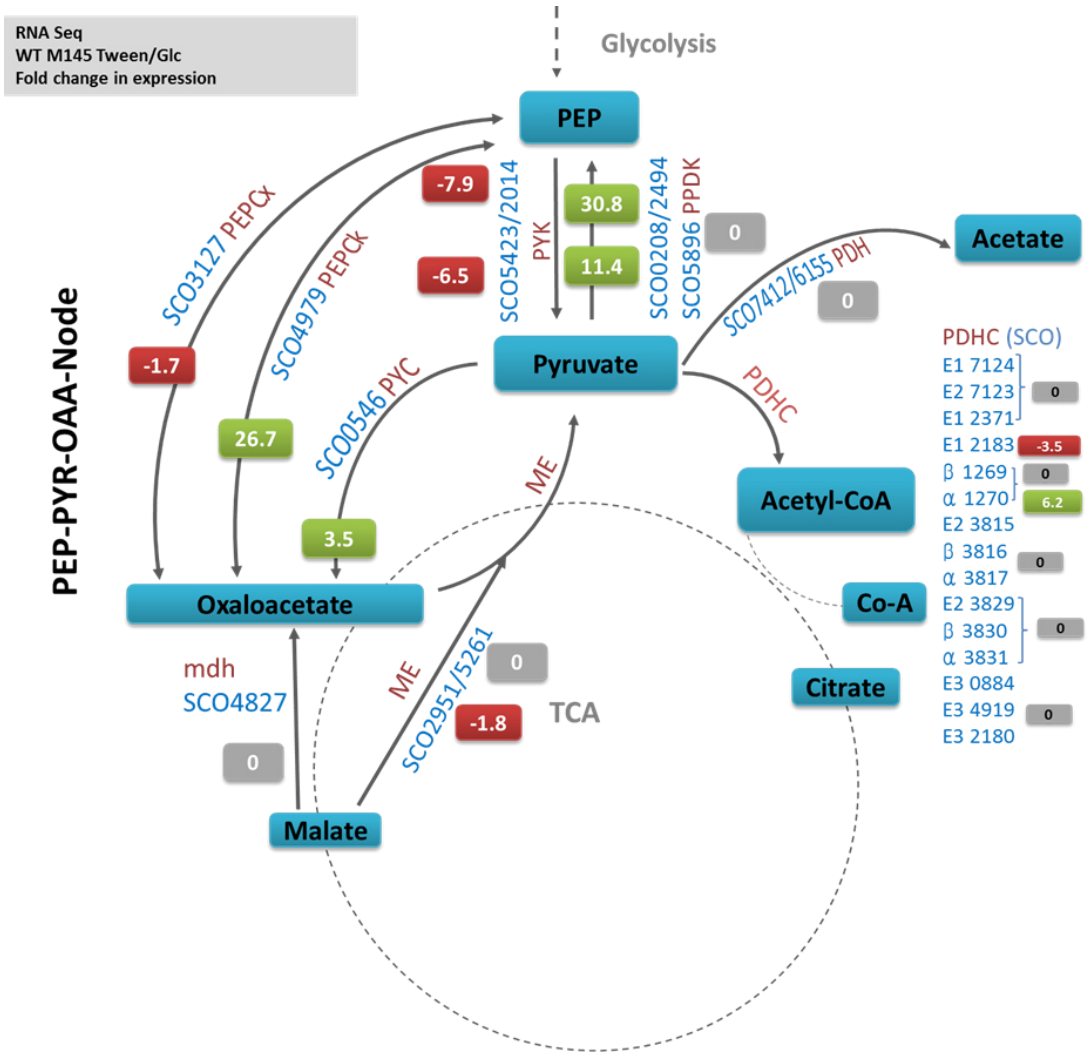



Figure 6.15 Overview of gene expression fold changes in the PEP-PYR-OAA node comparing Tween and glucose cultures

Table 6-7 Pathways with individual genes showing differential expression on Tween and Glucose given SCO number, gene function, fold change and p-value

| **Fatty acid utilisation** | | **Fold change** | **p-value** |
|---|---|---|---|
| SCO2131 | long chain fatty acid CoA ligase | 4.066 | 8.34E-08 |
| SCO2774 | acyl-CoA dehydrogenase | 7.418 | 5.59E-32 |
| SCO3079 | acetyl-CoA acetyltransferase | 4.745 | 4.95E-15 |
| SCO5420 | cholesterol esterase | 36.914 | 8.74E-93 |
| SCO6027 | acetyl-CoA acetyltransferase | 7.768 | 1.08E-34 |
| SCO6732 | fatty acid oxidative multifunctional enzyme | 3.791 | 1.21E-04 |
| SCO6026 | fatty acid oxidation complex alpha-subunit | 12.96 | 8.85E-103 |
| **Isorenieratene biosynthesis** | | **Fold change** | **p-value** |
| SCO0185 | geranylgeranyl pyrophosphate synthase | 8.825 | 2.52E-09 |
| SCO0186 | phytoene dehydrogenase | 5.551 | 4.17E-07 |
| SCO0187 | phytoene synthase | 4.294 | 5.07E-03 |
| SCO0188 | Methylesterase | 3.618 | 0.011 |
| SCO0191 | lycopene cyclase | 3.129 | 0.041 |
| **NRPS** | | **Fold change** | **p-value** |
| SCO6429 | hypothetical protein | -5.075 | 9.30E-19 |
| SCO6430 | hypothetical protein | -2.946 | 1.06E-06 |
| SCO6431 | peptide synthase | -3.279 | 1.72E-08 |
| SCO6432 | peptide synthase | -2.738 | 1.78E-04 |
| SCO6433 | hypothetical protein | -2.101 | 9.51E-03 |
| SCO6434 | Oxidoreductase | -2.152 | 0.019 |
| SCO6436 | tRNA synthetase | -2.592 | 1.69E-03 |
| SCO6437 | hypothetical protein | -3.221 | 5.24E-03 |
| SCO6438 | diaminopimelate decarboxylase | -2.041 | 8.70E-03 |
| **Actinorhodin** | | **Fold change** | **p-value** |
| SCO5071 | hydroxylacyl-CoA dehydrogenase | -4.529 | 1.12E-05 |
| SCO5072 | hydroxylacyl-CoA dehydrogenase | -2.995 | 3.01E-03 |
| SCO5073 | Oxidoreductase | -4.975 | 7.48E-05 |
| SCO5074 | Dehydratase | -4.069 | 4.01E-03 |
| SCO5075 | Oxidoreductase | -2.477 | 0.039 |
| SCO5076 | hypothetical protein | -2.814 | 5.56E-06 |
| SCO5077 | hypothetical protein | -2.623 | 3.79E-07 |
| SCO5078 | hypothetical protein | -3.803 | 5.48E-03 |
| SCO5079 | hypothetical protein | -5.067 | 6.55E-06 |
| SCO5080 | hydrolase | -4.329 | 2.39E-04 |
| SCO5081 | hypothetical protein | -4.623 | 7.01E-14 |
| SCO5082 | transcriptional regulator | -2.846 | 2.15E-05 |
| SCO5083 | actinorhodin transporter | -2.666 | 4.62E-03 |
| SCO5084 | hypothetical protein | -1.939 | 0.037 |
| SCO5086 | ketoacyl reductase | -4.485 | 5.25E-04 |
| SCO5087 | actinorhodin polyketide beta-ketoacyl synthase subunit alpha | -4.556 | 6.55E-05 |
| SCO5088 | actinorhodin polyketide beta-ketoacyl synthase subunit beta | -5.883 | 1.92E-10 |
| SCO5089 | actinorhodin polyketide synthase ACP | -11.493 | 1.47E-13 |
| SCO5090 | actinorhodin polyketide synthase bifunctional cyclase/dehydratase | -7.725 | 4.72E-11 |
| SCO5091 | cyclase | -7.458 | 1.30E-14 |
| SCO5092 | actinorhodin polyketide dimerase | -6.913 | 1.96E-09 |

**Differential expression analysis of regulatory genes**

The RNA-Seq data were also analysed for differential expression of regulatory genes. Of the 644 genes identified in total, 61 (9.3%) were were regulatory genes. The regulators were split into four different groups according to their annotations: regulatory, transcriptional factor, sigma factor and two component systems (Table 6-8).

The highest changes in the regulatory genes were SCO0174, a DNA binding protein showing 188-fold increase, SCO0168 with 100-fold increase with no annotated function, a LuxR regulator with 115 fold increase and SCO4017 with a nine fold decrease in expression being annotated as an iron dependent regulator.

Table 6-8 Expression changes in regulatory genes organised by type and decrease and increase of expression on Tween compared to glucose

| Number of genes | Up-regulated | Down-regulated |
|---|---|---|
| **Two component system** | 8 | 3 |
| **Regulator** | 5 | 2 |
| **Sigma factor** | 7 | 6 |
| **Transcriptional** | 8 | 22 |
| **total** | 28 | 33 |

Table 6-9 Differential expression analysis of regulatory genes comparing expression on Tween against glucose during log phase of growth

| regulatory proteins | | fold change | p-value |
|---|---|---|---|
| SCO4232 | transcriptional factor regulator | 1.504 | 0.018 |
| SCO6474 | transcriptional regulator | 2.234 | 0.023 |
| SCO4640 | TetR family transcriptional regulator | 2.245 | 1.86E-03 |
| SCO5418 | transcriptional regulator | 3.673 | 0.014 |
| SCO5483 | transcriptional regulator | 5.007 | 3.52E-03 |
| SCO0204 | LuxR family transcriptional regulator | 115.75 | 2.41E-81 |
| SCO0174 | DNA-binding protein | 188.052 | 3.21E-63 |
| SCO3320 | redox-sensing transcriptional repressor Rex | 1.571 | 0.039 |
| SCO3694 | transcriptional regulator | -17.387 | 1.63E-12 |
| SCO4303 | transcriptional regulator | -11.159 | 2.15E-08 |
| SCO2100 | transcriptional regulator | -5.084 | 2.40E-06 |
| SCO3696 | transcriptional regulator | -3.546 | 5.86E-03 |
| SCO6813 | transcription regulator ArsR | -3.003 | 1.48E-03 |
| SCO5819 | sporulation transcription factor, WhiH | -2.938 | 7.28E-03 |
| SCO6694 | transcriptional regulator | -2.933 | 6.25E-04 |
| SCO6008 | transcriptional repressor protein | -3.529 | 5.44E-10 |
| SCO4017 | iron dependent repressor | -8.918 | 0.049 |
| SCO3207 | TetR family transcriptional regulator | -2.698 | 3.23E-06 |
| SCO5517 | transcriptional regulator | -2.336 | 0.02 |
| SCO1699 | transcriptional regulator | -2.271 | 2.08E-03 |
| SCO2105 | transcriptional regulator | -2.174 | 3.17E-03 |
| SCO2686 | LuxR family transcriptional regulator | -2.148 | 8.50E-03 |
| SCO6020 | transcriptional regulator | -2.109 | 1.68E-03 |
| SCO7554 | LacI family transcriptional regulator | -2.068 | 0.037 |
| SCO4188 | GntR family transcriptional regulator | -2.052 | 0.021 |
| SCO5100 | GntR family transcriptional regulator | -2.003 | 9.64E-03 |
| SCO6812 | transcription regulator ArsR | -1.93 | 0.041 |
| SCO0971 | LysR family transcriptional regulator | -1.876 | 0.039 |
| SCO4159 | transcriptional regulator | -1.757 | 3.13E-03 |
| SCO3571 | transcriptional regulator | -1.638 | 0.044 |
| SCO5584 | nitrogen regulatory protein P-II | -3.313 | 3.75E-07 |
| SCO5749 | two-component regulator | 2.013 | 1.63E-04 |
| SCO0203 | two-component sensor | 3.951 | 1.64E-14 |
| SCO2517 | two-component system response regulator | 4.111 | 5.02E-04 |
| SCO5784 | two-component sensor | 2.181 | 8.93E-03 |
| SCO4768 | two-component regulator | 2.329 | 2.30E-03 |
| SCO4597 | two-component system sensor kinase | 8.093 | 5.09E-19 |
| SCO4598 | two-component system sensor kinase | 5.459 | 6.61E-11 |
| SCO4261 | response regulator | 2.991 | 3.94E-09 |
| SCO4020 | two component system response regulator | -3.702 | 1.96E-09 |
| SCO4156 | two-component systen response regulator | -2.824 | 0.02 |
| SCO1654 | two-component response regulator | -1.771 | 0.022 |
| SCO4034 | RNA polymerase sigma factor | 1.682 | 0.0062 |
| SCO0194 | sigma factor | 1.768 | 0.038 |
| SCO0600 | RNA polymerase sigma factor sig8 | 1.958 | 1.19E-04 |
| SCO4425 | sigma-like protein | 3.182 | 1.68E-03 |
| SCO2954 | RNA polymerase sigma factor SigL | 4.186 | 3.75E-07 |
| SCO5216 | RNA polymerase sigma factor RpoE | -1.641 | 2.01E-03 |
| SCO4908 | RNA polymerase sigma factor | -2.044 | 2.64E-04 |
| SCO5351 | regulatory protein | 2.108 | 7.65E-06 |
| SCO1541 | regulator | 2.702 | 0.016 |
| SCO3975 | regulator | 3.178 | 4.15E-09 |
| SCO7277 | regulator protein | 4.438 | 8.86E-12 |
| SCO2865 | regulatory protein | 8.968 | 3.60E-24 |
| SCO0166 | regulator | 13.311 | 6.05E-29 |
| SCO0168 | regulator protein | 103.568 | 5.68E-79 |
| SCO4942 | regulatory protein | -27.36 | 0.027 |
| SCO3691 | regulatory protein | -8.32 | 2.15E-03 |
| SCO7252 | regulatory protein | -2.823 | 2.78E-12 |
| SCO3606 | regulator | -2.329 | 1.24E-03 |
| SCO1488 | bifun pyrimidine reg prot PyrR uracil P-ribosyltransferase | -2.048 | 0.032 |
| SCO0608 | regulatory protein | -1.874 | 1.47E-03 |

## 6.4 Temporal Differential Gene Expression comparing Log and Stationary phase for Glycolytic and Gluconeogeneic Genes on Glucose and Tween

The RNA-Seq covered the analysis of a single time point at the log phase of growth in two different carbon sources representing glycolytic (glucose) and gluconeogeneic (Tween) conditions. The results of this was verified with qPCR (6.2). Additionally a second time point of the same growth curves from stationary phase were analysed by qPCR. qPCR was performed on the same five genes as for the log phase samples: *hrdB*, *pyk1*, *pyk2*, *ppdK1* and *ppdK2*.

For this qPCR experiment another set of six samples (three biological replicates per carbon source) was analysed. They were taken from the same growth curves but in stationary phase of growth: at 36 h in glucose samples and at 48 h in the Tween growth curve.

The reference gene *hrdB* did not show significant changes either in the log or stationary phase when comparing glucose and Tween, meaning that in both growth phases the expression of *hrdB* was the same independently of the carbon source. When comparing the log phase against stationary phase on the same medium however *hrdB* showed a 6.6 fold decrease during stationary phase to log phase on Tween medium. On glucose in contrast no significant change could be detected (Table 6-10).

The gene *pyk1* showed a decrease in gene expression on Tween compared to glucose both in log and stationary phases, of 8 fold and 2.6-fold respectively. On Tween medium the gene expression was five-fold increased in stationary phase compared to log phase, which was not observed for growth on glucose (Table 6-10).

The second pyruvate kinase encoding gene *pyk2* also showed a decrease of 4.5-fold of expression on Tween compared to glucose during the log phase. However no significant change was observed during stationary phase. This suggests, that the gene is similarly expressed in glucose and Tween during stationary phase. When comparing stationary phase against log phase on glucose, no significant expression change was observed. However in Tween *pyk2* expression decreased 5.5-fold (Table 6-10).

The gene *ppdK1* had 12-fold increased expression on Tween compared to glucose during the log phase, but no significant change was detected on stationary phase. When comparing stationary and log phase for each of the different carbon sources, *ppdK1* showed a 5.1 fold

decrease in stationary phase on glucose, but no significant change could be detected on Tween (Table 6-10).

The second gene encoding a pyruvate phosphate dikinase (*ppdK2*) showed an increase in expression of 25.7-fold and 51.3-fold on Tween compared to glucose respectively in both log and stationary phase. Nontheless it showed no change in expression during the two growth phases on glucose, but on Tween a 3.6 fold increase in expression was observed during stationary phase compared to log phase (Table 6-10).

Table 6-10 Fold change in gene expression of *hrdB, pyk1, pyk2, ppdK1* and *ppdK2* on glucose, Tween and log and stationary growth phases (ns = not significant)

| **Phase** | **Log** | | **stationary** | |
|---|---|---|---|---|
| **Fold change Twe vs Glc** | | | | |
| ***hrdB*** | -2.3 | ns | 3.2 | ns |
| ***pyk1*** | -7.9 | *** | -2.6 | * |
| ***pyk2*** | -4.5 | * | 1.3 | ns |
| ***ppdk1*** | 12.3 | *** | 5.5 | ns |
| ***ppdk2*** | 25.7 | *** | 51.3 | *** |
| **Medium** | **Glc** | | **Tween** | |
| **Fold change stationary phase vs log phase** | | | | |
| ***hrdB*** | -1.2 | ns | -6.6 | * |
| ***pyk1*** | 1.6 | ns | 5.0 | ** |
| ***pyk2*** | -1.1 | ns | -5.5 | * |
| ***ppdK1*** | -5.1 | ** | -2.3 | ns |
| ***ppdK2*** | 1.8 | ns | 3.6 | * |

## 6.5 Analysis of Central Carbon Metabolites, Amino Acids and Nucleosides in *S. coelicolor* on Glucose and Tween cultures

In addition to analysing transcriptional changes during growth on the two carbon sources and a superficial proteome analysis to check for differences in phosphorylation, biomass samples from the log phase were subjected to metabolite extraction in order to analyse any differences in metabolite pools.

Comprehensive metabolomics analysis without a specific research question is a difficult task. To analyse the data obtained, based on the experimental set up, the analysis focussed on metabolites from central carbon metabolism and amino acids in addition to cofactors and metabolites involved in energy metabolism, as these were most likely to be affected under the experimental conditions comparing two different carbon sources. For growth on glucose, the flux is driven mainly through glycolysis, whereas on Tween glycolysis is bypassed, which also includes bypassing the actitvity of pyruvate kinase. The question was whether key metabolite pools differed among the two carbon sources and if this revealed any insight into the regulation of pyruvate kinase at metabolic level.

Three different measurements were carried out for each of the six samples (three from each medium): one of the intracellular metabolites, one of the culture supernatant (used medium) and one of the fresh medium, which led in total to 18 measurements. This strategy was chosen to account for metabolite leakage and secretion of metabolites into the medium.

Each sample was taken from the culture, the biomass was filtered through a glass fibre paper by vacuum filtration and was washed twice with HMM medium without carbon source. The biomass was then transferred to pre-cooled -80˚C Methanol/Chloroform/d$H_2O$ (3:1:1) for quenching and the same mixture was used for metabolite extraction. The supernatant (spent medium) was obtained from the centrifugation of the proteome samples, which was necessary in order to account for metabolites that were secreted into medium from the biomass. The fresh medium was sterile HMM medium with the respective carbon source added without any biomass in it to establish the background.

Overall 1488 metabolites were identified in at least one of the replicate samples with a confidence score of at least 5 (Creek *et al.*, 2012). Any confidence score below 5 was automatically passed to the rejected metabolite sheet within the Ideom analysis tool. These 1488 identified metabolites consisted of 802 identified from the negative ionisation mode

and 686 from the positive ionisation mode. From these identified metabolites, those that were intermediates for central carbon metabolism, amino acids, nucleosides were filtered out and analysed in more detail.

From the three biological replicates of each medium however it was noticed that both in Tween and glucose one of the replicates differed significantly, as it was always the same sample glucose B was always much higher than the other two samples and Tween A as well. These were excluded from the analysis as they were identified as outliers.

No absolute amounts were determined, only relative intensities of the different metabolites were compared.

**Glycolysis**

In the glycolytic pathway seven intermediates could be identified with a sufficient confidence score. However glucose was disregarded given that it was an ingredient of one of the media. No intracellular pyruvate was detected in any of the samples, which reflects the transient nature of this metabolite intracellularly. However it was found in the spent medium with 6-fold higher levels of pyruvate during growth on glucose compared to Tween. Similar levels of fructose-1,6-bisphosphate and fructose-2,6-bisphosphate were detected intracellularly in both conditions. Approximately 8-fold higher intracellular levels of GAP were detected in the glucose samples when compared to the Tween samples. PEP levels were 1.5-fold higher in the Tween samples, which links with the data of the RNA-Seq showing increased expression of PpdK yielding PEP as one of its products (Chapter 6). Differences were statistically significant for all metabolites, apart from fructose-1,6-bisphosphate and fructose-2,6-bisphosphate (Figure 6.16).

**TCA cycle**

There was no difference in the metabolite levels in the metabolite levels of TCA intermediates between growth on glucose and Tween. This reflects the importance of it in biosynthesis such as precursors like oxaloacetate and α-ketoglutarate for amino acids which were necessary to be synthesized as nitrate was the sole nitrogen source provided (Berg *et al.*, 2002).

Lower levels of TCA metabolites were detected for both conditions for acetyl-CoA, isocitrate, succinyl-CoA and oxaloacetate compared to levels of citrate, α-ketoglutarate, succinate and

malate in the samples (Figure 6.16). On Tween it was expected to find the glyoxylate shunt active as opposed to glucose, however as no changes were detected by RNA-Seq and it was not possible to verify this on metabolic level (Tarhan *et al.*, 2011). The RNA-Seq detected the upregulation of the EM-CoA pathway on Tween.

**Pentose Phosphate Pathway**

Examining metabolites associated with the pentose-phosphate pathway indicated that there were no significant differences among the two carbon sources. The only metabolite that showed a statistically significant change was 3-phosphoglycerate, where the concentration was about two fold higher in glucose grown cultures than those grown on Tween. Low levels of D-glucoronolactone, seduheptulose, ribulose-1,5-phosphate were detected. S7P and R5P showed a higher concentration in glucose cultures, but it was not statistically significant (Figure 6.16).

**Cofactors and Nucleoside**

FAD, NADP and NADPH were detected in the samples. Only FAD showed a difference between the two media, with about three-fold higher levels in Tween than in glucose (Figure 6.17). The presence of higher amounts of FAD might reflect the increased β-oxidation in order to utilize the palmitic acid from Tween40 (Nelson and Cox, 2004).

No differences were detected in the nucleosides. Higher levels of AMP and ADP were measured compared to the other metabolites (Figure 6.17), which is similar to findings reported by a metabolomics study where AMP levels were more than doubled to ATP and ADP more than 5-fold (Wentzel *et al.*, 2012). The energy charge was of 0.27 on both media. However AMP was on the rejected list with confidence score of 0.2 out of 10 and thus is very questionable if it was correctly identified.

**Amino acids**

Most amino acids were detected intracellularly, apart from alanine, glutamate, isoleucine, cysteine and valine. However only glutamine and proline showed a difference in their levels between the two different carbon sources, with a higher concentration on glucose compared to Tween (Figure 6.17). Both glutamate and proline are derived from AKG. Glutamate and aspartate are the preferentially utilised for growth (D'Huys *et al.*, 2011). Pyruvate and other organic acids such as succinate as well as alanine have been found to be secreted into the

medium during growth on glucose (D'Huys *et al.*, 2011). Lactate is often more associated with oxygen limitation inside pellets (Nieminen *et al.*, 2013).

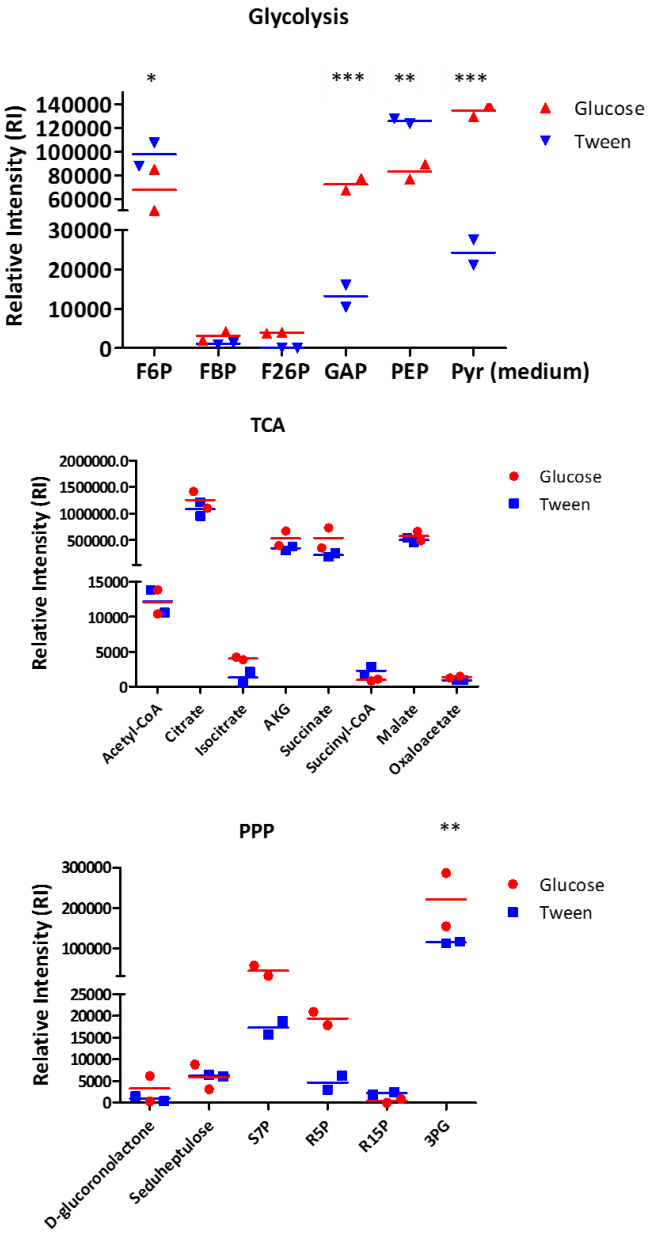



Figure 6.16 Relative Intensity of metabolite pools measured by LC-MS for glycolysis, TCA and PPP in cultures of *S. coelicolor* grown on either glucose or Tween40

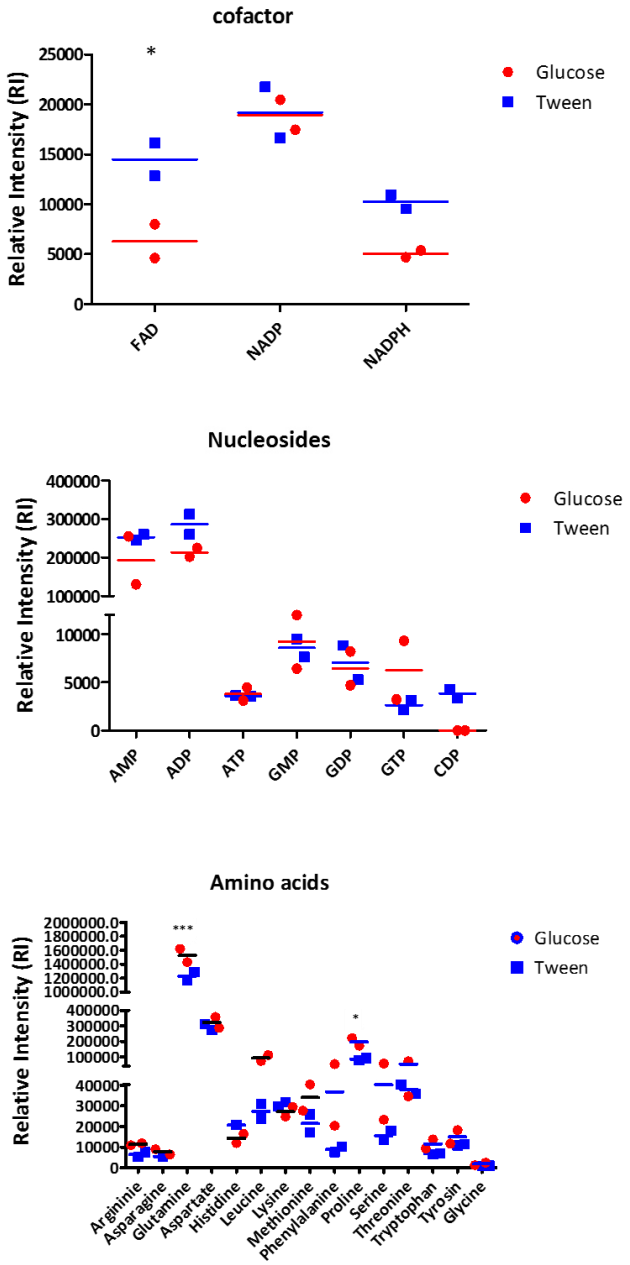



Figure 6.17 Relative Intensity of metabolite pools measured by LC-MS for cofactors, nucleosides and amino acids in cultures of *S. coelicolor* grown on either glucose or Tween40

## 6.6 SUMMARY

This Chapter was dedicated to the analysis of gene expression of central carbon metabolism on different carbon sources which were glucose and Tween with particular focus on genes that showed expansions and the analysis of central metabolites pools by LC-MS. Log and stationary phase samples were taken from growth curves in liquid medium and RNA was extracted. The log phase samples were analysed both by RNA-Seq and qPCR, whereas the stationary phase was only analysed by qPCR. The results obtained in the RNA-Seq were successfully verified by using qPCR on the same samples measuring the expression of one reference gene (*hrdB*), two genes from glycolysis (*pyk1* and *pyk2*) and two from gluconeogenesis (*ppdK1* and *ppdk2*).

In total 644 genes showed a significant change in expression on Tween compared to glucose during the log growth phase. Of those 644, 61 (9.3%) were regulatory genes and 251 (39%) were hypothetical genes. Furthermore some metal (zinc, iron and cobalt) and nitrogen metabolism genes (nitrate reductase) were expressed higher on Tween compared to glucose. Generally it was observed that gluconeogenesis, fatty acid degradation, isorenieratane biosynthesis were up-regulated on Tween, whereas glycolysis, fatty acid biosynthesis and actinorhodin and a NRPS cluster were down-regulated on Tween.

Central carbon metabolism showed the same trends for expanded genes during log growth for *pyk1*, *pyk2*, which were both down regulated and *ppdk1* and *ppdk2* were both up regulated. When considering different growth phase, there were different expression profiles observed in Tween medium: *pyk*1 was up-regulated in stationary phase compared to log phase and the opposite was true for *pyk2* which was less expressed during stationary phase. However in glucose no difference was visible. The two genes for PpdK showed that *ppdk1* expression was down-regulated during stationary phase compared to log phase and *ppdk2* expression was increased on both glucose and Tween.

Different trends were observed among genes annotated for the same function encoding GAPDH, where two down-regulated and one up-regulated on Tween. Phosphoglucose isomerase (Pgi) showed one down-regulated and the other unchanged on Tween. Triosephophaste isomerase (Tpi) had one gene up-regulated and the other one down-regulated on Tween. PDHC showed one up-regulated, one down-regulated gene and the rest unchanged whereas SUCDH had two genes higher expressed on Tween and the rest

unchanged. The malic enzymes had one gene down-regulated and the other unchanged in its expression when comparing Tween against glucose.

Since 9.3% of the differential expressed genes were regulatory genes, those were analysed superficially and there remains to be further explored. Some of these regulators were close to cluster being differentially expressed and might be a first indicator to their role such as SCO0203 in close proximity to *ppdk1* (SC00208).

In the analysis for differences in metabolite pools in central carbon metabolism, cofactors, nucleosides or amino acids only a few differences were observed. Pools of F6P and PEP were higher on Tween. Intracellular GAP, proline and glutamine and pyruvate excreted in the medium were higher during growth on glucose. Proline and glutamine are both derived from glutamate and ultimately from AKG. Pyruvate secretion is associated with an imbalance between carbon and nitrogen metabolism and is a common observation on glucose medium for *Streptomyces* (D'Huys *et al.*, 2011). F6P and PEP both serve as activators for ADP-Glc PPase, which is involved in glycogen catabolism (Asención Diez *et al.*, 2012). Furthermore PEP is the product of PpdK, an enzyme whose transcription was highly upregulated on Tweenwith more than twenty fold for Ppdk2.

# 7 DISCUSSION

## 7.1 EXPANSION OF CENTRAL CARBON AND AMINO ACID METABOLISM IN ACTINOBACTERIA

The analysis presented in this thesis was focussed on identifying targets for increasing polyketide production in *Streptomyces* by focussing on the interaction of primary metabolism and specialised metabolism. The primary metabolism of *Streptomyces* shows a large number of gene expansions and these were analysed in more detail. Gene expansion leads to the diversification of metabolism, a new function can be acquired either by gene duplication or horizontal gene transfer followed by divergence in both cases (Bentley and Parkhill, 2004; Konstantinidis and Tiedje, 2004; Treangen and Rocha, 2011; Schulenburg and Miller, 2014; Copley, 2014; Copley, 2015).

The most expanded pathway studied in *Streptomyces* was glycolysis. A general observation from the expansion of central carbon metabolism and amino acid metabolism in Actinobacteria was that those genera known for specialised metabolite production also showed a higher number of gene expansions in primary metabolism. Interestingly, from the 13 suborders analysed, Catenulisporineae and Pseudonocardineae showed the highest proportion of expansions overall. When analysing pathway specific expansion, it was observed that Streptomycineae and Catenulisporineae, had the highest number of gene expansions for glycolysis. Nevertheless, it has to be kept in mind that, for Catenulisporineae, only three genomes from two genera were included in the analysis and might not be seen as a sufficiently representative group. Streptomycineae included 303 species from three genera and in Pseudonocardineae 57 different species from 16 genera were included in the database.

Different patterns could be observed for each taxonomic group, indicating that there is no general pattern across all taxonomic groups. Whether or not each of these expansions play a role in increasing precursor supply for specialised metabolites remains to be determined or whether they simply reflect niche specific genes. It would be interesting to analyse, if there were a pattern in which pathways are expanded, if it can be linked to the chemical nature of the metabolites the group of bacteria is able to produce. Unfortunately many of the

biosynthetic clusters are silent under standard laboratory conditions, which makes addressing this question challenging. This analysis is a start and there are other efforts for mining the available genomes for unknown or silent producers currently by different bioinformatics approaches (Cimermancic *et al.*, 2014; Challis, 2014; Cruz-Morales *et al.*, 2015; Medema *et al.*, 2015). The acquisition of sequencing data has indicated from a number of studies that the actinobacteria and probably also other taxonomic groups harbour the genetic potential to make a variety of compounds that have not been described at this time (Watve *et al.*, 2001; Medema and Fischbach, 2015).

The discovery of new bioactive compounds is of great interest at the moment due to the emergence of antibiotic resistant strains causing disease and which has been highlighted by the WHO since 2001 (WHO, 2001). It is not only about discovering the compounds for drug applications, but also understanding the biology of specialised metabolite producers first, in order to systematically exploit the potential that they harbour. In order to do so, different levels of knowledge are needed to be combined - the bioinformatics tools allowing big data sets to be analysed, such as the study of metabolic expansions together with experimental data from phenotypes to know which metabolites are made under which conditions and discover if a general pattern emerges or even patterns for specific taxonomic groups. At this point, this is only a hypothesis. This might not be the case and the situation might be more complex as there are other factors, such as regulators, nutrient availability as well as interactions with other bacteria, which play important roles in the production of specialised metabolites too (Bibb, 1996; Bibb, 2005; den Hengst *et al.*, 2010; Traxler *et al.*, 2013).

A key point relating to these data is the specific role of specialised metabolites is still not fully understood. There are several theories relating to their production and evolutionary significance such as chemical weapons to gain advantage in the habitat (Waksman, 1941; Ratcliff and Denison, 2011; Raaijmakers and Mazzola, 2012) or as signalling molecules between the same or different species of microorganisms (Davies *et al.*, 2006; Linares *et al.*, 2006; Yim *et al.*, 2007; Yim *et al.*, 2010; Romero *et al.*, 2011). The classic defense theory has been challenged by the fact that antibiotics are almost undetectable in soil and if then at subinhibitory concentrations. In addition low concentration of a drug can increase biofilm formation (Hoffman *et al.*, 2005) or trigger the expression of virulence genes (Linares *et al.*, 2006; Yim *et al.*, 2010). Thus the theory of the metabolites mediating interactions between microorganisms arose from there. It is likely that given the wide range of ecological niches

occupied by bacteria, either of the hypotheses might be true depending on the metabolite studied. A difference in this 'social' role for specialised metabolism might be whether only the sender benefits and harms the receiver or both are benefiting from the signal. (Doroghazi and Metcalf, 2013; Davies, 2013; Gubbens *et al.*, 2014; Doroghazi *et al.*, 2014; Abrudan *et al.*, 2015; Eick and Pohnert, 2015; Kelsic *et al.*, 2015; Traxler and Kolter, 2015).

Another pertinent question that remains is how specialised metabolite pathways evolved. Evolutionary selective pressure would tend to favour the loss of such large, energetically expensive biosynthetic gene clusters if they did not serve to increase the fitness of the producing organism. Generally it would be expected that selection pressure would favour traits that enhance chemical diversity and subsequently retain it, as well as favour traits that will reduce the fitness cost (Firn and Jones, 2000) in order to increase chances to produce a metabolite with bioactivity. However the presence of low concentration is stabilising the biodiversity in the habitat and thus it might not only be the fitness of one particular organism but the community as a whole (Abrudan *et al.*, 2015).

An overlooked aspect for the exploitation of a microorganism for the production of a bioactive compound is the fact, that the actual ecosystem is significantly more complex than an artificially environment replicated in the lab. There are many species that are un-culturable in the laboratory, and the ones that can be cultivated express only a fraction of the biosynthetic gene clusters encoded in their genome. This might be because insufficient medium condition or because they live in a symbiosis and depend very much on the presence of another organism. There are studies investigating organisms in co-culture in order to access more of the chemical diversity discovered by metagenomics approaches and also to find methods to trigger expression of silent biosynthetic clusters (Traxler *et al.*, 2013; Abrudan *et al.*, 2015). Not only do correct conditions for metabolite cluster expression need to be identified but the demand for precursors produced during primary metabolism must also be met, thus understanding primary metabolism and its expansions is of importance to ensure precursor supply.

## 7.2 Role of Duplicated Pyruvate Kinase in *S. coelicolor*

Pyruvate kinase was one of the fourteen enzyme expansion events identified in the genus of *Streptomyces* through bioinformatics analysis. Examining the phylogenetic background of pyruvate kinase expansion was found to have arisen through duplication in Streptomycineae

comprising the genera *Streptomyces, Kitasatospora* and *Streptacidophilus* and is highly conserved. It is ikely to have occurred prior the radiation of *Streptomyces* speciation as 281 out 286 *Streptomyces* species show the presence of two pyruvate kinase and exactly one in each of the two branches (Chapter 3). There were three exceptions to this: *S. somaliensis, S. sp NRRL F5135* and *S. scrabrisporus* possessed a single pyruvate kinase, two of these grouping in the Pyk2 branch (*S. somaliensis* and *S. sp NRRL F1535*) indicating that Pyk2 might be the primary pyruvate kinase in *Streptomyces.*

The Gram- negative bacterium *E. coli* also possesses two pyruvate kinases which are allosterically regulated by AMP (PykA) and FBP (PykF) (Muirhead, 1990; Mattevi *et al.*, 1995; Knowles *et al.*, 2001), while two enzymes only share 35 % amino acid sequence identity and did not appear through duplication. Examining the literature, AMP seems to be the most prominent regulator in bacteria and PykF in *E. coli* seems to be rather the exception than the rule regarding its regulation by FBP in bacteria. FBP is the more prominent allosteric regulator in eukaryotic pyruvate kinases (Morgan *et al.*, 2014). In *E. coli* the presence of two pyruvate kinases does not seem to be the result of duplication as the sequences are very divergent. One might have been acquired through horizontal gene transfer. However this is an open question and remains to be answered as insufficient data are available to explain the presence of the two genes in *E. coli*.

In contrast to duplication, horizontal gene transfer is reported to be more frequent than duplications overall when discounting mobile elements and viral DNA (Treangen and Rocha, 2011). The two mechanisms play different roles in evolution. Horizontal gene transfer acquires new functions whereas gene duplication leads initially to higher gene dosage (Treangen and Rocha, 2011). Additionally duplications show slower evolution compared to horizontally acquired genes (Treangen and Rocha, 2011).

The retention of multiple gene copies in the genomes of the Streptomycineae suggests that selective advantage to fitness is conferred following duplication. Gene duplication usually does not have severe consequences for the fitness of the organism and can promote evolutionary changes over large time scales (Wagner, 2008). Generally it seems that the level of pathway connectivity of an enzyme is inversely correlated to the rate of evolution, meaning that a highly connected enzyme will evolve slower due to functional constraints on its diversification. Additionally it was found that enzymes in pathways with high level flux are faced with stricter evolutionary constraints and are more likely to retain duplicated copies of

genes (Vitkup *et al.*, 2006). For the presences of *pyk1* and *pyk2* in various *Streptomyces* species purifying selection was detected, and its characteristics agree with the literature as a highly connected, slowly evolving enzyme.

The pyruvate kinase phylogenetic tree (Figure 3.2) indicates that some species might have acquired an additional pyruvate kinase by horizontal gene transfer rather than duplication. It is important to note, however that a discrepancy was observed in the expansion analysis concerning how many pyruvate kinases were detected in the RAST functional role count analysis and those detected by blastp. This was likely due to the BLASTp algorithm solely searching for sequence similarities whereas RAST also compares sequences according to their functionality using the FIGfam database representing a manually curated database grouping proteins that share the same functional role, adding a second layer to the complexity search (Aziz *et al.*, 2008). A notable example of this is *C. glutamicum* that is reported throughout the literature to possess a single pyruvate kinase, but a second copy may have been missed during genome analysis (Gubler *et al.*, 1994; Becker *et al.*, 2008; Wieschalka *et al.*, 2012). Experimental work would be required to confirm biochemically that both copies encode active pyruvate kinase, but this was beyond the scope of this work. A quick and simple experiment would be to grow the *pyk* mutant in minimal medium with glucose as sole carbon source to determine if it is still able to grow. If it can grow, it might be because a second pyruvate kinase is present.

It was possible to deduce that both Pyk1 and Pyk2 of *S. coelicolor* are functional pyruvate kinase, as both were able to complement a *pyk* double mutant of *E. coli*, which was unable to grow on glucose as sole carbon source.

Mutants of *pyk1* and *pyk2* show phenotypic differences (Chapter 4). The *pyk1* mutant is more affected in specialised metabolite production as it overproduces both coelimycin and actinorhodin (Figure 4.16-Figure 4.17) whereas the *pyk2* mutant is negatively affected in growth on solid medium (Figure 4.18). Generally, detecting a phenotype for a central carbon metabolic enzyme is difficult as this part of metabolism is extremely robust, therefore changes might be subtle or even undetectable. It is therefore crucial to analyse the mutants during growth on a range of different carbon and nitrogen sources. The differences observed suggest that each of the pyruvate kinases in *Streptomyces* have distinct physiological roles. It was also observed that the germination yield was affected by the deletion of either of the pyruvate kinases pointing towards a role in spores, perhaps through storage compound

production during sporulation or during the germination and out growth stage of the lifecycle.

A gene dosage effect on the specific growth rate was observed for strains with different number of copies of pyruvate kinase mutants provided. The relationship was inverse, as the mutants showed a higher growth rate compared to the WT and the additional copy led to a decreased specific growth rate on glucose. The WT was somewhere between the two. This suggests that the flux through glycolysis and ultimately the specific growth rate is changed in the presence of several pyruvate kinases and the data presented here suggest different regulation of the two by AMP. The presence of the two pyruvate kinase activities might help distributing the flux towards growth or specialised metabolite production. A possible experiment to test this further might be to construct strains containing several copies of only *pyk1* or *pyk2* and also analyse the production of polyketides.

Both *pyk* genes are expressed throughout the life cycle and biochemically it was found that both are positively affected by AMP. The kinetics observed for both Pyk1 and Pyk2 were in agreement with that found for pyruvate kinases in other bacteria. Sigmoidal kinetics can be observed for the substrate PEP, as it is the case in *C. glutamicum* (Jetten *et al.*, 1994), *S. aureus* (Zoraghi *et al.*, 2010) and *Syneococcus PC6301* (Knowles *et al.*, 2001) as well as the allosteric regulation by AMP and R5P as observed with Pyk1 and Pyk2. The absolute values for the kinetics are difficult to compare directly as the assay conditions were different across the different publications. However they matched roughly: Pyk1 matched with a $K_{cat}$ of 4200 $s^{-1}$ to the $K_{cat}$ reported for *C. glutamicum* of 2540 $s^{-1}$ under similar conditions in the presence of AMP (Jetten *et al.*, 1994). Pyk2 was similar to the $K_{cat}$ for PEP in the absence and presence of AMP with 18 $s^{-1}$ and 336 $s^{-1}$ respectively to 98 $s^{-1}$ and 181 $s^{-1}$ in *S. aureus* (Zoraghi *et al.*, 2010).

Generally it is accepted that the fate of duplicated genes is that they diverge in either expression or function following duplication (Huminiecki and Wolfe, 2004). Pyk1 is highly activated in the presence of AMP, so at low energy states such as stationary phase, AMP levels rise and Pyk1 will be the major pyruvate kinase as opposed to Pyk2 which is active at low PEP and ADP concentrations. This analysis detected a physiological divergence between the two copies.

The current hypothesis therefore is that Pyk2 is the main pyruvate kinase with its low $S_{0.5}$ for both substrates and Pyk1 is the accessory pyruvate kinase augmenting the role of Pyk when AMP levels start to rise, which is likely during the entry into stationary phase when overall energy charge drops. This theory could be further supported by the results of the qPCR experiment showing that on medium with Tween as carbon source Pyk1 expression was increased in stationary phase opposed to Pyk2 showing decreased expression. This was not observed on glucose. However this fits very well with the natural habitat of *Streptomyces,* where glucose is a rare resource to find.

Furthermore the sizes of the proteins and CD measurements also show differences for the two enzymes. Pyk1 has a size of 206 kDa, which approximates to a tetramer as one monomer has a size of 51.6 kDa. In contrast Pyk2 was 307 kDa as determined by gel filtration, which suggests a hexameric structure when extrapolated from the 51.2 kDa monomer. The most common structure for pyruvate kinase is a homotetramer, with a single report of a leucoplast pyruvate kinase from developing castor bean endosperm showing a heterohexameric structure (Plaxton *et al.*, 1990). The CD analysis of the two proteins also suggests differences between the two enzymes- Pyk2 appears to have a mix of α helices and β sheets, whereas Pyk1 has predominantly β sheets (5.7).

## 7.3 Role of Pyruvate Phosphate Dikinase in *S. coelicolor*

The literature reports that pyruvate phosphate dikinase can also perform the pyruvate kinase reaction when no functional pyruvate kinase is available. Given that this is the case, pyruvate phosphate dikinase was included in the analysis, however at this time no mutant for this gene has been made and no information on its phenotype is available. It was identified as a gene expansion candidate for the genus of *Streptomyces*, although its phylogeny suggested that the two enzymes are divergent and have not arisen by duplication despite the almost 65% sequence identity on amino acid level between the two genes in *S. coelicolor*. No annotation for pyruvate phosphate dikinase was detected for 28 out of 83 genera included in the database and ten genera showed expansion which also all were spore forming genera: *Kribbella* (Kaewkla and Franco, 2013), *Micromonospora* (Hoskisson *et al.*, 2000), *Dactylosporangium* (Sharples and Williams, 1974; Vesselinova and Ensign, 1996), *Actinoplanes* (Parenti and Coronelli, 1979), *Lechevalieria* (Labeda *et al.*, 2001), *Actinomycetospora* (Jiang *et al.*, 2008), *Streptosporangium* (Lechevalier *et al.*, 1966), *Herbidospora* (Kudo *et al.*, 1993), *Actinospica* (Cavaletti *et al.*, 2006) and *Streptomyces* (Flärdh

and Buttner, 2009). The enzyme has been reported to be important during germination of plant seeds (Chastain *et al.*, 2006), it would be interesting to investigate if it also plays a role in spore germination of *Streptomyces* and the other genera showing expansion for this functional role.

Analysis of the gene expression of the two has been analysed by RNA-Seq and qPCR and revealed that both are highly expressed under gluconeogeneic conditions. However the profile was different when comparing log growth against stationary growth where only *ppdk1* expression was decreased on glucose and expression of *ppdK2* was increased on Tween. This suggests different roles for these two paralogues. However this needs to be confirmed with phenotypic studies of mutants.

## 7.4 Role of Pyruvate Dehydrogenase Complex in *S. coelicolor*

The mutant in SCO7124 that encodes $PDHC_{E1}$ resulted in the strongest overproduction of actinorhodin during the screening of interesting targets. $PDHC_{E1}$ is described as the rate limiting step from the complex formed by E1, E2 and E3. This enzyme was also highlighted as an expansion event in the genus of *Streptomyces* and is part of the node of interest.

Interestingly, *Streptomyces* has both gene conformations for $PDHC_{E1}$, which are typical for Gram- negative bacteria with a single gene coding for $PDHC_{E1}$ and at the same time two genes for $PDHC_{E1\alpha/\beta}$ which is the usual structure for Gram positive bacteria. *C. glutamicum* is described to have only a Gram-negative like $PDHC_{E1}$ (*aceE*) and the mutant was unable to grow on minimal medium with glucose, pyruvate or lactate as sole carbon sources (Schreiner *et al.*, 2005).

Phylogenetically, only the genes encoding $PDHC_{E1}$ were analysed as the alignments for $PDHC_{E\alpha/\beta}$ were too diverse to construct a meaningful phylogenetic tree. The alignment of the $PDHC_{E1}$ suggested divergent genes. Two of the genes (SCO7124 and SCO2183) were much closer related than to SCO2371. Generally this enzyme complex is challenging to study as it is closely related to other complexes of the 2-oxoacid dehydrogenase family leading to false positives on the phylogenetic tree as they will share significant similarity in parts. The family of 2-oxoacid dehydrogenases contains apart from pyruvate dehydrogenase also 2-oxoglutarate dehydrogenase, branched chain dehydrogenase, glycine and acetoin dehydrogenase complexes. However specificity for their substrate is very narrow (de Kok *et al.*, 1998).

The complementation of the mutant strain was unsuccessful. When analysing the immediate vicinity of the gene, it was noticed that upstream of SCO7124 sits an acyltransferase (SCO7123, $PDHC_{E2}$), a secreted protein, cytochrome B and C and a dehydrogenase and downstream mainly hypothetical or unknowns, except for a transmembrane efflux protein (probable ACT transporter, SCO7128). Due to the increased production of actinorhodin, maybe the knockout of this gene has polar effects on this efflux protein and maybe the cytochrome which are involved in ATP production. As hypothesised for pyruvate kinase perhaps the role of energy charge may impact on PDHC function.

As overall 14 different genes are present encoding the several components of this complex, yielding in theory three different $PDHC_{E1}$ , three different $PDHC_{E1\alpha/\beta}$, three different $PDHC_{E2}$ and three $PDHC_{E3}$ it would be important to know which of these form one complex in order to gain further insight. Some information could be obtained from transcriptomic and proteomic data to which genes were co-expressed or which proteins were present at the same time.

In the RNA-Seq experiment only two of the genes showed differential expression, both were $PDHC_{E1}$ which showed a decrease in expression on Tween and $PDHC_{E1\alpha}$ an increase in expression on Tween compared to glucose (Chapter 6). The remaining genes did not show any changes in expression under the tested conditions.

In a proteomics study in *S. lividans* homologous genes to *S. coelicolor* SCO2180 ($PDHC_{E2}$), SCO0884 and SCO4919 ($PDHC_{E3}$) were down-regulated in the *ppk* mutant (deficient in polyphosphate kinase) in P-limited R2YE medium (Le Maréchal *et al.*, 2013). This might indicate that those gene products interact. And in another proteomic study in *S. coelicolor* conducted by Gubbens *et al* (2012) SCO2183 (E1), SCO3815 (E2), SCO4919 (E3) and SCO2180 (E2) showed a slight increase in expression on mannitol. A decrease was observed for SCO3816-17 (E1α/β) and 0884 (E3) on mannitol in the WT. Since the genes SCO2180 abd SCO4919 showed similar behaviour in two independent studies it might indicate that they are part of the same complex

However it is difficult to draw general conclusions as some of the genes or enzymes are not always detected and thus interpreting the results is challenging, especially because limited phenotypic data from mutants is available. It also remains the question to how and why *Streptomyces* shows both types of E1.

## 7.5 Transcriptional Changes of Central Carbon Metabolism and Regulators on Glucose and Tween

The RNA-Seq data lacked sequencing depth, which was due to the small chip size (316v2) and by pooling three samples per chip as *Streptomyces* has a large genome. Nevertheless it was still possible to extract useful data as 30 x genome coverage was achieved on average across the genome. Additionally as the interest was in primary metabolic enzymes which are expressed at high levels and consistently, the data set proved useful (Alam *et al.*, 2010; Thomas *et al.*, 2012; Gubbens *et al.*, 2012).

The internal control with ERCC transcripts proved to be unusable as the analysis Plug In from the Ion Torrent Suite gave inconsistent results. Thus they were disregarded for the analysis and it is not advised to use this product with prokaryotic samples by Life Technologies upon trouble shooting with the technical support team.

The use of qPCR of five different genes was used to validate the dataset and aided the decision to use CLC Genomics Workbench for the data analysis as the data correlated best when analysing the expression of both pyruvate kinases, pyruvate phosphate dikinases and *hrdB* between the two methods.

A high proportion of the differentially expressed genes were hypotheticals (39%) and regulatory genes (9.3%). Furthermore some metal and nitrogen metabolism genes were expressed differently on Tween compared to glucose. Overall gluconeogenesis, fatty acid degradation, isorenieratane biosynthesis were up-regulated, whereas glycolysis, fatty acid biosynthesis and actinorhodin were down-regulated on Tween as would be expected, validating the approach to isolate these metabolic pathways using catabolic substrates.

Changes in glycolysis, fatty acid degradation and gluconeogenesis were as expected with the high proportion of regulators in *Streptomyces*- RNA Seq is a good first step to identify their effect on gene expression. There were three regulatory genes (SCO0168, SCO0174 and SCO0204) that were more than a 100 fold times higher expressed in Tween than in glucose. This might be a starting point, in particular SCO0204 that is in the immediate vicinity of *ppdk1* (SCO0208). There were some genes involved in nitrite/nitrate metabolism SCO0213 (nitrite/nitrate transporter), SCO0216 (nitrate reductase α) and SCO0217 (nitrate reductase β) which were also expressed more highly on Tween than on glucose.

The isorenieratane pathway (SCO0185-0191) is usually induced by blue light by SCO0193 (*litR*), but this regulator was not differentially expressed in Tween containing media. Carotenoids act as accessory pigments for light capturing complexes in phototrophic bacteria and they protect cells from photo-oxidation (Takano *et al.*, 2005). However the data here suggest that there is another signal that can induce expression of these genes, as both culture were exposed to the same light and the only difference was the carbon source. Maybe the cultivation on Tween caused an oxidative stress similar to that from photodamage and thus induced the expression of protective molecule.

A biosynthetic cluster (SCO6429-6438) encoding an unknown NRPS (Medema *et al.*, 2011; van Keulen and Dyson, 2014) is expressed less in Tween than in glucose. However, it is unclear if the product is present in the culture. It is difficult to identify metabolite presence without knowing its structure and further investigation is needed.

This experiment also identified different expression patterns for central carbon metabolic enzymes which have several genes annotated to encode that function. Further, different expression was observed for a number of 'redundant' (contingent) central metabolic enzymes. These observations support the hypothesis of different roles of the different isoenzymes. For example, the genes encoding GAPDH showed that expression of SCO1947/7511 was down-regulated on Tween, but SCO7040 expression was up-regulated on Tween. In the case of *pfk* SCO1214 expression was down-regulated and no change for the other two copies on Tween compared to glucose. Similar observation was seen for *zwf* with decreased expression of SCO6661 but unchanged for SCO1937 on Tween. The genes *pgi* showed that SCO6659 expression was down-regulated and the other gene unchanged similar to the genes for transketolase, where SCO6768 expression was increased and the other genes were unchanged on Tween. The PDHC gave increased expression for SCO2183 and decreased expression for SCO1269. The expression of genes for fumarase were decreased in the case of SCO5042 and unchanged for SCO5044. SCO0922 and SCO0924 encoding SUCDH were both down in their expression on Tween and the remaining genes were unchanged in their expression. The malic enzymes displayed differences too, SCO2951 the NAD dependent malic enzyme showed decreased expression as opposed to unchanged expression for SCO5621 which is NADP dependent.

## 7.6 Proteomic Analysis of Central Carbon Metabolism in Different Carbon Sources

As a proteomic analysis is lacking in this thesis, the literature was analysed to find a study which focussed on the central carbon metabolism.

The proteomes of *S. coelicolor* WT M145 and Δ*glkA* were analysed during growth on different carbon sources by Gubbens and co-workers (Gubbens *et al.*, 2012). The objective of the publication was to identify the glucose kinase dependent and independent carbon catabolite control in *S. coelicolor* using proteomics and some of the data were used to compare to the results obtained from RNA-Seq. Similar trends were observed for gene expression and protein expression regarding glycolytic and gluconeogenetic conditions for the pyruvate kinase, Ppdk2 and GAPDH. Only the medium comparison data sets were used when comparing the protein expression on glucose and mannitol against mannitol and glucose plus fructose mix against fructose for both the WT and Δ*glkA*. Mannitol is a repressing carbon source. The study that identified most central carbon metabolic enzymes.

Gubbens et al., 2012 conducted their experiments using *S. coelicolor* comparing WT or Δ*glkA* in NMMP supplemented with 0.5 % of casamino acids and either 1 % mannitol or fructose alone or additionally supplemented with 0.5 % glucose. Samples were taken at mid-log phase (18-20 h after inoculation) and proteins were extracted and analysed. The ratios of the protein expression when comparing two different conditions were expressed as 2-log ratios ranging from -5 to +5.

The WT showed an increase in protein expression for both pyruvate kinases when glucose was present in the medium. This is what was observed in the RNA-Seq experiment, where both genes were down-regulated on Tween. In the Δ*glkA* strain no increase in expression of Pyk1 (SCO2014) in the presence of glucose was observed on either fructose or mannitol. This suggests that Pyk1 (SCO2014) expression is dependent on GlkA, but not Pyk2 (SCO5423, Figure 7.1-Figure 7.2). This observation further supports the hypothesis of physiological divergent roles of the two enzymes (7.2).

The expression of PpdK2 (SCO2494) was highly repressed by glucose on both mannitol and fructose. This was observed for both strains. Ppdk1 (SCO0208) was detected 3.4 fold increase in the mutant on mannitol, but could not be detected in any of the other conditions (Figure 7.1-Figure 7.2). This observation was also in accordance with the results of the transcriptomic

analysis, which showed an increase in expression of both pyruvate phosphate dikinase on Tween.

PEPCk expression was repressed by glucose, however this repression was abolished in the Δ*glkA* strain both on mannitol and fructose, which was similar to what was seen in the RNA-Seq experiment (Table 6-6). PEPCx expression was increased in presence of glucose; however the effect was more pronounced on fructose than on mannitol. The increase in expression was less in the Δ*glkA* cultures. (Figure 7.1-Figure 7.2). No change in gene expression was observed between glucose and Tween cultures in the RNA-Seq experiment.

Pyruvate carboxylase could not be analysed as it was not detected in any of the tested conditions. It was repressed in presence of glucose on fructose, however it was not detected on mannitol. The mutant showed a minimal decrease on mannitol and a minimal increase on fructose in presence of glucose. An increase in expression of this gene on Tween was observed in the RNA-Seq (Figure 7.1-Figure 7.2).

The malic enzymes SCO2951 (NADH-dependent) showed a 2-log ratio change of -0.4 on mannitol and of 0.8 on fructose for the WT in comparison to the presence of glucose. The RNA-Seq detected a decrease in expression on Tween when compared to glucose.However, on both media, the mutant showed a slight decrease in expression (Figure 7.1-Figure 7.2).

The WT showed no change for SCO5261 (NADPH-dependent) on mannitol and a 1.5 2-log ratio decrease on fructose. No change in expression was detected in RNA-Seq for this gene. For the mutant this gene showed an increase on mannitol, but a decrease on fructose (Figure 7.1-Figure 7.2).

Another observation from the study was that GlcP levels increased in the presence of glucose which fits with the observation in the RNA-Seq experiments of an almost 26-fold decrease of transcription of SCO5578 on Tween compared to glucose. SCO7153 showed no change in expression.

The three GAPDH genes also showed a similar pattern as observed in the RNA-Seq experiments, SCO1947 and SCO7511 was induced by glucose and SCO7040 was repressed by glucose fitting the transcription profile.

The data from the RNA-Seq largely agreed with the proteomic data from the literature, although different media were employed, further supporting the data obtained. An

exception to this was SCO3127 (PEPCx) and SCO5261 (NADPH-me). However the changes observed in the proteome study might have been specific to the carbon source tested. Additionally the experiment did not analyse sole carbon source, but a mix of glucose with either mannitol or fructose against mannitol or fructose alone.

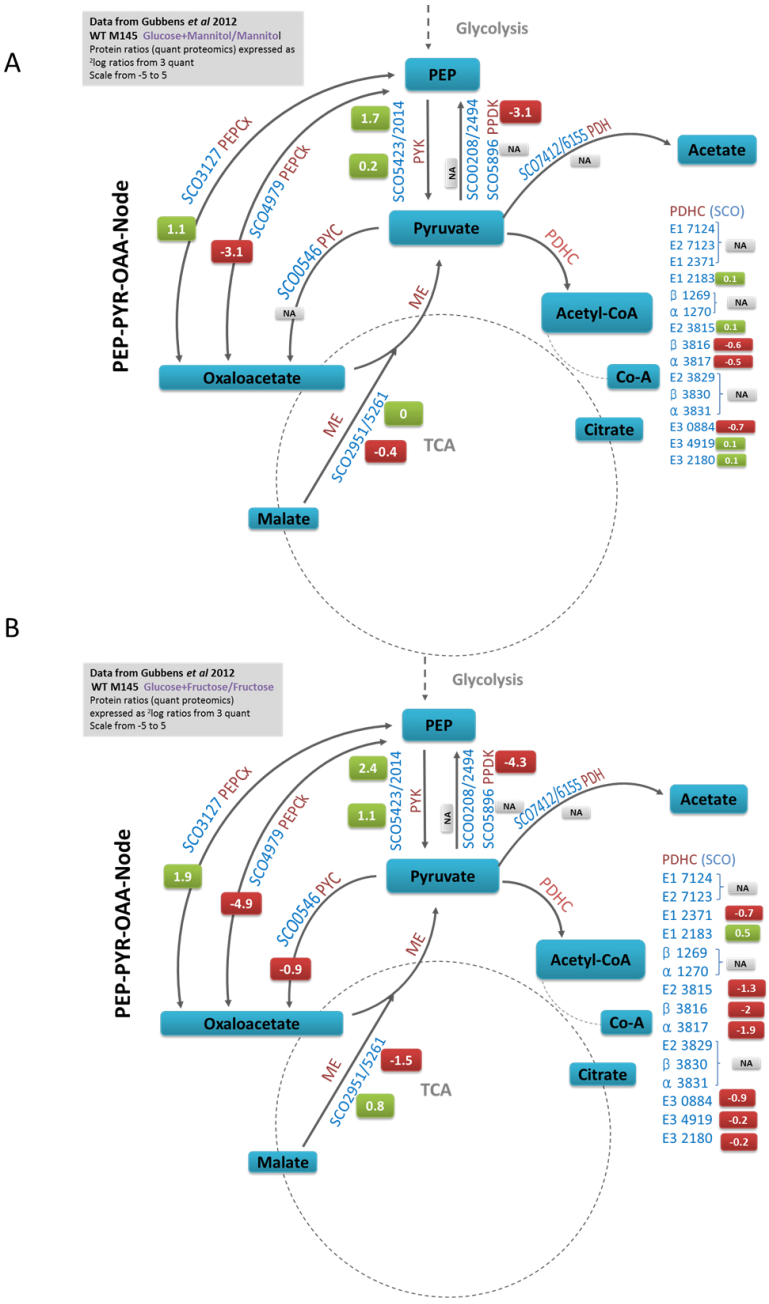



Figure 7.1 Proteomic changes in 2-log ratio of WT M145 grown on mannitol or mannitol plus glucose (Top) or on fructose or fructose plus glucose (bottom), data obtained from Gubbens *et al*., 2012

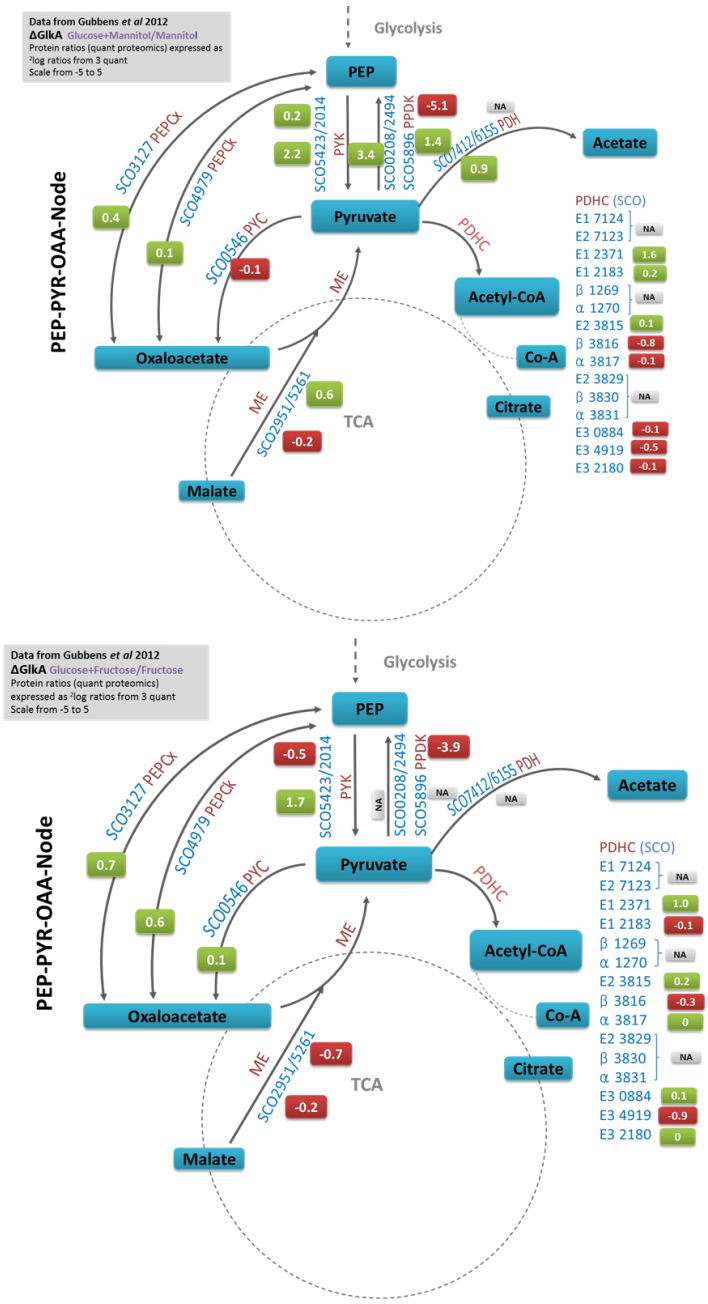



Figure 7.2 Proteomic changes in 2-log ratio of Δ*glkA* grown on mannitol or mannitol plus glucose (Top) or on fructose or fructose plus glucose (bottom), data obtained from Gubbens *et al*., 2012

## 7.7 Changes in Central Carbon Metabolite Pools during Growth on Glucose or Tween

When sampling a culture extraction of metabolomics it is important to sample rapidly and stop the metabolism by quenching as quickly as possible (Bolten *et al.*, 2007; Winder *et al.*, 2008). Here the sampling was carried out manually and the biomass was first filtered, so they were still metabolic active as the quenching of the metabolism was only done after filtration into cold methanol/ethanol. However as all samples were treated equally the error was judged to be the same for all samples collected. Sample collection took approximately 2 minutes.

Generally the metabolite analysis revealed only few differences among glucose and Tween. Most differences were observed for glycolytic intermediates. Unsurprisingly, the intermediates of the TCA cycle did not show any difference as it is essential under all conditions.

The concentration of F6P and PEP were increased in the Tween cultures. Both are reported as activators for the enzyme ADP-Glc PPase which is involved in glycogen catabolism (Asención Diez *et al.*, 2012). In the RNA-Seq experiments the glycogen branching enzymes GlgBI and GlgBII (Yeo and Chater, 2005) were not differentially expressed between the two media. No glycogen content was determined here, so it is difficult to prove this hypothesis. PEP is also the product of the reaction of PpdK, which expression is highly increased on Tween. Thus it could simply accumulate due to the increased gluconeogenesis.

Glyceraldehyde-3-phophate is an intermediate of glycolysis, gluconeogenesis and the pentose-phosphate-pathway and was present in higher amounts in the glucose cultures than in Tween. Extracellular pyruvate was found in high concentrations in the glucose cultures compared to Tween, under these conditions *S. coelicolor* is known to secrete organic acids into the medium. The most abundant of these are pyruvate and α-ketoglutarate (Ahmed *et al.*, 1984; Dekleva and Strohl, 1987; Hobbs *et al.*, 1992; Madden *et al.*, 1996; Karandikar *et al.*, 1997; D'Huys *et al.*, 2011). The volume of organic acid secreted depends on the nitrogen source. Nitrate leads to high pyruvate and low α-ketoglutarate, amino acids to low pyruvate and high α-ketoglutarate, whereas ammonium salts prevent organic acid formation (Madden *et al.*, 1996). In the cultivation medium nitrate was the nitrogen source utilised and the data obtained agree with the literature. Possible explanations for the secretions are that glucose

represses certain dehydrogenases (Ahmed *et al.*, 1984) or it represents an imbalance of carbon and nitrogen metabolism and biosynthetic fluxes (D'Huys *et al.*, 2011).

In the amino acids only proline and glutamine cellular pools showed a difference; both were increased in glucose cultures. Given that both these amino acids use AKG as a carbon skeleton, these may represent a convenient metabolic sink for AKG, or a futile cycle.

In this case, the metabolomics analysis did not help obtain further information, but it confirmed observations made in the other experiments. More information could be obtained by studying the enzymes themselves.

# 8 CONCLUSION AND FUTURE WORK

In conclusion it is possible to say that engineering the PEP-PYR-OAA node harbours potential to increase specialised metabolite production in several ways. By establishing a proof of concept in *S. coelicolor* using actinorhodin as model polyketide- it was possible to identify specific targets. In addition the *pyk1* mutant showed increased production of coelimycin on YEME and SMM plus pyruvate medium. As coelimycin has only recently had its structure resolved and a pathway proposed (Gomez-Escribano *et al.*, 2012), this might be an interesting starting point to further study this polyketide.

The screening of mutants from the metabolic node revealed that several members are interesting targets, many of which are involved directly in pyruvate metabolism. For pyruvate kinase, PFK (Borodina *et al.*, 2008; Siebring, 2010) and the malic enzyme (Rodríguez *et al.*, 2012), it could be shown experimentally that the presence of isoenzymes usually has different physiological roles and preliminary data on Ppdk also suggests different roles for the two enzymes. The two malic enzymes have different cofactor dependencies and it was demonstrated that the SCO2951 mutant and the double mutant displayed decreased actinorhodin production, due to reduced expression of *actIIORF4*. The double mutant also had decreased triacylglycerol storage levels during exponential growth (Rodríguez *et al.*, 2012). Previous work demonstrated activity for PfkA2 and PfkA3; however PfkA1 was found to exhibit no PFK activity (Siebring, 2010). PfkA2 and PfkA3 appear to have arisen through a gene duplication event (Figure 3.4), due to the topology of the tree. Deletion of *pfkA2* led to an increase of undecylprodigiosin and actinorhodin production on certain media, whereas deletion of either *pfkA1* or *pfkA3* did not show overproduction of either a specialised metabolite. Mutants in the *pfkA2* gene exhibited higher intracellular pools of glucose-6-phosphate and fructose-6-phosphate and radio-labeling experiments indicated increased flux through the pentose phosphate pathway (PPP) and concomitant increased levels of reduced nicotinamide adenine dinucleotide phosphate (NADPH, Borodina *et al*., 2008).

In case of the PDHC, the situation seems to be more complicated as there are two different type of $PDHC_{E1}$ either a heteromeric Gram-positive $PDHC_{E1}$ composed of an α and β chain or the Gram negative like homomeric $PDHC_{E1}$. Neither expression data nor literature proteomics data revealed any conclusive information. As only one mutant has been analysed this far it is necessary to construct the other mutants to see if they show similar overproduction of actinorhodin. Even if different carbon sources did not yield a change in expression of the

different genes, maybe a time course would be of interest or fluorescently tagged enzymes to localize them in the *Streptomyces* and observe if they are necessary. Not much information has been gained on the phylogeny as the sequences were too diverse yielding subsequently poor alignments. This might suggest that some have been acquired horizontally; however at this point it is only a hypothesis.

Metabolic expansion in central carbon metabolism could potentially play a role for precursor supply; however further experiments and passing the work performed here into industrial strains will be the next steps. Clearly the expansions do not seem to be redundant as the data suggests they have distinct roles in the physiology of the organism in several examples. The database created in this thesis is biased for the genus of *Streptomyces*, in order to analyse actinobacterial specific expansion events it could be useful to incorporate genomes of other bacteria and also include more species for underexplored genera to increase resolution. This would be a starting point in order to identify patterns for specific expansions and specialised metabolites.

In conclusion, targets for metabolic engineering were identified to increase polyketide production and the targets were analysed for their evolution, role in growth as well as production of polyketides. The analysis performed in this thesis also helped to understand better the presence of metabolic expansion in central carbon metabolism and how it helps the metabolic flexibilities and this knowledge can be exploited for strain construction.

Future experiments could be to fluorescently label Pyk1 and Pyk2, as well as for Ppdk1 and PpdK2 to check for different localisation in the different cell types of the *Streptomyces*. Additionally mutants for *ppdk1* and *ppdk2* could be constructed to study the phenotypes and potentially to determine the role of Ppdk in spore germination.

Regarding the $PDHC_{E1}$ mutant, future work would be to construct deletion mutants in all of the remaining genes, to see if the results are consistent with the actinorhodin overproducing phenotype. The failed complementation might be analysed by studying the expression of the actinorhodin pump to determine whether or not there is a difference in the WT and the mutant strain.

# 9 REFERENCES


Abrudan, M.I., Smakman, F., Grimbergen, A.J., Westhoff, S., Miller, E.L., Wezel, G.P. van, and Rozen, D.E. (2015) Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proc Natl Acad Sci U S A* **112**: 11054–11059.

Adamidis, T., and Champness, W.C. (1992) Genetic analysis of *absB*, a *Streptomyces coelicolor* locus involved in global antibiotic regulation. *J Bacteriol* **174**: 4622–4628.

Ahmed, Z.U., Shapiro, S., and Vining, L.C. (1984) Excretion of α-keto acids by strains of *Streptomyces venezuelae*. *Can J Microbiol* **30**: 1014–1021.

Aich, S., Imabayashi, F., and Delbaere, L.T.J. (2003) Expression, purification, and characterization of a bacterial GTP-dependent PEP carboxykinase. *Protein Expr Purif* **31**: 298–304.

Aínsa, J.A., Ryding, N.J., Hartley, N., Findlay, K.C., Bruton, C.J., and Chater, K.F. (2000) WhiA, a protein of unknown function conserved among gram-positive bacteria, is essential for sporulation in *Streptomyces coelicolor* A3(2). *J Bacteriol* **182**: 5470–5478.

Akpe San Roman, S., Facey, P.D., Fernandez-Martinez, L., Rodriguez, C., Vallin, C., Sol, R. Del, and Dyson, P. (2010) A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in *Streptomyces coelicolor*. *Microbiology* **156**: 1719–1729.

Alam, M.T., Merlo, M.E., Hodgson, D. a, Wellington, E.M.H., Takano, E., and Breitling, R. (2010) Metabolic modeling and analysis of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* **11**: 202–211.

Allert, S., Ernest, I., Poliszczak, A., Opperdoes, F.R., and Michels, P.A.M. (1991) Molecular cloning and analysis of two tandemly linked genes for pyruvate kinase of *Trypanosoma brucei*. *Eur J Biochem* **200**: 19–27.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Alves, A.M.C.R., Euverink, G.J., Bibb, M.J., and Dijkhuizen, L. (1997) Identification of ATP-dependent phosphofructokinase as a regulatory step in the glycolytic pathway of the actinomycete *Streptomyces coelicolor* A3 (2). *Appl Environ Microbiol* **63**: 956–961.

Alves, A.M.C.R., Euverink, G.J.W., and Santos, H. (2001) Different Physiological Roles of ATP- and PP i -Dependent Phosphofructokinase Isoenzymes in the Methylotrophic Actinomycete *Amycolatopsis methanolica* . *J Bacteriol* **183**: 7231–7240.

Aminov, R.I. (2010) A brief history of the antibiotic era: lessons learned and challenges for the future. *Front Microbiol* **1**: 1–7.

Aoki, Y., Matsumoto, D., Kawaide, H., and Natsume, M. (2011) Physiological role of germicidins in spore germination and hyphal elongation in *Streptomyces coelicolor* A3(2). *J Antibiot (Tokyo)* **64**: 607–611.

Asención Diez, M.D., Peirú, S., Demonte, A.M., Gramajo, H., and Iglesias, A.A. (2012) Characterization of recombinant UDP- and ADP-glucose pyrophosphorylases and glycogen synthase to elucidate glucose-1-phosphate partitioning into oligo- and polysaccharides in *Streptomyces coelicolor*. *J Bacteriol* **194**: 1485–1493.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 1–11.

Bailey, J.E. (1991) Toward a Science of Metabolic Engineering. *Science (80- )* **252**: 1668–1675.

Bascarán, V., Hardisson, C., and Braña, A.F. (1989) Regulation of nitrogen catabolic enzymes in *Streptomyces clavuligerus*. *J Gen Microbiol* **135**: 2465–2474.

Becker, J., Klopprogge, C., and Wittmann, C. (2008) Metabolic responses to pyruvate kinase deletion in lysine producing *Corynebacterium glutamicum*. *Microb Cell Fact* **7**: 8.

Bentley, S.D., and Parkhill, J. (2004) Comparative genomic structure of prokaryotes. *Annu Rev Genet* **38**: 771–792.

Bentley, S.D., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Harper, D., Bateman, A., Brown, S., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Larke, L., Murphy, L., Oliver, K., Rabbinowitsch, E., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Woodward, J., Barrell, B.G., and Parkhill, J. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141–147.

Berdy, J. (2005) Bioactive Microbial Metabolites A Personal View. *J Antibiot (Tokyo)* **58**: 1–26.

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002) The Citric Acid Cycle Is a Source of Biosynthetic Precursors. In *Biochemistry*. W H Freeman, p. 1050.

Bergemeyer, U.H., Gawehn, K., and Grassl, M. (1974) Enzymatic Assay of Pyruvate Kinase. In *Methods of Enzymatic Analysis*. Bergemeyer, U.H. (ed.). Academic Press Inc, pp. 509–510.

Bergstrom, C.T., and Kerr, B. (2015) Microbiology: Taking the bad with the good. *Nature* **1**: 10–11.

Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* **62**: 293–300.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**: 252–258.

Bibb, M.J. (1996) The regulation of antibiotic production in *Streptomyces coelicolor* A3(2). *Microbiology* **142**: 1335–1344.

Bibb, M.J. (2005) Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol* **8**: 208–215.

Bibb, M.J., Molle, V., and Buttner, M.J. (2000) BldN , an Extracytoplasmic Function RNA Polymerase Sigma Factor Required for Aerial Mycelium Formation in *Streptomyces*. *J Bacteriol* **182**: 4606–1466.

Bignell, D.R., Warawa, J.L., Strap, J.L., Chater, K.F., and Leskiw, B.K. (2000) Study of the *bldG* locus suggests that an anti-anti-sigma factor and an anti-sigma factor may be involved in *Streptomyces coelicolor* antibiotic production and sporulation. *Microbiology* **146**: 2161–2173.

Birch, A.J. (1968) Polyketide Metabolism. *Annu Rev Plant Physiol* **19**: 321–332.

Birch, A.J., Massy-Westropp, R., and Moye, C. (1955) Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx. *Aust J Chem* **8**: 539–544.

Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**: 1513–1523.

Bishop, A., Fielding, S., Dyson, P., and Herron, P.R. (2004) Systematic Insertional Mutagenesis of a Streptomycete Genome : A Link Between Osmoadaptation and Antibiotic Production. *Genome Res* **14**: 893–900.

Blombach, B., Arndt, A., Auchter, M., and Eikmanns, B.J. (2009) L-valine production during growth of pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum* in the presence of ethanol or by inactivation of the transcriptional regulator SugR. *Appl Environ Microbiol* **75**: 1197–1200.

Bolten, C.J., Kiefer, P., Letisse, F., Portais, J.-C., and Wittmann, C. (2007) Sampling for metabolome analysis of microorganisms. *Anal Chem* **79**: 3843–3849.

Borodina, I., Siebring, J., Zhang, J., Smith, C.P., Keulen, G. van, Dijkhuizen, L., and Nielsen, J. (2008) Antibiotic overproduction in *Streptomyces coelicolor* A3(2) mediated by phosphofructokinase deletion. *J Biol Chem* **283**: 25186–25199.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–254.

Bramwell, H., Nimmo, H.G., Hunter, I.S., and Coggins, J.R. (1993) Phosphoenolpyruvate carboxylase from *Streptomyces coelicolor* A3(2): purification of the enzyme, cloning of the ppc gene and over-expression of the protein in a streptomycete. *Biochem J* **293**: 131–136.

Braña, A.F., Méndez, C., Díaz, L.A., Manzanal, M.B., and Hardisson, C. (1986) Glycogen and trehalose accumulation during colony development in *Streptomyces antibioticus*. *J Gen Microbiol* **132**: 1319–1326.

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R.A., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., and Xia, F. (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* **5**: 8365.

Bruton, C.J., Plaskitt, K. a, and Chater, K.F. (1995) Tissue-specific glycogen branching isoenzymes in a multicellular prokaryote *Streptomyces coelicolor* A3(2). *Mol Microbiol* **18**: 89–99.

Buchholz, J., Schwentner, A., Brunnenkan, B., Gabris, C., Grimm, S., Gerstmeir, R., Takors, R., Eikmanns, B.J., and Blombach, B. (2013) Platform engineering of *Corynebacterium glutamicum* with reduced pyruvate dehydrogenase complex activity for improved production of L-lysine, L-valine, and 2-ketoisovalerate. *Appl Environ Microbiol* **79**: 5566–5575.

Butler, M.J., Bruheim, P., Jovetic, S., Marinelli, F., Postma, P.W., and Bibb, M.J. (2002) Engineering of Primary Carbon Metabolism for Improved Antibiotic Production in *Streptomyces lividans*. *Appl Environ Microbiol* **68**: 4731–4739.

Cai, X., Hu, H., and Li, X. (2009) A new measurement of sequence conservation. *BMC Genomics* **10**: 623.

Case, R.J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W.F., and Kjelleberg, S. (2007) Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* **73**: 278–288.

Cavaletti, L., Monciardini, P., Schumann, P., Rohde, M., Bamonte, R., Busti, E., Sosio, M., and Donadio, S. (2006) *Actinospica robiniae* gen. nov., sp. nov. and *Actinospica acidiphila* sp. nov.: proposal for Actinospicaceae fam. nov. and Catenulisporinae subord. nov. in the order Actinomycetales. *Int J Syst Evol Microbiol* **56**: 1747–1753.

Chakraburtty, R., and Bibb, M.J. (1997) The ppGpp synthetase gene (*relA*) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation. *J Bacteriol* **179**: 5854–5861.

Challis, G.L. (2008) Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* **154**: 1555–1569.

Challis, G.L. (2014) Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J Ind Microbiol Biotechnol* **41**: 219–232.

Challis, G.L., and Hopwood, D.A. (2003) Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc Natl Acad Sci U S A* **100**: 14555–14561.

Champness, W.C. (1988) New loci required for *Streptomyces coelicolor* morphological and physiological differentiation. *J Bacteriol* **170**: 1168–1174.

Chastain, C.J., Heck, J.W., Colquhoun, T. a, Voge, D.G., and Gu, X.-Y. (2006) Posttranslational regulation of pyruvate, orthophosphate dikinase in developing rice (*Oryza sativa*) seeds. *Planta* **224**: 924–934.

Chater, K.F. (1993) Genetics of differentiation in *Streptomyces*. *Annu Rev Microbiol* **47**: 685–713.

Chater, K.F., Biró, S., Lee, K.J., Palmer, T., and Schrempf, H. (2010) The complex extracellular biology of *Streptomyces*. *FEMS Microbiol Rev* **34**: 171–198.

Chater, K.F., and Chandra, G. (2006) The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev* **30**: 651–672.

Chi, W.-J., Jin, X.-M., Jung, S.-C., Oh, E. a, and Hong, S.-K. (2011) Characterization of Sgr3394 produced only by the A-factor-producin *Streptomyces griseus* IFO 13350, not by the A-factor deficient mutant. *J Microbiol* **49**: 155–160.

Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Wieland Brown, L.C., Mavrommatis, K., Pati, A., Godfrey, P.A., Koehrsen, M., Clardy, J., Birren, B.W., Takano, E., Sali, A., Linington, R.G., and Fischbach, M.A. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**: 412–421.

Ciria, R., Abreu-Goodger, C., Morrett, E., and Merino, E. (2004) GeConT: gene context analysis. *Bioinformatics* **20**: 2307–2308.

Clark, L.C., and Hoskisson, P. a (2011) Duplication and evolution of *devA*-like genes in *Streptomyces* has resulted in distinct developmental roles. *PLoS One* **6**: e25049.

Cochrane, V.W., and Dimmick, I. (1949) The metabolism of species of *Streptomyces*; the formation of succinic and other acids. *J Bacteriol* **58**: 723–730.

Collie, N., and Myers, W.S. (1893) VII.The formation of orcinol and other condensation products from dehydracetic acid. *J Chem Soc Trans* **63**: 122–128.

Copley, S.D. (2003) Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. *Curr Opin Chem Biol* **7**: 265–272.

Copley, S.D. (2012) Moonlighting is mainstream: Paradigm adjustment required. *BioEssays* **34**: 578–588.

Copley, S.D. (2014) An evolutionary perspective on protein moonlighting. *Biochem Soc Trans* **42**: 1684–1691.

Copley, S.D. (2015) An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci* **40**: 72–78.

Cornish Bowen, A. (2012) *Fundamentals of Enzyme Kinetics*. 4th ed., Wiley-Blackwell, .

Craney, A., Ozimok, C., Pimentel-Elardo, S.M., Capretta, A., and Nodwell, J.R. (2012) Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem Biol* **19**: 1020–1027.

Creek, D.J., Jankevics, A., Burgess, K.E. V, Breitling, R., and Barrett, M.P. (2012) IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics* **28**: 1048–1049.

Cruz-Morales, P., Martínez-Guerrero, C.E., Morales-Escalante, M.A., Yáñez-Guerra, L.A., Kopp, J.F., Feldmann, J., Ramos-Aboites, H.E., and Barona-Gómez, F. (2015) Recapitulation of the evolution of biosynthetic gene clusters reveals hidden chemical diversity on bacterial genomes. *BioRxiv* .

Cunningham, C.C., and Hager, L.P. (1975) Reactivation of the lipid-depleted pyruvate oxidase system from *Escherichia coli* with cell envelope neutral lipids. *J Biol Chem* **250**: 7139–7146.

D'Alia, D., Eggle, D., Nieselt, K., Hu, W.-S., Breitling, R., and Takano, E. (2011) Deletion of the signalling molecule synthase ScbA has pleiotropic effects on secondary metabolite biosynthesis, morphological differentiation and primary metabolism in *Streptomyces coelicolor* A3(2). *Microb Biotechnol* **4**: 239–251.

D'Huys, P.-J., Lule, I., Hove, S. Van, Vercammen, D., Wouters, C., Bernaerts, K., Anné, J., and Impe, J.F.M. Van (2011) Amino acid uptake profiling of wild type and recombinant *Streptomyces lividans* TK24 batch fermentations. *J Biotechnol* **152**: 132–143.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640–6645.

Davies, J. (2013) Specialized microbial metabolites: Functions and Origins. *J Antibiot (Tokyo)* **66**: 361–364.

Davies, J., and Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**: 417–433.

Davies, J., Spiegelman, G.B., and Yim, G. (2006) The world of subinhibitory antibiotic concentrations. *Curr Opin Mcrobiology* **9**: 445–453.

Dekleva, M.L., and Strohl, W.R. (1987) Glucose-stimulated acidogenesis by *Streptomyces peucetius*. *Can J Microbiol* **33**: 1129–1132.

Demain, A.L., and Sanchez, S. (2009) Microbial drug discovery: 80 years of progress. *J Antibiot (Tokyo)* **62**: 5–16.

Dhillon, N., Hale, R.S., Cortes, J., and Leadlay, P.F. (1989) Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. *Mol Microbiol* **3**: 1405–1414.

Dilks, K., Rose, R.W., Hartmann, E., and Pohlschröder, M. (2003) Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey. *J Bacteriol* **185**: 1478–1483.

Doroghazi, J.R., Albright, J.C., Goering, A.W., Ju, K.-S., Haines, R.R., Tchalukov, K. a, Labeda, D.P., Kelleher, N.L., and Metcalf, W.W. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* **10**: 963–968.

Doroghazi, J.R., and Metcalf, W.W. (2013) Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**: 611.

Doull, J.L., and Vining, L.C. (1990) Nutritional control of actinorhodin production by *Streptomyces coelicolor* A3 (2): suppressive effects of nitrogen and phosphate. *Appl Microbiol Biotechnol* **32**: 449–454.

Driessen, A.J.M., and Nouwen, N. (2008) Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem* **77**: 643–667.

Dunn, M.F., Araíza, G., and Finan, T.M. (2001) Cloning and characterization of the pyruvate carboxylase from *Sinorhizobium meliloti* Rm1021. *Arch Microbiol* **176**: 355–363.

Edgar, R.C. (2004a) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Edgar, R.C. (2004b) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Eick, K., and Pohnert, G. (2015) Simplifying Complexity in Metabolomics. *Chem Biol* **22**: 567–568.

Eikmanns, B.J., and Blombach, B. (2014) The pyruvate dehydrogenase complex of *Corynebacterium glutamicum*: an attractive target for metabolic engineering. *J Biotechnol* **192**: 339–345.

Embley, T.M., and Stackebrandt, E. (1994) The molecular phylogeny and systematics of the actinomycetes. *Annu Rev Microbiol* **48**: 257–289.

Emmerling, M., Bailey, J.E., and Sauer, U. (1999) Glucose catabolism of Escherichia colistrains with increased activity and altered regulation of key glycolytic enzymes. *Metab Eng* **1**: 117–127.

Ernest, I., Callens, M., Opperdoes, F.R., and Michels, P.A. (1994) Pyruvate kinase of *Leishmania mexicana mexicana*. Cloning and analysis of the gene, overexpression in *Escherichia coli* and characterization of the enzyme. *Mol Biochem Parasitol* **64**: 43–54.

Ernest, I., Callens, M., Uttaro, A.D., Chevalier, N., Opperdoes, F.R., Muirhead, H., and Michels, P.A. (1998) Pyruvate kinase of *Trypanosoma brucei*: overexpression, purification, and functional characterization of wild-type and mutated enzyme. *Protein Expr Purif* **13**: 373–382.

Evans, H.J., and Wood, H.G. (1968) The mechanism of the pyruvate, phosphate dikinase reaction. *Proc Natl Acad Sci U S A* **61**: 1448–1453.

Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**: 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175–185.

Feitelson, J.S., Malpartida, F., and Hopwood, D.A. (1985) Genetic and biochemical characterization of the red gene cluster of *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **131**: 2431–2441.

Feng, X.-M., Cao, L.-J., Adam, R.D., Zhang, X.-C., and Lu, S.-Q. (2008) The catalyzing role of PPDK in *Giardia lamblia*. *Biochem Biophys Res Commun* **367**: 394–398.

Fernández-Martínez, L.T., Sol, R. Del, Evans, M.C., Fielding, S., Herron, P.R., Chandra, G., and Dyson, P.J. (2011) A transposon insertion single-gene knockout library and new ordered cosmid library for the model organism *Streptomyces coelicolor* A3(2). *Antonie Van Leeuwenhoek* **99**: 515–522.

Fink, D., Weissschuh, N., Reuther, J., Wohlleben, W., and Engels, A. (2002) Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **46**: 331–347.

Firn, R.D., and Jones, C.G. (2000) MicroOpinion The evolution of secondary metabolism ± a unifying model. *Mol Microbiol* **37**: 989–994.

Firn, R.D., and Jones, C.G. (2009) A Darwinian view of metabolism: molecular properties determine fitness. *J Exp Bot* **60**: 719–726.

Fisher, R.A. (1934) Statistical Methods for Research Workers. In *Biological Monographs and Manuals*. Crew, F.A.E., and Cutler, D.W. (eds). Edinburgh. pp. 1–336.

Flärdh, K., and Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7**: 36–49.

Forrester, P.I., and Gaucher, G.M. (1972) Conversion of 6-methylsalicylic acid into patulin by *Penicillium urticae*. *Biochemistry* **11**: 1102–1107.

Fowler-Goldsworthy, K., Gust, B., Mouz, S., Chandra, G., Findlay, K.C., and Chater, K.F. (2011) The actinobacteria-specific gene *wblA* controls major developmental transitions in *Streptomyces coelicolor* A3(2). *Microbiology* **157**: 1312–1328.

Gao, C., Mulder, D., Yin, C., and Elliot, M.A. (2012) Crp Is a Global Regulator of Antibiotic Production in *Streptomyces*. **3**: 407–412.

Gatewood, M.L., and Jones, G.H. (2010) (p) ppGpp Inhibits Polynucleotide Phosphorylase from *Streptomyces* but Not from *Escherichia coli* and Increases the Stability of Bulk mRNA in *Streptomyces coelicolor*. **192**: 4275–4280.

Girard, G., Traag, B.A., Sangal, V., Mascini, N., Hoskisson, P.A., Goodfellow, M., and Wezel, G.P. van (2013) A novel taxonomic marker that discriminates between morphologically complex actinomycetes. *Open Biol* **3**: 130073.

Glauert, A.M., and Hopwood, D.A. (1960) The fine structure of *Streptomyces coelicolor*. I. The cytoplasmic membrane system. *J Biophys Biochem Cytol* **7**: 479–488.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**: R86.

Gomez-Escribano, J.P., and Bibb, M.J. (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* **4**: 207–215.

Gomez-Escribano, J.P., Song, L., Fox, D.J., Yeo, V., Bibb, M.J., and Challis, G.L. (2012) Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. *Chem Sci* **3**: 2716.

Gottelt, M., Kol, S., Gomez-Escribano, J.P., Bibb, M.J., and Takano, E. (2010) Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology* **156**: 2343–2353.

Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci U S A* **87**: 4645–4649.

Gubbens, J., Janus, M., Florea, B.I., Overkleeft, H.S., and Wezel, G.P. van (2012) Identification of glucose kinase-dependent and -independent pathways for carbon control of primary metabolism, development and antibiotic production in *Streptomyces coelicolor* by quantitative proteomics. *Mol Microbiol* **86**: 1490–1507.

Gubbens, J., Zhu, H., Girard, G., Song, L., Florea, B.I., Aston, P., Ichinose, K., Filippov, D. V, Choi, Y.H., Overkleeft, H.S., Challis, G.L., and Wezel, G.P. van (2014) Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products. *Chem Biol* **21**: 707–718.

Gubler, M.E., Jetten, M.S., Lee, S.H., and Sinskey, a J. (1994) Cloning of the pyruvate kinase gene (*pyk*) of *Corynebacterium glutamicum* and site-specific inactivation of *pyk* in a lysine-producing *Corynebacterium lactofermentum* strain. *Appl Environ Microbiol* **60**: 2494–2500.

Guerra, S.M., Rodríguez-García, A., Santos-Aberturas, J., Vicente, C.M., Payero, T.D., Martín, J.F., and Aparicio, J.F. (2012) LAL regulators SCO0877 and SCO7173 as pleiotropic modulators of phosphate starvation response and actinorhodin biosynthesis in *Streptomyces coelicolor*. *PLoS One* **7**: e31475.

Guex, N., and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**: 2714–2723.

Gunnarsson, N., Eliasson, A., and Nielsen, J. (2004) Control of fluxes towards antibiotics and the role of primary metabolism in production of antibiotics. *Adv Biochem Eng Biotechnol* **88**: 137–178.

Guo, C.-J., and Wang, C.C.C. (2014) Recent advances in genome mining of secondary metabolites in Aspergillus terreus. *Front Microbiol* **5**: 717.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **100**: 1541–1546.

Gust, B., Kieser, T., and Chater, K.F. (2002) PCR targeting system in Streptomyces coelicolor A3(2). *Protoc John Innes Cent* **1**: 1–39.

Hartmann, T. (2008) The lost origin of chemical ecology in the late 19th century. *Proc Natl Acad Sci U S A* **105**: 4541–4546.

Hatanaka, H., Kino, T., Asano, M., Goto, T., Tanaka, H., and Okuhara, M. (1989) FK-506 related compounds produced by *Streptomyces tsukubaensis* No. 9993. *J Antibiot (Tokyo)* **42**: 620–622.

Hatch, M.D., and Slack, C.R. (1969) Studies on the mechanism of activation and inactivation of pyruvate, phosphate dikinase. A possible regulatory role for the enzyme in the C4 dicarboxylic acid pathway of photosynthesis. *Biochem J* **112**: 549–558.

Hengst, C.D. den, Tran, N.T., Bibb, M.J., Chandra, G., Leskiw, B.K., and Buttner, M.J. (2010) Genes essential for morphological development and antibiotic production in *Streptomyces coelicolor* are targets of BldD during vegetative growth. *Mol Microbiol* **78**: 361–379.

Hesketh, A., Chandra, G., Shaw, a D., Rowland, J.J., Kell, D.B., Bibb, M.J., and Chater, K.F. (2002) Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*. *Mol Microbiol* **46**: 917–932.

Higo, A., Horinouchi, S., and Ohnishi, Y. (2011) Strict regulation of morphological differentiation and secondary metabolism by a positive feedback loop between two global regulators AdpA and BldA in *Streptomyces griseus*. *Mol Microbiol* **81**: 1607–1622.

Hiltner, J.K., Hunter, I.S., and Hoskisson, P.A. (2015) Tailoring Specialized Metabolite Production in Streptomyces. In *Advances in Applied Microbiology*. Sariaslani, S., and Gadd, G.M. (eds). Elsevier Ltd, pp. 237–255.

Hobbs, G., Frazer, C.M., Gardner, D.C.J., Cullum, J.A., and Oliver, S.G. (1989) Dispersed growth of Streptomyces in liquid culture. *Appl Microbiol Biotechnol* **31**: 272–277.

Hobbs, G., Frazer, C.M., Gardner, D.C.J., Flett, F., and Oliver, S.G. (1990) Pigmented antibiotic production by *Streptomyces coelicolor* A3(2): kinetics and the influence of nutrients. *J Gen Microbiol* **136**: 2291–2296.

Hobbs, G., Obanye, A.I., Petty, J., Mason, J.C., Barratt, E., Gardner, D.C., Flett, F., Smith, C.P., Broda, P., and Oliver, S.G. (1992) An integrated approach to studying regulation of production of the antibiotic methylenomycin by *Streptomyces coelicolor* A3(2). *J Bacteriol* **174**: 1487–1494.

Hodgson, D.A. (2000) Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol* **42**: 47–238.

Hoffman, L.R., D'Argenio, D.A., MacCoss, M.J., Zhang, Z., Jones, R.A., and Miller, S.I. (2005) Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature* **436**: 1171–1175.

Hopwood, D.A. (1967) Genetic analysis and genome structure in *Streptomyces coelicolor*. *Bacteriol Rev* **31**: 373–403.

Hopwood, D.A. (1999) Forty years of genetics with Streptomyces : from *in vivo* through *in vitro* to in silico. *Microbiology* **145**: 2183–2202.

Hopwood, D.A. (2007) Streptomyces *in Nature and Medicine: The Antibiotic Makers*. .

Hopwood, D.A., and Wright, H.M. (1983) CDA is a new chromosomally-determined antibiotic from *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **129**: 3575–3579.

Horinouchi, S., Malpartida, F., Hopwood, D.A., and Beppu, T. (1989) *afsB* stimulates transcription of the actinorhodin biosynthetic pathway in *Streptomyces coelicolor* A3(2) and *Streptomyces lividans*. *Mol Gen Genet* **215**: 355–357.

Hoskisson, P.A., England, R., Sharples, G.P., and Hobbs, G. (2004) Modulation of glycogen and trehalose levels in *Micromonospora echinospora* (ATCC 15837). *Antonie Van Leeuwenhoek* **86**: 225–233.

Hoskisson, P.A., Hobbs, G., and Sharples, G.P. (2000) Response of *Micromonospora echinospora* (NCIMB 12744) spores to heat treatment with evidence of a heat activation phenomenon. *Lett Appl Microbiol* **30**: 114–117.

Hoskisson, P.A., Rigali, S., Fowler, K., Kim, C., Buttner, M.J., and Findlay, K.C. (2006) DevA , a GntR-Like Transcriptional Regulator Required for Development in *Streptomyces coelicolor*. *J Bacteriol* **188**: 5014–5023.

Howe, K., Bateman, A., and Durbin, R. (2002) QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics* **18**: 1546–1547.

Hu, L., Olsen, C., Maddux, N., Joshi, S.B., Volkin, D.B., and Middaugh, C.R. (2011) Investigation of protein conformational stability employing a multimodal spectrometer. *Anal Chem* **83**: 9399–9405.

Huang, H., and Grove, A. (2013) The transcriptional regulator TamR from *Streptomyces coelicolor* controls a key step in central metabolism during oxidative stress. *Mol Microbiol* **87**: 1151–1166.

Huang, H., Sivapragasam, S., and Grove, A. (2015) The regulatory role of *Streptomyces coelicolor* TamR in central metabolism. *Biochem J* **466**: 347–358.

Huang, J., Shi, J., Molle, V., Sohlberg, B., Weaver, D., Bibb, M.J., Karoonuthaisiri, N., Lih, C.-J., Kao, C.M., Buttner, M.J., and Cohen, S.N. (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol Microbiol* **58**: 1276–1287.

Huggett, J.F., Foy, C.A., Benes, V., Emslie, K., Garson, J.A., Haynes, R., Hellemans, J., Kubista, M., Mueller, R.D., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., and Bustin, S.A. (2013) The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. *Clin Chem* **59**: 892–902.

Huminiecki, L., and Wolfe, K.H. (2004) Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome Res* **14**: 1870–1879.

Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* **18**: 486.

Hyeon, J.E., Kang, D.H., Kim, Y.I., You, S.K., and Han, S.O. (2012) GntR-type transcriptional regulator PckR negatively regulates the expression of phosphoenolpyruvate carboxykinase in *Corynebacterium glutamicum*. *J Bacteriol* **194**: 2181–2188.

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., Mueller, O., Schroeder, A., and Auffray, C. (2005) Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res* **33**: e56.

Ish-Horowicz, D., and Burke, J.F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Res* **9**: 2989–2998.

Ives, P.R., and Bushell, M.E. (1997) Manipulation of the physiology of clavulanic acid production in *Streptomyces clavuligerus*. *Microbiology* **143**: 3573–3579.

Jenke-Kodama, H., and Dittmann, E. (2005) Combinatorial polyketide biosynthesis at higher stage. *Mol Syst Biol* **1**: 25–26.

Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005) Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* **22**: 2027–2039.

Jetten, M.S., Gubler, M.E., Lee, S.H., and Sinskey, A.J. (1994) Structural and functional analysis of pyruvate kinase from *Corynebacterium glutamicum*. *Appl Environ Microbiol* **60**: 2501–2507.

Jiang, Y., Wiese, J., Tang, S.-K., Xu, L.-H., Imhoff, J.F., and Jiang, C.-L. (2008) *Actinomycetospora chiangmaiensis* gen. nov., sp. nov., a new member of the family Pseudonocardiaceae. *Int J Syst Evol Microbiol* **58**: 408–413.

Jin, Y.-Y., Cheng, J., Yang, S.H., Meng, L., Palaniyandi, S.A., Zhao, X.-Q., and Suh, J.-W. (2011) S-adenosyl-L-methionine activates actinorhodin biosynthesis by increasing autophosphorylation of the Ser/Thr protein kinase AfsK in Streptomyces coelicolor A3(2). *Biosci Biotechnol Biochem* **75**: 910–913.

Jurica, M.S., Mesecar, A., Heath, P.J., Shi, W., Nowak, T., and Stoddard, B.L. (1998) The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure* **6**: 195–210.

Kaewkla, O., and Franco, C.M.M. (2013) *Kribbella endophytica* sp. nov., an endophytic actinobacterium isolated from the surface-sterilized leaf of a native apricot tree. *Int J Syst Evol Microbiol* **63**: 1249–1253.

Kale, S., Arjunan, P., Furey, W., and Jordan, F. (2007) A dynamic loop at the active center of the *Escherichia coli* pyruvate dehydrogenase complex E1 component modulates substrate utilization and chemical communication with the E2 component. *J Biol Chem* **282**: 28106–28116.

Kämpfer, P., and Glaeser, S.P. (2012) Prokaryotic taxonomy in the sequencing era - the polyphasic approach revisited. *Environ Microbiol* **14**: 291–317.

Kang, S.G., Jin, W., Bibb, M.J., and Lee, K.J. (1998) Actinorhodin and undecylprodigiosin production in wild-type and *relA* mutant strains of *Streptomyces coelicolor* A3(2) grown in continuous culture. *FEMS Microbiol Lett* **168**: 221–226.

Karandikar, A., Sharples, G.P., and Hobbs, G. (1997) Differentiation of *Streptomyces coelicolor* A3(2) under nitrate-limited conditions. *Microbiology* **143**: 3581–3590.

Kayne, F.J., and Price, N.C. (1973) Amino acid effector binding to rabbit muscle pyruvate kinase. *Arch Biochem Biophys* **159**: 292–296.

Kelemen, G.H., Brown, G.L., Kormanec, J., Potúcková, L., Chater, K.F., and Buttner, M.J. (1996) The positions of the sigma-factor genes, *whiG* and *sigF*, in the hierarchy controlling the development of spore chains in the aerial hyphae of *Streptomyces coelicolor* A3(2). *Mol Microbiol* **21**: 593–603.

Kelsic, E.D., Zhao, J., Vetsigian, K., and Kishony, R. (2015) Counteraction of antibiotic production and degradation stabilizes microbial communities. *Nat Lett* In Press.

Keulen, G. van, and Dyson, P.J. (2014) Production of specialized metabolites by *Streptomyces coelicolor* A3(2). In *Advances in Applied Microbiology*. Sariaslani, S., and Gadd, G.M. (eds). Elsevier Inc., pp. 217–266.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000) *Practical Streptomyces Genetics*. 1st ed., John Innes Foundation, Norwich, UK.

Kim, E.S., Hong, H.J., Choi, C.Y., and Cohen, S.N. (2001) Modulation of actinorhodin biosynthesis in *Streptomyces lividans* by glucose repression of *afsR2* gene transcription. *J Bacteriol* **183**: 2198–2203.

Kim, J., and Copley, S.D. (2007) Why metabolic enzymes are essential or nonessential for growth of *Escherichia coli* K12 on glucose. *Biochemistry* **46**: 12501–12511.

Klaffl, S., and Eikmanns, B.J. (2010) Genetic and functional analysis of the soluble oxaloacetate decarboxylase from *Corynebacterium glutamicum*. *J Bacteriol* **192**: 2604–2612.

Knowles, V.L., Dennis, D.T., and Plaxton, W.C. (1989) Purification of a novel pyruvate kinase from a green alga. *FEBS Lett* **259**: 130–132.

Knowles, V.L., Smith, C.S., Smith, C.R., and Plaxton, W.C. (2001) Structural and regulatory properties of pyruvate kinase from the *Cyanobacterium synechococcus* PCC 6301. *J Biol Chem* **276**: 20966–20972.

Koffas, M.A.G., and Stephanopoulos, G. (2005) Strain improvement by metabolic engineering: lysine production as a case study for systems biology. *Curr Opin Biotechnol* **16**: 361–366.

Kok, A. de, Hengeveld, A.F., Martin, A., and Westphal, A.H. (1998) The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria. *Biochim Biophys Acta* **1385**: 353–366.

Koland, J.G., Miller, M.J., and Gennis, R.B. (1984) Reconstitution of the membrane-bound, ubiquinone-dependent pyruvate oxidase respiratory chain of *Escherichia coli* with the cytochrome d terminal oxidase. *Biochemistry* **23**: 445–453.

Kolbe, S., Fischer, S., Becirevic, A., Hinz, P., and Schrempf, H. (1998) The *Streptomyces reticuli* α-chitin-binding protein CHB2 and its gene. *Microbiology* **144**: 1291–1297.

Konstantinidis, K.T., and Tiedje, J.M. (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci U S A* **101**: 3160–3165.

Krügel, H., Krubasik, P., Weber, K., Saluz, H.P., and Sandmann, G. (1999) Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. *Biochim Biophys Acta* **1439**: 57–64.

Kudo, T., Itoh, T., Miyadoh, S., Shomura, T., and Seino, A. (1993) *Herbidospora* gen. nov., a new genus of the family Streptosporangiaceae Goodfellow *et al*. 1990. *Int J Syst Bacteriol* **43**: 319–328.

Labeda, D.P., Goodfellow, M., Brown, R., Ward, a C., Lanoot, B., Vanncanneyt, M., Swings, J., Kim, S.-B., Liu, Z., Chun, J., Tamura, T., Oguchi, a, Kikuchi, T., Kikuchi, H., Nishii, T., Tsuji, K., Yamaguchi, Y., Tase, a, Takahashi, M., Sakane, T., Suzuki, K.I., and Hatano, K. (2012) Phylogenetic study of the species within the family Streptomycetaceae. *Antonie Van Leeuwenhoek* **101**: 73–104.

Labeda, D.P., Hatano, K., Kroppenstedt, R.M., and Tamura, T. (2001) Revival of the genus *Lentzea* and proposal for *Lechevalieria* gen. nov. *Int J Syst Evol Microbiol* **51**: 1045–1050.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Lechevalier, H.A., Lechevalier, M.P., and Holbert, P.E. (1966) Electron microscopic observation of the sporangial structure of strains of Actinoplanaceae. *J Bacteriol* **92**: 1228–1235.

Li, R., and Townsend, C. a (2006) Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in *Streptomyces clavuligerus*. *Metab Eng* **8**: 240–252.

Linares, J.F., Gustafsson, I., Baquero, F., and Martinez, J.L. (2006) Antibiotics as intermicrobial signaling agents instead of weapons. *Proc Natl Acad Sci U S A* **103**: 19484–19489.

Lu, Y., He, J., Zhu, H., Yu, Z., Wang, R., Chen, Y., Dang, F., Zhang, W., Yang, S., and Jiang, W. (2011) An orphan histidine kinase, OhkA, regulates both secondary metabolism and morphological differentiation in *Streptomyces coelicolor*. *J Bacteriol* **193**: 3020–3032.

Lucas, X., Senger, C., Erxleben, A., Grüning, B. a, Döring, K., Mosch, J., Flemming, S., and Günther, S. (2013) StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Res* **41**: 1130–1136.

Luti, K.J.K., and Mavituna, F. (2011) *Streptomyces coelicolor* increases the production of undecylprodigiosin when interacted with *Bacillus subtilis*. *Biotechnol Lett* **33**: 113–118.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**: 61–68.

Madden, T., Ward, J.M., and Ison, A.P. (1996) Organic acid excretion by *Streptomyces lividans* TK24 during growth on defined carbon and nitrogen sources. *Microbiology* **142**: 3181–3185.

Maloy, S.R., Stewart, V.J., and Taylor, R.K. (1996) Genetic analysis of pathogenic bacteria: a laboratory manual. In *Genetic analysis of pathogenic bacteria: a laboratory manual.* Cold Spring Harbor Laboratory Press., p. 603.

Malpartida, F., and Hopwood, D.A. (1986) Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). *Mol Gen Genet* **205**: 66–73.

Maréchal, P. Le, Decottignies, P., Marchand, C.H., Degrouard, J., Jaillard, D., Dulermo, T., Froissard, M., Smirnov, A., Chapuis, V., and Virolle, M.-J. (2013) Comparative proteomic analysis of *Streptomyces lividans* wild-type and ppk mutant strains reveals the importance of storage lipids for antibiotic biosynthesis. *Appl Environ Microbiol* **79**: 5907–5917.

Marshall, J.S., Ashton, A.R., Govers, F., and Hardham, A.R. (2001) Isolation and characterization of four genes encoding pyruvate, phosphate dikinase in the oomycete plant pathogen *Phytophthora cinnamomi*. *Curr Genet* **40**: 73–81.

Martín, J.F., and Gil, J.A. (1984) Cloning and Expression of Antibiotic Production Genes. *Nat Biotechnol* **2**: 63–72.

Martin, M.C., Schneider, D., Bruton, C.J., Chater, K.F., and Hardisson, C. (1997) A *glgC* gene essential only for the first of two spatially distinct phases of glycogen synthesis in *Streptomyces coelicolor* A3(2). *J Bacteriol* **179**: 7784–7789.

Mattevi, A., Valentini, G., Rizzi, M., Speranza, M.L., Bolognesi, M., and Coda, A. (1995) Crystal structure of *Escherichia coli* pyruvate kinase type I: molecular basis of the allosteric transition. *Structure* **3**: 729–741.

Mazodier, P., Petter, R., and Thompson, C. (1989) Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *J Bacteriol* **171**: 3583–3585.

McBride, M.J., and Ensign, J.C. (1987) Metabolism of endogenous trehalose by *Streptomyces griseus* spores and by spores or cells of other actinomycetes. *J Bacteriol* **169**: 5002–5007.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. a, Vanderpool, C.K., and Tjaden, B. (2013) Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res* **41**: 140.

McGregor, J.F. (1954) Nuclear division and the life cycle in a *Streptomyces* sp. *J Gen Microbiol* **11**: 52–56.

Medema, M.H., Blin, K., Cimermancic, P., Jager, V. de, Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., and Breitling, R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* **39**: 339–346.

Medema, M.H., and Fischbach, M.A. (2015) Computational approaches to natural product discovery. *Nat Chem Biol* **11**: 639–648.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Glöckner, F.O., *et al.* (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol* **11**: 625–631.

Meyer, F., Overbeek, R.A., and Rodriguez, A. (2009) FIGfams: yet another set of protein families. *Nucleic Acids Res* **37**: 6643–6654.

Meza, E., Becker, J., Bolivar, F., Gosset, G., and Wittmann, C. (2012) Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*. *Microb Cell Fact* **11**: 127.

Mikulík, K., Palecková, P., Felsberg, J., Bobek, J., Zídková, J., and Halada, P. (2008) SsrA genes of streptomycetes and association of proteins to the tmRNA during development and cellular differentiation. *Proteomics* **8**: 1429–1441.

Modak, H. V, and Kelly, D.J. (1995) Acetyl-CoA-dependent pyruvate carboxylase from the photosynthetic bacterium *Rhodobacter capsulatus*: rapid and efficient purification using dye-ligand affinity chromatography. *Microbiology* **141**: 2619–2628.

Moons, a, Valcke, R., and Montagu, M. Van (1998) Low-oxygen stress and water deficit induce cytosolic pyruvate orthophosphate dikinase (PPDK) expression in roots of rice, a C3 plant. *Plant J* **15**: 89–98.

Morgan, H.P., O'Reilly, F.J., Wear, M.A., O'Neill, J.R., Fothergill-Gilmore, L.A., Hupp, T., and Walkinshaw, M.D. (2013) M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. *Proc Natl Acad Sci U S A* **110**: 5881–5886.

Morgan, H.P., Zhong, W., McNae, I.W., Michels, P.A.M., Fothergill-Gilmore, L.A., and Walkinshaw, M.D. (2014) Structures of pyruvate kinases display evolutionarily divergent allosteric strategies. *R Soc Open Sci* **1**: 140120.

Morrison, T.B., Weis, J.J., and Wittwer, C.T. (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* **24**: 954–958.

Muirhead, H. (1990) Isoenzymes of pyruvate kinase. *Biochem Soc Trans* **18**: 193–196.

Muñoz, M.E., and Ponce, E. (2003) Pyruvate kinase: Current status of regulatory and functional properties. *Comp Biochem Physiol - B Biochem Mol Biol* **135**: 197–218.

Negm, F.B., Cornel, F.A., and Plaxton, W.C. (1995) Suborganellar Localization and Molecular Characterization of Nonproteolytic Degraded Leukoplast Pyruvate Kinase from Developing Castor Oil Seeds. *Plant Physiol* **109**: 1461–1469.

Nei, M., and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.

Nelson, D.A., and Cox, M.M. (2004) *Lehninger Principles of Biochemistry*. 4th ed., Freeman, W H, .

Neveling, U., Bringer-Meyer, S., and Sahm, H. (1998) Gene and subunit organization of bacterial pyruvate dehydrogenase complexes. *Biochim Biophys Acta* **1385**: 367–372.

Nieminen, L., Webb, S., Smith, M.C.M., and Hoskisson, P.A. (2013) A flexible mathematical model platform for studying branching networks: experimentally validated using the model actinomycete, *Streptomyces coelicolor*. *PLoS One* **8**: 54316.

Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, Ø.M., Sletta, H., Alam, M.T., Merlo, M.E., Moore, J., Omara, W. a M., Morrissey, E.R., Juarez-Hermosillo, M. a, Rodríguez-García, A., Nentwich, M., Thomas, L., Iqbal, M., Legaie, R., Gaze, W.H., Challis, G.L., Jansen, R.C., Dijkhuizen, L., Rand, D. a, Wild, D.L., Bonin, M., Reuther, J., Wohlleben, W., Smith, M.C.M., Burroughs, N.J., Martín, J.F., Hodgson, D. a, Takano, E., Breitling, R., Ellingsen, T.E., and Wellington, E.M.H. (2010) The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* **11**: 10–18.

Noda-García, L., and Barona-Gómez, F. (2013) Enzyme evolution beyond gene duplication: A model for incorporating horizontal gene transfer. *Mob Genet Elements* **3**: 26439.

Noda-García, L., Camacho-Zarco, A.R., Medina-Ruíz, S., Gaytán, P., Carrillo-Tripp, M., Fülöp, V., and Barona-Gómez, F. (2013) Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Mol Biol Evol* **30**: 2024–2034.

Nowak, T., and Suelter, C. (1981) Pyruvate kinase: activation by and catalytic role of the monovalent and divalent cations. *Mol Cell Biochem* **35**: 65–75.

O'Brien, T.A., Schrock, H.L., Russell, P., Blake, R., and Gennis, R.B. (1976) Preparation of *Escherichia coli* pyruvate oxidase utilizing a thiamine pyrophosphate affinity column. *Biochim Biophys Acta* **452**: 13–29.

O'Hagan, D. (1992) Biosynthesis of polyketide metabolites. *Nat Prod Rep* **9**: 447–479.

Obanye, A.I.C., Hobbs, G., Gardner, D.C.J., and Oliver, S.G. (1996) Correlation between carbon flux through the pentose phosphate pathway and production of the antibiotic methylenomycin in *Streptomyces coelicolor* A3 (2). *Microbiology* **142**: 133–137.

Okamoto, S., Taguchi, T., Ochi, K., and Ichinose, K. (2009) Biosynthesis of actinorhodin and related antibiotics: discovery of alternative routes for quinone formation encoded in the act gene cluster. *Chem Biol* **16**: 226–36.

Okanishi, M., Suzuki, K., and Umezawa, H. (1974) Formation and reversion of Streptomycete protoplasts: cultural condition and morphological study. *J Gen Microbiol* **80**: 389–400.

Olano, C., Lombó, F., Méndez, C., and Salas, J. a (2008) Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab Eng* **10**: 281–292.

Olukoshi, E.R., and Packter, N.M. (1994) Importance of stored triacylglycerols in *Streptomyces*: possible carbon source for antibiotics. *Microbiology* **140**: 931–943.

Oria-Hernández, J., Cabrera, N., Pérez-Montfort, R., and Ramírez-Sílva, L. (2005) Pyruvate kinase revisited: the activating effect of K+. *J Biol Chem* **280**: 37924–37939.

Oria-Hernández, J., Riveros-Rosas, H., and Ramírez-Sílva, L. (2006) Dichotomic phylogenetic tree of the pyruvate kinase family: K+ -dependent and -independent enzymes. *J Biol Chem* **281**: 30717–30724.

Overbeek, R.A., Begley, T., Butler, R.M., Choudhuri, J. V, Chuang, H.-Y., Cohoon, M., Crécy-Lagard, V. de, Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691–5702.

Overbeek, R.A., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., and Stevens, R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**: 206–214.

Owen, O.E., Kalhan, S.C., and Hanson, R.W. (2002) The key role of anaplerosis and cataplerosis for citric acid cycle function. *J Biol Chem* **277**: 30409–30412.

Parenti, F., and Coronelli, C. (1979) Members of the genus *Actinoplanes* and their antibiotics. *Annu Rev Microbiol* **33**: 389–411.

Pawluk, A., Scopes, R.K., and Griffiths-Smith, K. (1986) Isolation and properties of the glycolytic enzymes from *Zymomonas mobilis*. The five enzymes from glyceraldehyde-3-phosphate dehydrogenase through to pyruvate kinase. *Biochem J* **238**: 275–281.

Peters-Wendisch, P.G., Kreutzer, C., Kalinowski, J., Pátek, M., Sahm, H., and Eikmanns, B.J. (1998) Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the *pyc* gene. *Microbiology* **144**: 915–927.

Peters-Wendisch, P.G., Schiel, B., Wendisch, V.F., Katsoulidis, E., Möckel, B., Sahm, H., and Eikmanns, B.J. (2001) Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. *J Mol Microbiol Biotechnol* **3**: 295–300.

Plaxton, W.C. (1991) Leucoplast Pyruvate Kinase from Developing Castor Oil Seeds : Characterization of the Enzyme's Degradation by a Cysteine Endopeptidase. *Plant Physiol* **97**: 1334–1338.

Plaxton, W.C., Dennis, D.T., and Knowles, V.L. (1990) Purification of leucoplast pyruvate kinase from developing castor bean endosperm. *Plant Physiol* **94**: 1528–1534.

Plou, F.J., Ferrer, M., Nuero, O.M., Calvo, M. V., Alcalde, M., Reyes, F., and Ballesteros, A. (1998) Analysis of Tween 80 as an esterase/ lipase substrate for lipolytic activity assay. *Biotechnol Tech* **12**: 183–186.

Ponce, E., Flores, N., Martinez, A., Valle, F., and Bolívar, F. (1995) Cloning of the two pyruvate kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these enzymes in pyruvate biosynthesis. *J Bacteriol* **177**: 5719–5722.

Ponce, E., Martínez, A., Bolívar, F., and Valle, F. (1998) Stimulation of glucose catabolism through the pentose pathway by the absence of the two pyruvate kinase isoenzymes in *Escherichia coli*. *Biotechnol Bioeng* **58**: 292–295.

Pratt, J., Cooley, J.D., Purdy, C.W., and Straus, D.C. (2000) Lipase activity from strains of *Pasteurella multocida*. *Curr Microbiol* **40**: 306–309.

Projan, S.J. (2003) Why is big Pharma getting out of antibacterial drug discovery? *Curr Opin Microbiol* **6**: 427–430.

Raaijmakers, J.M., and Mazzola, M. (2012) Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. *Annu Rev Phytopathol* **50**: 403–424.

Ratcliff, W.C., and Denison, R.F. (2011) Microbiology. Alternative actions for antibiotics. *Science (80- )* **332**: 547–548.

Redenbach, M., Kieser, H.M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., and Hopwood, D.A. (1996) A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol Microbiol* **21**: 77–96.

Reeves, R.E. (1968) A new enzyme with the glycolytic function of pyruvate kinase. *J Biol Chem* **243**: 3202–3204.

Reeves, R.E. (1971) Pyruvate,phosphate dikinase from *Bacteroides symbiosus*. *Biochem J* **125**: 531–539.

Ren, Y., He, J., Feng, L., Liao, X., Jin, J., Li, Y., Cao, Y., Wan, J., and He, H. (2011) Structure-based rational design of novel hit compounds for pyruvate dehydrogenase multienzyme complex E1 components from *Escherichia coli*. *Bioorg Med Chem* **19**: 7501–7506.

Rico, S., Yepes, A., Rodríguez, H., Santamaría, J., Antoraz, S., Krause, E.M., Díaz, M., and Santamaría, R.I. (2014) Regulation of the AbrA1/A2 Two-Component System in *Streptomyces coelicolor* and the Potential of Its Deletion Strain as a Heterologous Host for Antibiotic Production. *PLoS One* **9**: e109844.

Ridley, C.P., Lee, H.Y., and Khosla, C. (2008) Evolution of polyketide synthases in bacteria. *Proc Natl Acad Sci U S A* **105**: 4595–4600.

Riedel, C., Rittmann, D., Dangel, P., Möckel, B., Petersen, S., Sahm, H., and Eikmanns, B.J. (2001) Characterization of the phosphoenolpyruvate carboxykinase gene from *Corynebacterium glutamicum* and significance of the enzyme for growth and amino acid production. *J Mol Microbiol Biotechnol* **3**: 573–583.

Rigali, S., Nothaft, H., Noens, E.E.E., Schlicht, M., Colson, S., Müller, M., Joris, B., Koerten, H.K., Hopwood, D. a, Titgemeyer, F., and Wezel, G.P. van (2006) The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol Microbiol* **61**: 1237–1251.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A.W., Hopwood, D.A., and Wezel, G.P. van (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep* **9**: 670–675.

Ringnér, M. (2008) What is principal component analysis? *Nat Biotechnol* **26**: 303–304.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Robinson, M.D., and Smyth, G.K. (2007) Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**: 2881–2887.

Robinson, M.D., and Smyth, G.K. (2008) Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**: 321–332.

Rocha, E.P.C., Smith, J.M., Hurst, L.D., Holden, M.T.G., Cooper, J.E., Smith, N.H., and Feil, E.J. (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* **239**: 226–235.

Rodríguez, E., Banchio, C., Diacovich, L., Bibb, M.J., and Gramajo, H. (2001) Role of an essential acyl coenzyme A carboxylase in the primary and secondary metabolism of *Streptomyces coelicolor* A3(2). *Appl Environ Microbiol* **67**: 4166–4176.

Rodríguez, E., Navone, L., Casati, P., and Gramajo, H. (2012) Impact of malic enzymes on antibiotic and triacylglycerol production in *Streptomyces coelicolor*. *Appl Environ Microbiol* **78**: 4571–4579.

Rokem, J.S., Lantz, A.E., and Nielsen, J. (2007) Systems biology of antibiotic production by microorganisms. *Nat Prod Rep* **24**: 1262–1287.

Romero, D., Traxler, M.F., López, D., and Kolter, R. (2011) Antibiotics as signal molecules. *Chem Rev* **111**: 5492–5505.

Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**: 539–542.

Rose, R.W., Bruser, T., Kissinger, J.C., and Pohlschroder, M. (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* **45**: 943–950.

Rossa, C.A., White, J., Kuiper, A., Postma, P.W., Bibb, M.J., and Gulik, V. (2002) Carbon Flux Distribution in Antibiotic-Producing Chemostat Cultures of. **150**: 138–150.

Ryding, N.J., Bibb, M.J., Molle, V., Findlay, K.C., Chater, K.F., and Buttner, M.J. (1999) New sporulation loci in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**: 5419–5425.

Ryu, Y., Butler, M.J., Chater, K.F., and Lee, K.J. (2006) Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in *Streptomyces coelicolor*. **72**: 7132–7139.

Saavedra-Lira, E., Ramirez-Silva, L., and Perez-Montfort, R. (1998) Expression and characterization of recombinant pyruvate phosphate dikinase from *Entamoeba histolytica*. *Biochim Biophys Acta* **1382**: 47–54.

Sakai, H., Suzuki, K., and Imahori, K. (1986) Purification and properties of pyruvate kinase from *Bacillus stearothermophilus*. *J Biochem* **99**: 1157–1167.

Sakai, Y., Hayatsu, M., and Hayano, K. (2002) Use of tween 20 as a substrate for assay of lipase activity in soils. *Soil Sci Plant Nutr* **48**: 729–734.

Sambrook, J., and Russell, D. (2000) *Molecular Cloning: A Laboratory Manual*. 3rd ed., Cold Spring Harbor Laboratory Press, .

Santos-Beneit, F., Rodríguez-García, A., and Martín, J.F. (2011) Complex transcriptional control of the antibiotic regulator afsS in *Streptomyces*: PhoP and AfsR are overlapping, competitive activators. *J Bacteriol* **193**: 2242–2251.

Sauer, U., and Eikmanns, B.J. (2005) The PEP-Pyruvate-Oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev* **29**: 765–794.

Schaftingen, E. van, Opperdoes, F.R., and Hers, H.G. (1985) Stimulation of *Trypanosoma brucei* pyruvate kinase by fructose 2,6-bisphosphate. *Eur J Biochem* **153**: 403–406.

Scheltema, R.A., Jankevics, A., Jansen, R.C., Swertz, M.A., and Breitling, R. (2011) PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal Chem* **83**: 2786–2793.

Schneider, D., Bruton, C.J., and Chater, K.F. (2000) Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2). *Mol Gen Genet* **263**: 543–553.

Schnellmann, J., Zeltins, A., Blaak, H., and Schrempf, H. (1994) The novel lectin-like protein CHB1 is encoded by a chitin-inducible *Streptomyces olivaceoviridis* gene and binds specifically to crystalline α-chitin of fungi and other organisms. *Mol Microbiol* **13**: 807–819.

Schramm, A., Siebers, B., Tjaden, B., Brinkmann, H., and Hensel, R. (2000) Pyruvate kinase of the hyperthermophilic crenarchaeote *Thermoproteus tenax*: Physiological role and phylogenetic aspects. *J Bacteriol* **182**: 2001–2009.

Schreiner, M.E., Fiur, D., Holátko, J., Pátek, M., and Eikmanns, B.J. (2005) E1 Enzyme of the Pyruvate Dehydrogenase Complex in *Corynebacterium glutamicum* : Molecular Analysis of the Gene and Phylogenetic Aspects. *J Bacteriol* **187**: 6005–6018.

Schulenburg, C., and Miller, B.G. (2014) Enzyme recruitment and its role in metabolic expansion. *Biochemistry* **53**: 836–845.

Sharples, G.P., and Williams, S.T. (1974) Fine structure of the globose bodies of *Dactylosporangium thailandense* (actinomycetales). *J Gen Microbiol* **84**: 219–222.

Shen, B., and Hutchinson, C.R. (1993) Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. *Science (80- )* **262**: 1535–1540.

Shen, X.-L., Dong, H.-J., Hou, X.-P., Guan, W.-J., and Li, Y.-Q. (2008) FtsY affects sporulation and antibiotic production by whiH in *Streptomyces coelicolor*. *Curr Microbiol* **56**: 61–65.

Shigekawa, K., and Dower, W.J. (1988) Electroporation of eukaryotes and prokaryotes: a general approach to the introduction of macromolecules into cells. *Biotechniques* **6**: 742–751.

Siebring, J. (2010) The Phosphofructokinases of *Streptomyces coelicolor*. .

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**: 779–787.

Smyth, G.K., and Verbyla, A.P. (1996) A Conditional Likelihood Approach to Residual Maximum Likelihood Estimation in Generalized Linear Models on JSTOR. *J R Stat Soc* **58**: 565–572.

Sola-Landa, A., Rodríguez-García, A., Apel, A.K., and Martín, J.F. (2008) Target genes and structure of the direct repeats in the DNA-binding sequences of the response regulator PhoP in *Streptomyces coelicolor*. *Nucleic Acids Res* **36**: 1358–1368.

Spencer, J.B., and Jordan, P.M. (1992) Purification and properties of 6-methylsalicylic acid synthase from *Penicillium patulum*. *Biochem J* **288**: 839–846.

Staunton, J. (1998) Combinatorial biosynthesis of erythromycin and complex polyketides. *Curr Opin Chem Biol* **2**: 339–345.

Staunton, J., and Weissman, K.J. (2001) Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**: 380–416.

Stirrett, K., Denoya, C., and Westpheling, J. (2009) Branched-chain amino acid catabolism provides precursors for the Type II polyketide antibiotic, actinorhodin, via pathways that are nutrient dependent. *J Ind Microbiol Biotechnol* **36**: 129–137.

Strauch, E., Takano, E., Baylis, H.A., and Bibb, M.J. (1991) The stringent response in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **5**: 289–298.

Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207–234.

Studier, F.W., and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**: 113–130.

Suarez, J.E., Barbes, C., and Hardisson, C. (1980) Germination of spores of *Micromonospora chalcea*: physiological and biochemical changes. *J Gen Microbiol* **121**: 159–167.

Suzuki, K., Ito, S., Shimizu-Ibuka, A., and Sakai, H. (2008) Crystal structure of pyruvate kinase from *Geobacillus stearothermophilus*. *J Biochem* **144**: 305–312.

Świątek, M. a, Tenconi, E., Rigali, S., and Wezel, G.P. van (2012) Functional analysis of the N-acetylglucosamine metabolic genes of *Streptomyces coelicolor* and role in control of development and antibiotic production. *J Bacteriol* **194**: 1136–1144.

Takano, H., Obitsu, S., Beppu, T., and Ueda, K. (2005) Light-Induced Carotenogenesis in *Streptomyces coelicolor* A3(2): Identification of an Extracytoplasmic Function Sigma Factor That Directs Photodependent Transcription of the Carotenoid Biosynthesis Gene Cluster. *J Bacteriol* **187**: 1825–1832.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725–2759.

Tarhan, L., Kayali, H.A., Sazak, A., and Sahin, N. (2011) The correlations between TCA-glyoxalate metabolite and antibiotic production of *Streptomyces* sp. M4018 grown in glycerol, glucose, and starch mediums. *Appl Biochem Biotechnol* **164**: 318–337.

Thomas, L., Hodgson, D.A., Wentzel, A., Nieselt, K., Ellingsen, T.E., Moore, J., Morrissey, E.R., Legaie, R., Consortium, T.S., Rodríguez-garcía, A., Martín, J.F., Burroughs, N.J., and Mh, E. (2012) Metabolic switches and adaptations deduced from the proteomes of *Streptomyces coelicolor* wild type and *phoP* mutant grown in batch culture. *Mol Cell Proteomics* **11**: M111.013797.

Thompson, C.J., Ward, J.M., and Hopwood, D.A. (1980) DNA cloning in *Streptomyces*: resistance genes from antibiotic-producing species. *Nature* **286**: 525–527.

Titgemeyer, F., Walkenhorst, J., Reizer, J., Stuiver, M.H., Cui, X., and Saier, M.H. (1995) Identification and characterization of phosphoenolpyruvate:fructose phosphotransferase systems in three *Streptomyces* species. *Microbiology* **141**: 51–58.

Tjaden, B., Plagens, A., Dörr, C., Siebers, B., and Hensel, R. (2006) Phosphoenolpyruvate synthetase and pyruvate, phosphate dikinase of *Thermoproteus tenax*: key pieces in the puzzle of archaeal carbohydrate metabolism. *Mol Microbiol* **60**: 287–298.

Tokovenko, B., Rebets, Y., Luzhetskyy, A., Metabolic, A., and Group, E. (2016) Automating Assessment of the Undiscovered Biosynthetic Potential of Actinobacteria. *BioRxiv* 1–27.

Traxler, M.F., and Kolter, R. (2015) Natural products in soil microbe interactions and evolution. *Nat Prod Rep* **32**: 956–970.

Traxler, M.F., Watrous, J.D., Alexandrov, T., Dorrestein, P.C., and Kolter, R. (2013) Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome. *MBio* **4**: e00459–00413.

Treangen, T.J., and Rocha, E.P.C. (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* **7**: e1001284.

Turło, J., Gajzlerska, W., Klimaszewska, M., Król, M., Dawidowski, M., and Gutkowska, B. (2012) Enhancement of tacrolimus productivity in *Streptomyces tsukubaensis* by the use of novel precursors for biosynthesis. *Enzyme Microb Technol* **51**: 388–395.

Uguru, G.C., Stephens, K.E., Stead, J.A., Towle, J.E., Baumberg, S., and McDowall, K.J. (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol Microbiol* **58**: 131–150.

Ullmann, A., Jacob, F., and Monod, J. (1967) Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. *J Mol Biol* **24**: 339–343.

Utter, M.F., and Keech, D.B. (1963) Pyruvate carboxylase I. Nature of the reaction. *J Biol Chem* **238**: 2603–2608.

Valentini, G., Chiarelli, L., Fortin, R., Speranza, M.L., Galizzi, A., and Mattevi, A. (2000) The allosteric regulation of pyruvate kinase. *J Biol Chem* **275**: 18145–18152.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., and Sinderen, D. van (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**: 495–548.

Vesselinova, N.I., and Ensign, J.C. (1996) Flagellar proteins of motile spores of Actinomycetes. *J Ind Microbiol* **16**: 377–382.

Viollier, P.H., Minas, W., Dale, G.E., Thompson, C.J., Folcher, M., and Thompson, C.J. (2001) Role of Acid Metabolism in *Streptomyces coelicolor* Morphological Differentiation and Antibiotic Biosynthesis. *J Bacteriol* **183**: 3184–3192.

Vitkup, D., Kharchenko, P., and Wagner, A. (2006) Influence of metabolic network structure and function on enzyme evolution. *Genome Biol* **7**: R39.

Vorísek, J., Powell, A.J., and Vaněk, Z. (1970) Regulation of biosynthesis of secondary metabolites. 13. Specific allosteric properties of phosphoenolpyruvate carboxylase in *Streptomyces aureofaciens*. *Folia Microbiol (Praha)* **15**: 153–159.

Vos, M., Quince, C., Pijl, A.S., Hollander, M. de, and Kowalchuk, G.A. (2012) A comparison of *rpoB* and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. *PLoS One* **7**: e30600.

Wagner, A. (2008) Gene duplications, robustness and evolutionary innovations. *BioEssays* **30**: 367–373.

Waksman, S.A. (1941) Antagonistic relations of microorganisms. *Bacteriol Rev* **5**: 231–291.

Walter, S., Wellmann, E., and Schrempf, H. (1998) The cell wall-anchored *Streptomyces reticuli* avicel-binding protein (AbpS) and its gene. *J Bacteriol* **180**: 1647–1654.

Wang, J., Wang, W., Wang, L., Zhang, G., Fan, K., Tan, H., and Yang, K. (2011) A novel role of "pseudo"γ-butyrolactone receptors in controlling γ-butyrolactone biosynthesis in *Streptomyces*. *Mol Microbiol* **82**: 236–250.

Warren, A.S., Aurrecoechea, C., Brunk, B., Desai, P., Emrich, S., Giraldo-Calderón, G.I., Harb, O., Hix, D., Lawson, D., Machi, D., Mao, C., McClelland, M., Nordberg, E., Shukla, M., Vosshall, L.B., Wattam, A.R., Will, R., Yoo, H.S., and Sobral, B. (2015) RNA-Rocket: an RNA-Seq analysis resource for infectious disease research. *Bioinformatics* **31**: 1496–1498.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Watve, M.G., Tickoo, R., Jog, M.M., and Bhole, B.D. (2001) How many antibiotics are produced by the genus *Streptomyces*? *Arch Microbiol* **176**: 386–390.

Weissman, K.J., and Leadlay, P.F. (2005) Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**: 925–936.

Wentzel, A., Bruheim, P., Øverby, A., Jakobsen, Ø.M., Sletta, H., Omara, W. a M., Hodgson, D. a, and Ellingsen, T.E. (2012) Optimized submerged batch fermentation strategy for systems scale studies of metabolic switching in *Streptomyces coelicolor* A3(2). *BMC Syst Biol* **6**: 59.

Wentzel, A., Sletta, H., Consortium, S., Ellingsen, T.E., and Bruheim, P. (2012) Intracellular Metabolite Pool Changes in Response to Nutrient Depletion Induced Metabolic Switching in *Streptomyces coelicolor*. *Metabolites* **2**: 178–194.

Wezel, G.P. van, Mahr, K., König, M., Traag, B.A., Pimentel-Schmitt, E.F., Willimek, A., and Titgemeyer, F. (2005) GlcP constitutes the major glucose uptake system of *Streptomyces coelicolor* A3(2). *Mol Microbiol* **55**: 624–636.

White, J., and Bibb, M.J. (1997) *bldA* dependence of undecylprodigiosin production in *Streptomyces coelicolor*A3(2) involves a pathway-specific regulatory cascade. *J Bacteriol* **179**: 627–633.

WHO (2001) WHO Global Strategy for Containment of Antimicrobial Resistance. .

Widdick, D.A., Dilks, K., Chandra, G., Bottrill, A., Naldrett, M., Pohlschröder, M., and Palmer, T. (2006) The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A* **103**: 17927–17932.

Wieschalka, S., Blombach, B., and Eikmanns, B.J. (2012) Engineering *Corynebacterium glutamicum* for the production of pyruvate. *Appl Microbiol Biotechnol* **94**: 449–459.

Winder, C.L., Dunn, W.B., Schuler, S., Broadhurst, D., Jarvis, R., Stephens, G.M., and Goodacre, R. (2008) Global metabolic profiling of *Escherichia coli* cultures: an evaluation of methods for quenching and extraction of intracellular metabolites. *Anal Chem* **80**: 2939–2948.

Wittwer, C.T., Ririe, K.M., Andrew, R. V, David, D.A., Gundry, R.A., and Balis, U.J. (1997) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* **22**: 176–181.

Wright, L.F., and Hopwood, D.A. (1976) Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **96**: 289–297.

Wu, C., Dunaway-Mariano, D., and Mariano, P.S. (2013) Design, synthesis, and evaluation of inhibitors of pyruvate phosphate dikinase. *J Org Chem* **78**: 1910–2192.

Wu, J., Zhang, Q., Deng, W., Qian, J., Zhang, S., and Liu, W. (2011) Toward improvement of erythromycin A production in an industrial *Saccharopolyspora erythraea* strain via facilitation of genetic manipulation with an artificial *attB* site for specific recombination. *Appl Environ Microbiol* **77**: 7508–7516.

Wu, Y.J. (2000) Highlights of semi-synthetic developments from erythromycin A. *Curr Pharm Des* **6**: 181–223.

Yang, Y.-H., Song, E., Lee, B.-R., Kim, E., Park, S.-H., Kim, Y.-G., Lee, C.-S., and Kim, B.-G. (2010) Rapid functional screening of *Streptomyces coelicolor* regulators by use of a pH indicator and application to the MarR-like regulator AbsC. *Appl Environ Microbiol* **76**: 3645–3656.

Yeo, M., and Chater, K.F. (2005) The interplay of glycogen metabolism and differentiation provides an insight into the developmental biology of *Streptomyces coelicolor*. *Microbiology* **151**: 855–861.

Yepes, A., Rico, S., Rodríguez-García, A., Santamaría, R.I., and Díaz, M. (2011) Novel two-component systems implied in antibiotic production in *Streptomyces coelicolor*. *PLoS One* **6**: e19980.

Yim, G., McClure, J., Surette, M.G., and Davies, J.E. (2010) Modulation of *Salmonella* gene expression by subinhibitory concentrations of quinolones. *J Antibiot (Tokyo)* **64**: 73–78.

Yim, G., Wang, H.H., and Davies, J. (2007) Antibiotics as signalling molecules. *Philos Trans R Soc London* **362**: 1195–1200.

Yokoyama, A., Murata, M., Oshima, Y., Iwashita, T., and Yasumoto, T. (1988) Some chemical properties of maitotoxin, a putative calcium channel agonist isolated from a marine dinoflagellate. *J Biochem* **104**: 184–187.

Yu, Z., Zhu, H., Zheng, G., Jiang, W., and Lu, Y. (2014) A genome-wide transcriptomic analysis reveals diverse roles of the two-component system DraR-K in the physiological and morphological differentiation of *Streptomyces coelicolor*. *Appl Microbiol Biotechnol* **98**: 9351–9363.

Zhang, R., Watson, D.G., Wang, L., Westrop, G.D., Coombs, G.H., and Zhang, T. (2014) Evaluation of mobile phase characteristics on three zwitterionic columns in hydrophilic interaction liquid chromatography mode for liquid chromatography-high resolution mass spectrometry based untargeted metabolite profiling of *Leishmania* parasites. *J Chromatogr* **1362**: 168–179.

Zoraghi, R., See, R.H., Gong, H., Lian, T., Swayze, R., Finlay, B.B., Brunham, R.C., McMaster, W.R., and Reiner, N.E. (2010) Functional analysis, overexpression, and kinetic characterization of pyruvate kinase from methicillin-resistant *Staphylococcus aureus*. *Biochemistry* **49**: 7733–7747.

Zoraghi, R., Worrall, L., See, R.H., Strangman, W., Popplewell, W.L., Gong, H., Samaai, T., Swayze, R.D., Kaur, S., Vuckovic, M., Finlay, B.B., Brunham, R.C., McMaster, W.R., Davies-Coleman, M.T., Strynadka, N.C., Andersen, R.J., and Reiner, N.E. (2011) Methicillin-resistant *Staphylococcus aureus* (MRSA) pyruvate kinase as a target for bis-indole alkaloids with antibacterial activities. *J Biol Chem* **286**: 44716–44725.

# 10 APPENDIX

## 10.1 SCRIPTS FOR DATABASE GENERATION AND DATA ANLYSIS

All scripts utilised were written in Perl.

### Download of Genomes From NCBI and Upload to RAST

This script can be found on github as "1.NCBI_and_RAST.pl" and "2.Batch_RetrieveFiles.pl" were both written by Nelly Selem (https://github.com/nselem/perlas/tree/master/TREE)

```
README FILE:

This scripts can
          1) Download genomes from NCBI using NCBI-id from the data bases Genome,
Nucleotide or GI
          2) Upload genomes to RAST using myRAST

################################################################################
###################
### Usage  ##############################################

$perl ControlRAST.pl <file>

The file must contain a list of
id        Organism Name

##############################################
## Requirments
##############################################

For download genomes from NCBI You dont need to install anything.
The scripts use the Eutils NCBI software, (no need to install).
You can look specifically at the function UploadGenomes

For upload genomes to RAST (Ubuntu)
If you want to use them you will need to
          1) Have a proper RAST account
          2) Install myRAST using linux installation (not Ubuntu because it has
library dependence)

The script of myRAST that is used to batch the genome upload is svr_submit_RAST_job:
$svr_submit_RAST_job -user user -passwd password -fasta <file> -domain Bacteria -
bioname <Genus Species> -genetic_code 11 -gene_caller rast

################################################################################
####################
```

---

1.NCBI_and_RAST.pl

```
#################################################################
############  Declare Functions  ##########################
#################################################################
use LWP::Simple;

my %ORGS;
sub ReadFile;
sub ReadID;
```

```
sub DownloadGenome;
sub UploadGenome;

#####################################################################
############                    Main    ##############################
%ORGS=ReadFile;
foreach my $ID (keys %ORGS){
        #print ("$id => $ORGS{$id}\n")
        $OrgName=$ORGS{$ID};
        DownloadGenome($ID, $OrgName);
        UploadGenome($ID, $OrgName);
}
#ReadID;

#####################################################################
#####################################################################
sub ReadFile{
        open FILE,  "$ARGV[0]" or die "I can not open the input FILE\n";
        my %orgs;
        while (my $line=<FILE>){
                chomp $line;
                #print "$line\n";
                my @content= split(/\t/,$line);
                #print"$content[0] => $content[1]\n";
                $orgs{$content[0]}=$content[1]." ".$content[0];
                #print"$content[0]=>$ORGS{$content[0]}\n\n";
                }
        return %orgs;
}
#####################################################################
#####################################################################
## 0 Reading the ID
sub ReadID{
        my $ID=shift;
        my $Flag="";

        #my $ID="ACGD01";
        #my $ORGNAME="Corynebacterium Accolens Atcc 49725 $ID";
        #my $ID="NZ_JODT01";
        #my $ORGNAME="Streptomyces Achromogenes Subsp. Achromogenes $ID";

        if ($ID=~/^[0-9]+$/){   #only match numbers then is Gi
                #print ("$ID is GI\n");
                $Flag="GI";
                }
        elsif($ID=~/_/){        #match _ then is nucleotide
                #print ("$ID is nucleotide ID\n");
                $Flag="NU";
                }
        elsif(substr($ID,0,4)=~/^[A-Za-z]{4}$/ and substr($ID,-2)=~/^[0-9]+$/){         # first 4 letters, last 2 numbers then is
genome
                print("$ID is genome ID\n");
                $Flag="GE";
        }
        else {
                print("Please Provide a valid ID. $ID is not valid\n");
                }
        return($Flag);
}

#####################################################################
##1      Download Genome from NCBI using Unique Genome identifier
sub DownloadGenome{
        my $ID=shift;
        my $Flag=ReadID($ID);
        print ("$ID,$Flag\n");
        my $ORGNAME=shift;
        my $file=$ID.".genome";
```

```
	##(Identifiers previously collected by lab members)
	if ($Flag eq "GE"){
		##1.1.1	If Genome Id (Use of wget)
		`wget http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=$ID.1.fsa_nt.gz`;
		##1.2 Decompress the file
		`gunzip index.html?download=$ID.1.fsa_nt.gz`;
		`mv index.html?download=$ID.1.fsa_nt $file`;
		}
	elsif($Flag eq "NU"){
		##1.1.2	If Nucleotide Id (Use of Entrez)
		$base = 'http://eutils.ncbi.nlm.nih.gov/entrez/eutils/';
		$url = $base . "efetch.fcgi?db=Nucleotide&id=$ID&rettype=fasta";
		$output = get($url);
		#$output = "hello";
		open(GENOME,'>',$file) or die "Could not open file $!";
		print GENOME ($output);
		close GENOME;
		}
	elsif($Flag eq "GI"){
		##1.1.2	If GI Id (Use of Entrez)
		}
}

##2	Upload Genomes to RAST using SVR svr_submit_RAST_job.pl and my account
sub UploadGenome{
	my $ID=shift;
	my $ORGNAME=shift;
	`svr_submit_RAST_job -user nselem35 -passwd q8Vf6ib -fasta $ID.genome -domain Bacteria -bioname
"$ORGNAME" -genetic_code 11 -gene_caller rast`;
	`rm $ID.genome`;
	## Remove Genome file from computer!!
}
```

---

## 2.Batch_RetrieveFiles.pl

```
#################################################################
############  Declare Functions  ############################
#################################################################
use LWP::Simple;

my %ORGS;
sub ReadFile;
sub RetrieveGenome;


#################################################################
############                Main    ##############################
%ORGS=ReadFile;
foreach my $ID (keys %ORGS){
	print"Retrieving RAST job $ID corresponding to the org:\n";
	#$OrgName=$ORGS{$ID};
	RetrieveGenome($ID);
}
#ReadID;

#################################################################
#################################################################

sub ReadFile{
	open FILE,  "$ARGV[0]" or die "I can not open the input FILE\n";
	my %orgs;
	while (my $line=<FILE>){
		chomp $line;
		print "$line\n";
		my @content= split(/\t/,$line);
		print "$content[0]=>$content[1]=>$content[2]\n";
```

```
                        #print"$content[0] => $content[1]\n";
                        $orgs{$content[0]}=$content[1];
                        #$orgs{$content[0]}=($content[1],$content[2]);
                        #print"$content[0]=>$ORGS{$content[0]}\n\n";
                        }
        return %orgs;
}

#################################################################
##2      Upload Genomes to RAST using SVR svr_submit_RAST_job.pl and my account
sub RetrieveGenome{
        my $ID=shift;
        `svr_retrieve_RAST_job nselem35 q8Vf6ib $ID table_txt > $ID.cvs`;
}
```

---

**Extraction of protein sequences**

The following script extracts all the protein sequences from each genome (.csv file) and combines them into one file creating a database. It can be run by the command extract_proteins.pl <inputFilewithID.txt>

Extract_proteins.pl

```
open FILE, "$ARGV[0]" or die "I can't open th input file\n";
while ($filename=<FILE>){
  $filename=~/(\d+)\s+(\d+\.\d+)\s+(.+)/;

  $file="$1.cvs";
  $specie="$3";


  open CVS, $file or die "I can't open the input file\n";
  while($line=<CVS>){
    if ($line=~/\.peg\./){
    my @st=split("\t",$line);
                        $seqaa=$st[12];
        $id=$st[1];
        $id=~s/fig\|6666666\.//;
        $id=~s/\.peg\./_/;
        $specie=~s/\s/_/g;
        $specie=~s/sp\.//g;

        print ">$id\_$specie\n$seqaa";
  }

  }
}
```

---

Creation database of protein sequence database and Tree construction

In order to obtain a tree for each of the genes of interest, the following commands need to be executed in Linux command line:

makeblastdb -in protein_DB.fasta -dbtype prot -parse_seqids # will make protein database to be used for the blast

blastp -query query.fasta -db protein_DB.fasta -evalue 0.001 -outfmt 6 -max_target_seqs 1000000|cut -f 2 > hits.txt

sort hits.txt | uniq > hitsNR.txt

blastdbcmd -dbtype prot -db protein_DB.fasta -entry_batch hitsNR.txt > hits.fasta

# then clean hits.fasta in Text Editor (replace lcl| and unnamed protein product with nothing)
muscle -in hits.fasta -out hits.aln #aligns the hits from the blast search

#check the alignment and delete if necessary respective sequences and rerun alignment

#For tree building:

#1.-Gblock to clean up the alignment
/opt/Gblocks_0.91b/Gblocks hitsHeader.aln -b4=5 -b5=H -b3=10>LOGhits.GBLOCks

# 2.-change format to stockholm
/opt/hmmer-3.1b1-linux-intel-x86_64/binaries/esl-reformat --informat afa stockholm hitsHeader.aln-gb >hitsHeader.stock

nohup /opt/quicktree_1.1/bin/quicktree -in a -out t -boot 10000 hitsHeader.stock > XYZ.tree &

#manual colouring of each tree by family in FigTree

---

## Expansion analysis

Perl scripts for the expansion analysis can be found under

https://github.com/nselem/perlas/tree/master/ORTHO

Download of files from RAST using F0_ExtractFiles.pl

```
use strict;
use warnings;
#####################################################
######### Description
#### nselem84@gmail.com
#####################################################
#  1. Creates a list that compares functions between GENOMES
## 2. The table contains the number of pegs that are doing the same function
#    Uses bindings files from the tar download from RAST

######### Subs and variables ######################
my @GENOMES=qx/ls *.bindings/; ## Read all cvs
my @ID=qx/ls *.txt/; ## Read all text files
my $N=scalar @GENOMES; ## Number of files
my %GENOME_NAMES; ## Hash Genome_Id -> Organisms Name
my %FUNC;
my %GENES;
```

```
my @functions;


sub readOrganisms;
sub CountFunctions;
sub headers;
sub Output;


#####################################################
##################################
###### Main program
#####################################################
%GENOME_NAMES=readOrganisms(@ID); ##Read txt files and fill Hash Genome_Id -> Organisms Name
%FUNC=CountFunctions($N,\%GENES,@GENOMES); ##Store in FUNC The number of fucntion and in GENES the pegs of the
functions
headers(\%GENOME_NAMES,@GENOMES);
Output($N,\%GENES,%FUNC);
#####################################################
#####################################################


sub readOrganisms{
        my @ID=shift;
        my %GENOME_NAMES;
        foreach my $ID(@ID){
                open(FILE,$ID) or die "Could not open file $ID $!";
                while ( my $line = <FILE> ) {
                        $line=~s/\n|\r//g;
                        my @st=split("\t",$line);
                        my $Id=$st[0];
                        my $name=$st[2];
                        $GENOME_NAMES{$Id}=$name;
                        #print "$Id => $GENOME_NAMES{$Id}\n";
                        }
                close FILE;
                }
        return (%GENOME_NAMES);
        }


sub CountFunctions{
        my $N=shift;
        my $refGENES=shift;
        my @GENOMES=@_;
        my $count=0;
        my %FUNC;

        foreach my $genome(@GENOMES){
                chomp $genome;
                print "genome: $genome count $count\n";
                open(FILE,$genome) or die "Could not open file $genome $!";
                my %FUNCTION; ##For each genome a Hash of array HAsh keys:functions Array Contents: gene with
that function

                while ( my $line = <FILE> ) {
                        my @st=split("\t",$line);
                        my $func=$st[1];
                        my $gen=$st[2];
                        chomp $gen;
                        #print "#$gen# GEn \n";

                        if(!exists $FUNCTION{$func}) {
                        #print "New Function $func !!! \n ";
                                $FUNCTION{$func}=[];
                                }

                        if (!($gen~~$FUNCTION{$func})){
```

```
                                        push (@{$FUNCTION{$func}}, $gen);
                                        my $size=scalar @{$FUNCTION{$func}};
                                        #print "Size $size \n";
                                        #print "$func => @{$FUNCTION{$func}}\n";
                                        }

                                }

                        close FILE;

                        foreach my $element(keys %FUNCTION){  ##Looking the function of the genome number $count
                                if(!exists $FUNC{$element}) {
                                        ## IF not exists FUNC function creates an array for all the genomes with zeros
                                        $FUNC{$element} = [];
                                        $refGENES->{$element}=[];
                                        for (my $i=0;$i<$N;$i++){
                                                $FUNC{$element}[$i]=0;
                                                $refGENES->{$element}[$i]="";
                                                }
                                        }
                                ##In the corresponding genome fill the number of gene with that function
                                $FUNC{$element}[$count]=scalar @{$FUNCTION{$element}};
                                $refGENES->{$element}[$count]=join(',', @{ $FUNCTION{$element} } );
                                }
                        $count++;

                        }

                return %FUNC;
                }


sub headers{
                my $refNAMES=shift;
                my @GENOMES=@_;

                open (FILE,">FunctionTable.cvs") or die "Could not open file Salida $!";
                open (PEGS,">PegsTable.cvs") or die "Could not open file Salida $!";
                my $headers="FUNCTION\t";
                foreach my $genome(@GENOMES){
                        chomp $genome;
                        $genome=~s/\.bindings//;
                        print "#$genome# => #$refNAMES->{$genome}#\n";
                        $headers=$headers.$refNAMES->{$genome}."\t";
                        }
                print FILE "$headers\n";
                print PEGS "$headers\n";
                close FILE;
                close PEGS;
                }


sub Output{
                my $N=shift;
                my $refGENES=shift;
                my %FUNC=(@_);

                open (FILE,">>FunctionTable.cvs") or die "Could not open file Salida $!";
                open (PEGS,">>PegsTable.cvs") or die "Could not open file Salida $!";

                foreach my $func(sort keys %FUNC){
                        #print "$func\n";
                        my $line = $func."\t";
                        my $peg= $func."\t";

                        for (my $i=0;$i<$N;$i++){
                                my $numberFunc=scalar $FUNC{$func}[$i];
                                $line=$line.$numberFunc."\t";
```

```
                        $peg=$peg.$refGENES->{$func}[$i]."\t";
                        }

                print FILE "$line\n";
                print PEGS "$peg\n";
                }
        close FILE;
        close PEGS;
        }
```

---

## Counting of number of genes annotated per functional roles in each genome

### F1_FunctionCountingBindigs.pl

```
use strict;
use warnings;
#######################################################
######### Description
#### nselem84@gmail.com
#######################################################
#  1. Creates a list that compares functions between GENOMES
## 2. The table contains the number of pegs that are doing the same function
#    Uses bindings files from the tar download from RAST

######### Subs and variables ######################
my @GENOMES=qx/ls *.bindings/; ## Read all cvs
my @ID=qx/ls *.txt/; ## Read all text files
my $N=scalar @GENOMES; ## Number of files
my %GENOME_NAMES; ## Hash Genome_Id -> Organisms Name
my %FUNC;
my %GENES;
my @functions;


sub readOrganisms;
sub CountFunctions;
sub headers;
sub Output;


#######################################################
###################################
###### Main program
#######################################################
%GENOME_NAMES=readOrganisms(@ID); ##Read txt files and fill Hash Genome_Id -> Organisms Name
%FUNC=CountFunctions($N,\%GENES,@GENOMES); ##Store in FUNC The number of fucntion and in GENES the pegs of the
functions
headers(\%GENOME_NAMES,@GENOMES);
Output($N,\%GENES,%FUNC);
#######################################################
#######################################################


sub readOrganisms{
        my @ID=shift;
        my %GENOME_NAMES;
        foreach my $ID(@ID){
                open(FILE,$ID) or die "Could not open file $ID $!";
                while ( my $line = <FILE> ) {
                        $line=~s/\n|\r//g;
                        my @st=split("\t",$line);
                        my $Id=$st[0];
                        my $name=$st[2];
                        $GENOME_NAMES{$Id}=$name;
                        #print "$Id => $GENOME_NAMES{$Id}\n";
                        }
```

```
		close FILE;
		}
	return (%GENOME_NAMES);
	}


sub CountFunctions{
	my $N=shift;
	my $refGENES=shift;
	my @GENOMES=@_;
	my $count=0;
	my %FUNC;

	foreach my $genome(@GENOMES){
		chomp $genome;
		print "genome: $genome count $count\n";
		open(FILE,$genome) or die "Could not open file $genome $!";
		my %FUNCTION; ##For each genome a Hash of array HAsh keys:functions Array Contents: gene with
that function

		while ( my $line = <FILE> ) {
			my @st=split("\t",$line);
			my $func=$st[1];
			my $gen=$st[2];
			chomp $gen;
			#print "#$gen# GEn \n";

			if(!exists $FUNCTION{$func}) {
			#print "New Function $func !!! \n ";
				$FUNCTION{$func}=[];
				}

			if (!($gen~~$FUNCTION{$func})){
				push (@{$FUNCTION{$func}}, $gen);
				my $size=scalar @{$FUNCTION{$func}};
				#print "Size $size \n";
				#print "$func => @{$FUNCTION{$func}}\n";
				}

			}

		close FILE;

		foreach my $element(keys %FUNCTION){  ##Looking the function of the genome number $count
			if(!exists $FUNC{$element}) {
				## IF not exists FUNC function creates an array for all the genomes with zeros
				$FUNC{$element} = [];
				$refGENES->{$element}=[];
				for (my $i=0;$i<$N;$i++){
					$FUNC{$element}[$i]=0;
					$refGENES->{$element}[$i]="";
					}
				}
			##In the corresponding genome fill the number of gene with that function
			$FUNC{$element}[$count]=scalar @{$FUNCTION{$element}};
			$refGENES->{$element}[$count]=join(',', @{ $FUNCTION{$element} } );
			}
		$count++;

		}

	return %FUNC;
	}


sub headers{
	my $refNAMES=shift;
	my @GENOMES=@_;
```

```
	open (FILE,">FunctionTable.cvs") or die "Could not open file Salida $!";
	open (PEGS,">PegsTable.cvs") or die "Could not open file Salida $!";
	my $headers="FUNCTION\t";
	foreach my $genome(@GENOMES){
		chomp $genome;
		$genome=~s/\.bindings//;
		print "#$genome# => #$refNAMES->{$genome}#\n";
		$headers=$headers.$refNAMES->{$genome}."\t";
		}
	print FILE "$headers\n";
	print PEGS "$headers\n";
	close FILE;
	close PEGS;
	}


sub Output{
	my $N=shift;
	my $refGENES=shift;
	my %FUNC=(@_);

	open (FILE,">>FunctionTable.cvs") or die "Could not open file Salida $!";
	open (PEGS,">>PegsTable.cvs") or die "Could not open file Salida $!";

	foreach my $func(sort keys %FUNC){
		#print "$func\n";
		my $line = $func."\t";
		my $peg= $func."\t";

		for (my $i=0;$i<$N;$i++){
			my $numberFunc=scalar $FUNC{$func}[$i];
			$line=$line.$numberFunc."\t";
			$peg=$peg.$refGENES->{$func}[$i]."\t";
			}

		print FILE "$line\n";
		print PEGS "$peg\n";
		}
	close FILE;
	close PEGS;
	}
```

## 10.2 RNA-SEQ DATA ANALYSIS SETTINGS

The raw data files were downloaded in BAM format from the Ion Torrent Server and were uploaded onto the RNA Rocket Server (http://RNA-Seq.pathogenportal.org/). Each file was analysed using the tools and settings as specified in Table 10-1.

Table 10-1 RNA data analysis settings using the RNA Rocket Galaxy pipeline

| **Tool** | **Settings** | **Purpose** |
|---|---|---|
| FastQ | default | determination of Q score |
| Trim (sickle) | single end<br>quality type: Sanger<br>Threshold Quality:20/Length:20 | removal of low quality sequence |
| Bowtie2 v2.0.2 | prokaryotic single-end analysis<br>Reference: *S. coelicolor* A3(2) | alignment |
| SAM Stat | default | quality report of alignment |
| cufflinks v2.0.2 | default<br>perform quartile normalization<br>assemble only transcripts matching<br>annotation | assemble transcripts |
| cuffdiff/cummRbund v0.0.7 | quartile normalization<br>disp. Estimation method: pooled<br>FDR: 0.05<br>use multiread correct<br>perform bias correction<br>include read group database<br>build cummRbund database | analyse differential expression |
| Boxplot<br>heatmap<br>PCA | default<br>default<br>default | graphical visualisation of results |

The fastq files of the raw data were imported into CLC Genomics workbench and analysed by using the tools available in the package as specified in Table 10-2.

Table 10-2 Tools and Settings used for RNA-Seq data analysis in CLC Genomics Workbench 7.5

| **Tool** | **Settings** | **Purpose** |
|---|---|---|
| Trim | limit: 0.02<br>ambigious trim = yes<br>ambigious limit = 2<br>quality trim = yes<br>quality limit = 0.02 | removal of low quality sequences |
| RNA Seq Analysis | Reference: *S. coelicolor* A3(2)<br>mismatches = 2<br>length fraction = 0.9<br>similarity fraction = 0.8<br>Expression levels: Genes<br>Expression values: total gene reads | alignment |
| Differential expression analysis (edgeR) | total count filter cutoff: 5.0<br>analysis: total gene reads<br>estimate tagwise dispersion = yes<br>comparison = all pairs<br>Bonferroni correction = yes<br>FDR correction = yes | differential gene expression analysis |

## 10.3 List of Genomes in Actinobacterial Database

This is an alphabetical list of all genomes and their identifier of the deposition in the whole genome sequence deposit of NCBI

http://www.ncbi.nlm.nih.gov/Traces/wgs/

***Actinoalloteichus cyanogriseus DSM 43889 AUBJ01***
***Actinoalloteichus cyanogriseus JOAL01***
***Actinoalloteichus cyanogriseus JODV01***

***Actinobaculum massiliae ACS-171-V-Col2 AGWL01***
***Actinobaculum schaalii DSM 15541 AUBK01***
***Actinobaculum schaalii FB123-CNA-2 AGWM01***
***Actinobaculum sp. oral taxon 183 str. F0552 AWSB01***
***Actinobaculum urinale DSM 15805 ATUY01***

***Actinocatenispora sera JOEG01***
***Actinokineospora enzanensis DSM 44649 ARFV01***
***Actinokineospora inagensis DSM 44258 AXWW01***
***Actinokineospora sp. EG49 AYXG01***
***Actinomadura atramentaria DSM 43919 ARMS01***
***Actinomadura flavalba DSM 45200 ARFO01***
***Actinomadura madurae LIID-AJ290 AWOO02***
***Actinomadura oligospora ATCC 43269 JADG01***
***Actinomadura rifamycini DSM 43936 AULB01***
***Actinomyces cardiffensis F0333 AQHZ01***
***Actinomyces coleocanis DSM 15436 ACFG01***
***Actinomyces europaeus ACS-120-V-Col10b AGWN01***
***Actinomyces georgiae DSM 6843 AUBM01***
***Actinomyces gerencseriae DSM 6844 AUBN01***
***Actinomyces graevenitzii C83 ACRN01***
***Actinomyces israelii DSM 43320 JONS01***
***Actinomyces massiliensis 4401292 AKIO01***
***Actinomyces neuii subsp. neuii DSM 8576 ATUW01***
***Actinomyces odontolyticus ATCC 17982 AAYI02***
***Actinomyces oris K20 BABV01***
***Actinomyces slackii ATCC 49928 AUAK01***
***Actinomyces sp. HPA0247 ATCA01***
***Actinomyces sp. ICM54 JDFI01***
***Actinomyces sp. MS2 CCXH01***
***Actinomyces sp. oral taxon 170 str. F0386 AFBL01***
***Actinomyces suimastitidis DSM 15538 AUBF01***
***Actinomyces timonensis DSM 23838 AKGF01***
***Actinomyces turicensis ACS-279-V-Col4 AGWQ01***
***Actinomyces vaccimaxillae DSM 15804 ATUX01***
***Actinomyces viscosus C505 ACRE02***
***Actinomycetospora chiangmaiensis DSM 45062 ARBI01***

***Actinoplanes globisporus DSM 43857 ARBJ01***
***Actinoplanes subtropicus JOJL01***
***Actinoplanes utahensis JRTT01***

***Actinopolymorpha alba DSM 45243 AREV01***
***Actinopolyspora erythraea JPMV01***
***Actinopolyspora halophila DSM 43834 AQUI01***
***Actinopolyspora iraqiensis IQ-H1 AICW01***
***Actinopolyspora mortivallis DSM 44261 AQZN01***

*Actinospica acidiphila JNYX01*
*Actinospica robiniae DSM 44927 AZAN01*
*Actinotalea fermentans ATCC 43279 = JCM 9966 = DSM 3133 AXCX01*
*Actinotalea ferrariae CF5-4 AXCW01*
*Aeromicrobium marinum DSM 15272 ACLF03*
*Aeromicrobium massiliense JC14 CAHG01*
*Agromyces italicus DSM 16388 ATXF01*
*Agromyces subbeticus DSM 16689 ATXG01*
*Allokutzneria albata JOEF01*

*Amycolatopsis alba DSM 44262 ARAF01*
*Amycolatopsis azurea DSM 43854 ANMG01*
*Amycolatopsis balhimycina FH 1894 ARBH01*
*Amycolatopsis benzoatilytica AK 16/65 ARPK01*
*Amycolatopsis decaplanina DSM 44594 AOHO01*
*Amycolatopsis halophila YIM 93223 AZAK01*
*Amycolatopsis jejuensis JNYZ01*
*Amycolatopsis lurida NRRL 2430 JFBM01*
*Amycolatopsis mediterranei JMQG01*
*Amycolatopsis methanolica 239 AQUL01*
*Amycolatopsis nigrescens CSC17Ta-90 ARVW01*
*Amycolatopsis orientalis DSM 40040 = KCTC 9412 ASJB01*
*Amycolatopsis orientalis DSM 43388 ASXG01*
*Amycolatopsis rifamycinica JMQI01*
*Amycolatopsis sp. ATCC 39116 AFWY03*
*Amycolatopsis sp. JGI 0001006-E10 AXXI01*
*Amycolatopsis sp. MJM2582 JPLW01*
*Amycolatopsis taiwanensis DSM 45107 JAFB01*
*Amycolatopsis thermoflava N1165 AXBH01*
*Amycolatopsis vancoresmycina JNYY01*

*Arthrobacter albus DNF00011 JRNH01*
*Arthrobacter castelli DSM 16402 AUMN01*
*Arthrobacter gangotriensis Lz1y AOCK01*
*Arthrobacter phenanthrenivorans JWTB01*
*Arthrobacter siccitolerans CAQI01*
*Arthrobacter sp. MA-N2 AQRI01*
*Arthrobacter sp. MWB30 JPZK01*
*Arthrobacter sp. PAO19 ATKN01*
*Arthrobacter sp. UNC362MFTsu5.1 JMLE01*
*Arthrobacter sp. W1 JWMD01*
*Bifidobacterium bifidum NCIMB 41171 AKCA01*
*Bifidobacterium breve HPH0326 ATCB01*
*Bifidobacterium dentium ATCC 27678 ABIX02*
*Blastococcus sp. URHD0036 JNIK01*
*Brevibacterium album DSM 18261 AUFJ01*
*Brevibacterium casei S18 AMSP01*
*Brevibacterium linens JTJZ01*
*Brevibacterium massiliense 5401308 CAJD01*
*Brevibacterium senegalense CAHK01*
*Brevibacterium sp. VCM10 JAJB01*
*Catenulispora acidiphila DSM 44928 NC_013131.1*
*Catenuloplanes japonicus JNXY01*
*Cellulomonas massiliensis JC225 CAHD01*
*Cellulomonas sp. HZM JEOE01*
*Cellulomonas sp. KRMCY2 JAGF01*
*Cellulomonas sp. URHD0024 AUEW01*

*Cellulomonas sp. URHE0023 JIAN01*
*Citricoccus sp. CH26A AFXQ01*
*Clavibacter cf. michiganensis LMG 26808 AZQZ01*
*Clavibacter michiganensis JROD01*
*Corynebacterium caspium DSM 44850 ARBM01*
*Corynebacterium diphtheriae str. Aberdeen AUZO01*
*Corynebacterium glutamicum ATCC 14067 AGQQ02*
*Corynebacterium jeikeium JFCR01*
*Corynebacterium propinquum DSM 44285 AQXC01*
*Corynebacterium pseudotuberculosis JPJB01*
*Corynebacterium ulcerans NCTC 12077 AYUJ01*
*Dactylosporangium aurantiacum JNYJ01*
*Dermabacter hominis 1368 JDRS01*
*Dermabacter sp. HFH0086 ATFO01*
*Dietzia alimentaria 72 AGFF01*
*Dietzia sp. UCD-THP AOSR01*
*Frankia sp. Allo2 JPHT01*
*Frankia sp. BCU110501 ARDT01*
*Frankia sp. BMG5.12 ARFH01*
*Frankia sp. BMG5.23 JDWE01*
*Frankia sp. CcI6 AYTZ01*
*Frankia sp. CeD JPGU01*
*Frankia sp. CN3 AGJN02*
*Frankia sp. Iso899 ATXA01*
*Frankia sp. Thr JENI01*
*Glycomyces arizonensis DSM 44726 AXWO01*
*Glycomyces sp. NRRL B-16210 JOGR01*
*Glycomyces tenuis DSM 44171 ATYW01*
*Gordonia Kroppenstedtii Dsm 45133 AQYG01*
*Gordonia Neofelifaecis Nrrl B-59395 AEUD01*
*Gordonia Shandongensis Dsm 45094 AUHE01*
*Hamadaea tsunoensis DSM 44101 AUAX01*
*Herbidospora cretacea JODQ01*
*Intrasporangium chromatireducens Q5-1 AWQS01*
*Intrasporangium oryzae NRRL B-24470 AWSA01*
*Kibdelosporangium sp. MJ126-NF4 CDME01*
*Kineococcus radiotolerans SRS30216 = ATCC BAA-149 NC_009664.2*
*Kitasatospora arboriphila JAIY01*
*Kitasatospora azatica KCTC 9699 JNBY01*
*Kitasatospora cheerisanensis KCTC 2395 JNWZ01*
*Kitasatospora sp. MBT63 JNYV01*
*Kitasatospora sp. NRRL B-11411 JQMO01*
*Knoellia aerolata DSM 18566 AVPL01*
*Knoellia flava TL1 AVPI01*
*Knoellia sinensis KCTC 19936 AVPJ01*
*Knoellia subterranea KCTC 19937 AVPK01*
*Kocuria marina JROM01*
*Kocuria palustris PEL ANHZ02*
*Kocuria polaris JSUH01*
*Kocuria rhizophila JWTC01*
*Kocuria sp. UCD-OTCP AOSQ01*
*Kocuria sp. ZOR0020 JROX01*
*Kribbella catacumbae DSM 19601 AQUZ01*
*Kutzneria albida JNYH01*
*Kutzneria sp. 744 ACQP01*

*Lechevalieria aerocolonigenes JOFI01*
*Leifsonia aquatica H1aii AYMR01*
*Leifsonia rubra CMS 76R ATIA01*
*Leifsonia sp. 109 AQYK01*
*Lentzea albidocapillata JOEA01*
*Leucobacter chironomi DSM 19883 ATXU01*
*Leucobacter salsicius M1-8 AOCN01*
*Leucobacter sp. PH1c AYMV01*
*Leucobacter sp. UCD-THU APJM01*
*Longispora albida DSM 44784 ARBS01*
*Marmoricola aequoreus JOJN01*
*Marmoricola sp. URHB0036 AUEP01*
*Microbacterium barkeri 2011-R4 AKVP01*
*Microbacterium gubbeenense DSM 15944 AUGQ01*
*Microbacterium hominis JWSZ01*
*Microbacterium luticocti DSM 19459 AULS01*
*Microbacterium mangrovi JTDK01*
*Microbacterium oleivorans JFYO01*
*Microbacterium paraoxydans 77MFTsu3.2 AQYI01*
*Microbacterium profundi JPSY01*
*Microbacterium sp. 11MF ARTJ01*
*Microbacterium sp. CF335 JUGQ01*
*Microbacterium sp. CH12i JHET01*
*Microbacterium sp. Cr-K1W JARF01*
*Microbacterium sp. Cr-K20 JARE01*
*Microbacterium sp. Cr-K29 JARC01*
*Microbacterium sp. G3 CDAR01*
*Microbacterium sp. KROCY2 JAGG01*
*Microbacterium sp. TS-1 BASQ01*
*Microbacterium sp. UCD-TDU AOSO01*
*Microbispora rosea subsp. nonnitritogenes JNZQ01*
*Microbispora sp. ATCC PTA-5024 AWEV01*
*Microbispora sp. NBRC 110460 BBOZ01*
*Microbispora sp. NRRL B-24597 JOAF01*
*Micromonospora chokoriensis JOAN01*
*Micromonospora globosa JNZR01*
*Micromonospora lupini str. Lupac 08 CAIE01*
*Micromonospora parva JNZS01*
*Micromonospora purpureochromogenes JOEU01*
*Micromonospora sp. ATCC 39149 ACES01*
*Micromonospora sp. CNB394 ARGW01*
*Microtetraspora glauca JOFO01*
*Modestobacter sp. KNN45-2b JPMX01*
*Mycetocola saprophilus JOEC01*
*Mycobacterium bovis BCG str. ATCC 35733 AEZF01*
*Mycobacterium smegmatis MC2 51 JAJD01*
*Mycobacterium smegmatis MKD8 AOCJ01*
*Mycobacterium tuberculosis KZN 4207 ACVS02*

*Nocardia abscessus NBRC 100374 BAFP01*
*Nocardia aobensis NBRC 100429 BAFQ01*
*Nocardia asteroides NBRC 15531 BAFO02*
*Nocardia brasiliensis IFM 10847 BAUA01*
*Nocardia brasiliensis NBRC 14402 BAFT02*
*Nocardia brevicatena NBRC 12119 BAFU01*
*Nocardia carnea NBRC 14403 BAFV01*

*Nocardia concava NBRC 100430 BAFX01*
*Nocardia exalbida NBRC 100660 BAFZ01*
*Nocardia higoensis NBRC 100133 BAGA01*
*Nocardia jiangxiensis NBRC 101359 BAGB01*
*Nocardia niigatensis NBRC 100131 BAGC01*
*Nocardia otitidiscaviarum IFM 11049 BATZ01*
*Nocardia otitidiscaviarum NBRC 14405 BAGD01*
*Nocardia paucivorans NBRC 100373 BAGE01*
*Nocardia pneumoniae NBRC 100136 BAGF01*
*Nocardia rhamnosiphila JOAJ01*
*Nocardia rhamnosiphila JOEJ01*
*Nocardia sp. 348MFTsu5.1 ARTM01*
*Nocardia sp. BMG51109 JAFQ01*
*Nocardia sp. CNY236 AXVD01*
*Nocardia sp. NRRL WC-3656 JOJF01*
*Nocardia sp. W9851 JNFP01*
*Nocardia takedensis NBRC 100417 BAGG01*
*Nocardia testacea NBRC 100365 BAGJ01*
*Nocardia transvalensis NBRC 15921 BAGL01*
*Nocardia veterana NBRC 100344 BAGM01*
*Nocardia vinacea NBRC 16497 BAGN01*

*Nocardioides alkalitolerans DSM 16699 AUFN01*
*Nocardioides halotolerans DSM 19273 AUGT01*
*Nocardioides sp. CF8 ASEP01*
*Nocardioides sp. Iso805N AQOZ01*
*Nocardioides sp. UNC345MFTsu5.1 JQJL01*
*Nocardioides sp. URHA0020 JIAR01*
*Nocardioides sp. URHA0032 JIAV01*
*Nocardiopsis alba DSM 43377 ANAC01*
*Nocardiopsis alkaliphila YIM 80379 ANBD01*
*Nocardiopsis baichengensis YIM 90130 ANAS01*
*Nocardiopsis baichengensis YIM 90130 ANAS01*
*Nocardiopsis ganjiahuensis DSM 45031 ANBA01*
*Nocardiopsis prasina DSM 43845 ANAE01*
*Nocardiopsis sp. CNS639 ARMI01*
*Nocardiopsis sp. CNT312 AZXF01*
*Nocardiopsis xinjiangensis YIM 90004 ANBE01*
*Pilimelia anulata JOFP01*
*Pimelobacter simplex strain VKM Ac-2033D NZ_CP009896.1*
*Prauserella rugosa JOIJ01*
*Prauserella sp. Am3 JTJI01*
*Propionibacterium acidifaciens AUFR01*
*Propionibacterium acidipropionici ATYU01*
*Propionibacterium acnes NC_006085.1*
*Propionibacterium avidum AGBA01*
*Propionibacterium granulosum AOST01*
*Propionibacterium humerusii AFAM01*
*Propionibacterium jensenii AUDD01*
*Propionibacterium propionicum NC_018142.1*
*Propionibacterium thoenii AUHZ01*
*Propionicicella superfundia DSM 22317 AUIA01*

*Pseudonocardia acaciae DSM 45401 JIAI01*
*Pseudonocardia asaccharolytica DSM 44247 = NBRC 16224 AUII01*
*Pseudonocardia autotrophica JNYD01*
*Pseudonocardia sp. P2 AEGE01*

*Pseudonocardia spinosispora DSM 44797 AUBB01*

*Rhodococcus defluvii JPOC01*
*Rhodococcus erythropolis R138 ASKF01*
*Rhodococcus erythropolis SK121 ACNO01*
*Rhodococcus erythropolis XP AGCF01*
*Rhodococcus fascians A44A JMEX01*
*Rhodococcus imtechensis RKJ300 = JCM 13270 AJJH01*
*Rhodococcus opacus JOIM01*
*Rhodococcus pyridinivorans AK37 AHBW01*
*Rhodococcus rhodnii JOAA01*
*Rhodococcus rhodnii LMG 5362 APMY01*
*Rhodococcus rhodochrous ATCC 21198 AZHI01*
*Rhodococcus rhodochrous JNWS01*
*Rhodococcus sp. 114MFTsu3.1 ARTN01*
*Rhodococcus triatomae BKS 15-14 AODO01*
*Rhodococcus wratislaviensis NBRC 100605 BAWF01*

*Rothia aeria F0184 AXZG01*
*Rothia dentocariosa JPVS01*
*Rothia dentocariosa M567 ADDW01*
*Rothia mucilaginosa ATCC 25296 ACVO01*
*Saccharomonospora azurea NA-128 AGIU02*
*Saccharomonospora cyanea NA-134 AHLY01*
*Saccharomonospora halophila 8 AICX01*
*Saccharomonospora marina XMU15 AHLX01*
*Saccharomonospora sp. CNQ490 AZUM01*
*Saccharomonospora viridis JRZE01*
*Saccharomonospora xinjiangensis XJ-54 AICV01*
*Saccharopolyspora erythraea D AVCN01*
*Saccharopolyspora erythraea NRRL 2338 ABFV01*
*Saccharopolyspora rectivirgula JNVU01*
*Saccharopolyspora spinosa NRRL 18395 AEYC01*
*Saccharothrix sp. NRRL B-16314 JNXC01*
*Saccharothrix syringae JNYO01*
*Salinispora arenicola DSM 45545 ARBB01*
*Salinispora pacifica DSM 45543 AQZB01*
*Salinispora tropica CNS416 ARHQ01*
*Sciscionella marina DSM 45152 ARAX01*
*Sciscionella sp. SE31 JALM01*
*Segniliparus rugosus ATCC BAA-974 ACZI02*
*Serinicoccus marinus DSM 15273 ATWM01*
*Serinicoccus profundi MCCC 1A05965 AFYF01*

*Streptacidiphilus albus JL83 JQML01*
*Streptacidiphilus anmyonensis BBPQ01*
*Streptacidiphilus carbonis BBPM01*
*Streptacidiphilus jeojiense JOEH01*
*Streptacidiphilus jiangxiensis BBPN01*
*Streptacidiphilus melanogenes BBPP01*
*Streptacidiphilus neutrinimicus BBPO01*
*Streptacidiphilus oryzae TH49 JQMQ01*
*Streptacidiphilus rugosus AM-16 JQMJ01*
*Streptomonospora alba JROO01*
*Streptomyces achromogenes subsp. achromogenes JODT01*
*Streptomyces acidiscabies 84-104 AHBF01*
*Streptomyces Afghaniensis 772 AOPY01*
*Streptomyces albidoflavus JOII01*

***Streptomyces alboflavus JNXT01***
***Streptomyces alboviridis JNWU01***
***Streptomyces albulus CCRC 11814 AROY02***
***Streptomyces albus J1074 ABYC01***
***Streptomyces anulatus JOEZ01***
***Streptomyces atratus JQJU01***
***Streptomyces atroolivaceus JNXG01***
***Streptomyces aurantiacus JA 4570 AOPZ01***
***Streptomyces auratus AGR0001 AJGV01***
***Streptomyces aureocirculatus JOAR01***
***Streptomyces aureofaciens JNWR01***
***Streptomyces aureofaciens JODU01***
***Streptomyces aureofaciens JOER01***
***Streptomyces aureus JNZD01***
***Streptomyces avellaneus JOFK01***
***Streptomyces avermitilis MA-4680 = NBRC 14893 BAVY01***
***Streptomyces avicenniae JOEK01***
***Streptomyces baarnensis JNZV01***
***Streptomyces bicolor JOFS01***
***Streptomyces bikiniensis JNWL01***
***Streptomyces bikiniensis JOAU01***
***Streptomyces bottropensis ATCC 25435 ARTP01***
***Streptomyces brasiliensis JNXB01***
***Streptomyces californicus JOFG01***
***Streptomyces canus 299MFChir4.1 ARTQ01***
***Streptomyces capuensis JNWP01***
***Streptomyces carneus JNZF01***
***Streptomyces catenulae JODY01***
***Streptomyces chartreusis NRRL 12338 AGDE01***
***Streptomyces clavuligerus ATCC 27064 ADGD01***
***Streptomyces coelicoflavus ZG0656 AHGS01***
***Streptomyces coelicolor A3(2) NC_003888.3***
***Streptomyces cyaneofuscatus JOEM01***
***Streptomyces durhamensis JNXR01***
***Streptomyces erythrochromogenes JNZG01***
***Streptomyces erythrochromogenes JNZG01***
***Streptomyces exfoliatus JNZP01***
***Streptomyces flaveolus JNWV01***
***Streptomyces Flaveus JOCU01***
***Streptomyces flavidovirens DSM 40150 AUBE01***
***Streptomyces flavochromogenes JNZO01***
***Streptomyces Flavotricini JNXV01***
***Streptomyces flavovirens JOAB01***
***Streptomyces floridae JOAC01***
***Streptomyces fradiae JNAD01***
***Streptomyces fulvoviolaceus JOEY01***
***Streptomyces fulvoviridis JNXH01***
***Streptomyces galbus JRHJ01***
***Streptomyces gancidicus BKS 13-15 AOHP01***
***Streptomyces ghanaensis ATCC 14672 ABYA01***
***Streptomyces Globisporus C-1027 AJUO01***
***Streptomyces globisporus subsp. globisporus JNZK01***
***Streptomyces griseoaurantiacus M045 AEYX01***
***Streptomyces griseoflavus Tu4000 ACFA01***
***Streptomyces griseofuscus JOFU01***

*Streptomyces griseolus JOFC01*
*Streptomyces griseoluteus JOBE01*
*Streptomyces griseorubens JJMG01*
*Streptomyces griseus subsp. griseus JOGA01*
*Streptomyces Griseus Subsp. Griseus JOIT01*
*Streptomyces griseus subsp. rhodochrous JOFE01*
*Streptomyces griseus XylebKG-1 ADFC02*
*Streptomyces halstedii JOAZ01*
*Streptomyces hygroscopicus subsp jinggangensis 5008*
*Streptomyces iakyrus JNXI01*
*Streptomyces Ipomoeae 91-03 AEJC01*
*Streptomyces katrae JNZY01*
*Streptomyces lavendulae subsp. lavendulae JOBU01*
*Streptomyces leeuwenhoekii AZSD01*
*Streptomyces lilacinus JNXU01*
*Streptomyces lividans 1326 APVM01*
*Streptomyces lydicus JNZA01*
*Streptomyces mediolani JOJK01*
*Streptomyces megasporus JODL01*
*Streptomyces mirabilis JQLE01*
*Streptomyces mobaraensis NBRC 13819 = DSM 40847 AORZ01*
*Streptomyces monomycini JNYL01*
*Streptomyces mutabilis JNFQ01*
*Streptomyces natalensis JNYF01*
*Streptomyces niger JOFQ01*
*Streptomyces niveus NCIMB 11891 AWQW01*
*Streptomyces novaecaesareae JNWQ01*
*Streptomyces ochraceiscleroticus JOAX01*
*Streptomyces olindensis JJOH01*
*Streptomyces olivaceus JNWM01*
*Streptomyces Olivaceus JOFH01*
*Streptomyces peucetius JOCK01*
*Streptomyces pluripotens JTDH01*
*Streptomyces puniceus JOFA01*
*Streptomyces purpeochromogenes JOBG01*
*Streptomyces purpeofuscus JODS01*
*Streptomyces purpureus KA281 ARAD01*
*Streptomyces Resistomycificus JOBA01*
*Streptomyces rimosus R6-500 JJNO01*
*Streptomyces roseochromogenes subsp. oscitans DS 12.976 AWQX01*
*Streptomyces Roseosporus Nrrl 11379 ABYX02*
*Streptomyces roseoverticillatus JOFL01*
*Streptomyces ruber JOAQ01*
*Streptomyces scabrisporus DSM 41855 ARCJ01*
*Streptomyces sclerotialus JODX01*
*Streptomyces scopuliridis RB72 JOEI01*
*Streptomyces seoulensis JNXP01*
*Streptomyces somaliensis DSM 40738 AJJM01*
*Streptomyces sp. 142MFCol3.1 AUKV01*
*Streptomyces sp. 150FB JTHL01*
*Streptomyces sp. 303MFCol5.2 ARTR01*
*Streptomyces sp. 351MFTsu5.1 ARTS01*
*Streptomyces sp. AA0539 ALNP01*
*Streptomyces sp. AA1529 ALAP01*
*Streptomyces sp. AcH 505 JTIY01*

*Streptomyces sp. Amel2xE9 ARPE01*
*Streptomyces sp. ATexAB-D23 AREH01*
*Streptomyces Sp. Aw19m42 CBRG01*
*Streptomyces sp. CNB091 ARJI01*
*Streptomyces sp. CNH099 AZWL01*
*Streptomyces sp. CNH189 AZUS01*
*Streptomyces sp. CNH287 AUKQ01*
*Streptomyces sp. CNQ329 AXVU01*
*Streptomyces sp. CNQ431 JTCK01*
*Streptomyces sp. CNQ-525 JNID01*
*Streptomyces sp. CNQ865 AUKP01*
*Streptomyces sp. CNR698 AZXC01*
*Streptomyces sp. CNS335 ARHS01*
*Streptomyces sp. CNS606 AUFD01*
*Streptomyces sp. CNS615 AQPE01*
*Streptomyces sp. CNS654 JNLT01*
*Streptomyces sp. CNT302 ARIM01*
*Streptomyces sp. CNT318 AUKN01*
*Streptomyces sp. CNT360 AUKO01*
*Streptomyces sp. CNT371 AZWZ01*
*Streptomyces sp. CNT372 ARHT01*
*Streptomyces sp. CNY228 ARIN01*
*Streptomyces sp. CNY243 ARHU01*
*Streptomyces sp. CT34 JSFP01*
*Streptomyces sp. DvalAA-83 ARPI01*
*Streptomyces sp. e14 ACUR01*
*Streptomyces sp. FxanaD5 ARDY01*
*Streptomyces sp. FXJ7.023 APIV01*
*Streptomyces sp. GBA 94-10 ASHF01*
*Streptomyces sp. GXT6 AYRX01*
*Streptomyces sp. HCCB10043 AWOQ01*
*Streptomyces sp. HGB0020 AGER01*
*Streptomyces sp. HmicA12 ARED01*
*Streptomyces sp. HPH0547 ATCE01*
*Streptomyces sp. JS01 JPWW01*
*Streptomyces sp. KhCrAH-244 AREA01*
*Streptomyces sp. KhCrAH-337 AZVF01*
*Streptomyces sp. KhCrAH-340 ARDX01*
*Streptomyces sp. KhCrAH-40 AZVG01*
*Streptomyces sp. LaPpAH-108 AREG01*
*Streptomyces sp. LaPpAH-165 AREB01*
*Streptomyces sp. LaPpAH-202 ARDM01*
*Streptomyces sp. LaPpAH-95 AQWQ01*
*Streptomyces sp. M10 AMZL01*
*Streptomyces sp. Mg1 ATCJ01*
*Streptomyces sp. MspMP-M5 AREE01*
*Streptomyces sp. MUSC 125 JUIG01*
*Streptomyces sp. NBRC 110027 BBNO01*
*Streptomyces sp. NBRC 110035 BBNN01*
*Streptomyces sp. NRRL B-11253 JNWN01*
*Streptomyces sp. NRRL B-12105 JNZW01*
*Streptomyces sp. NRRL B-1322 JOHF01*
*Streptomyces sp. NRRL B-1347 JOJM01*
*Streptomyces sp. NRRL B-1381 JOHG01*
*Streptomyces sp. NRRL B-24051 JOAE01*

***Streptomyces sp. NRRL B-24484 JOAI01***
***Streptomyces sp. NRRL B-24720 JOJO01***
***Streptomyces sp. NRRL B-2790 JOHI01***
***Streptomyces sp. NRRL B-3229 JOGP01***
***Streptomyces sp. NRRL B-3253 JOGQ01***
***Streptomyces sp. NRRL B-5680 JOGG01***
***Streptomyces sp. NRRL F-2202 JOIH01***
***Streptomyces sp. NRRL F-2305 JOFW01***
***Streptomyces sp. NRRL F-2580 JOIR01***
***Streptomyces sp. NRRL F-2664 JOFX01***
***Streptomyces sp. NRRL F-2747 JOIS01***
***Streptomyces sp. NRRL F-2799 JOIF01***
***Streptomyces sp. NRRL F-2890 JOIG01***
***Streptomyces sp. NRRL F-3213 JOIQ01***
***Streptomyces sp. NRRL F-3218 JOIP01***
***Streptomyces sp. NRRL F-3273 JOIO01***
***Streptomyces sp. NRRL F-3307 JOIC01***
***Streptomyces sp. NRRL F-4335 JOFY01***
***Streptomyces sp. NRRL F-4474 JOIB01***
***Streptomyces sp. NRRL F-4835 JOIE01***
***Streptomyces sp. NRRL F-5008 JOHW01***
***Streptomyces sp. NRRL F-5053 JOHT01***
***Streptomyces sp. NRRL F-5065 JOHV01***
***Streptomyces sp. NRRL F-5123 JOHY01***
***Streptomyces sp. NRRL F-5126 JOFZ01***
***Streptomyces sp. NRRL F-5135 JOHR01***
***Streptomyces sp. NRRL F-5193 JOHZ01***
***Streptomyces sp. NRRL F-525 JNXE01***
***Streptomyces sp. NRRL F-5527 JOHL01***
***Streptomyces sp. NRRL F-5555 JOHJ01***
***Streptomyces sp. NRRL F-5630 JOGO01***
***Streptomyces sp. NRRL F-5635 JOGJ01***
***Streptomyces sp. NRRL F-5638 JOGL01***
***Streptomyces sp. NRRL F-5639 JOGK01***
***Streptomyces sp. NRRL F-5650 JOGV01***
***Streptomyces sp. NRRL F-5681 JOHA01***
***Streptomyces sp. NRRL F-5702 JOHD01***
***Streptomyces sp. NRRL F-5727 JOGX01***
***Streptomyces sp. NRRL F-5917 JOHQ01***
***Streptomyces sp. NRRL F-6131 JOHN01***
***Streptomyces sp. NRRL F-6628 JOHX01***
***Streptomyces sp. NRRL F-6674 JNXF01***
***Streptomyces sp. NRRL F-6676 JOHS01***
***Streptomyces sp. NRRL F-6677 JOHP01***
***Streptomyces sp. NRRL S-1022 JOIU01***
***Streptomyces sp. NRRL S-118 JOBL01***
***Streptomyces sp. NRRL S-1314 JOHU01***
***Streptomyces sp. NRRL S-1448 JOGE01***
***Streptomyces sp. NRRL S-146 JOAW01***
***Streptomyces sp. NRRL S-15 JOGS01***
***Streptomyces sp. NRRL S-1777 JOGY01***
***Streptomyces sp. NRRL S-1813 JOHB01***
***Streptomyces sp. NRRL S-1824 JOGZ01***
***Streptomyces sp. NRRL S-1831 JOGT01***
***Streptomyces sp. NRRL S-1833 JOHC01***

*Streptomyces sp. NRRL S-1868 JOGD01*
*Streptomyces sp. NRRL S-1896 JOIV01*
*Streptomyces sp. NRRL S-20 JOBK01*
*Streptomyces sp. NRRL S-237 JODA01*
*Streptomyces sp. NRRL S-241 JOCY01*
*Streptomyces sp. NRRL S-244 JOCX01*
*Streptomyces sp. NRRL S-31 JOCB01*
*Streptomyces sp. NRRL S-325 JOIW01*
*Streptomyces sp. NRRL S-337 JOIX01*
*Streptomyces sp. NRRL S-340 JOIY01*
*Streptomyces sp. NRRL S-350 JOHO01*
*Streptomyces sp. NRRL S-37 JOIZ01*
*Streptomyces sp. NRRL S-378 JOBX01*
*Streptomyces sp. NRRL S-384 JOJA01*
*Streptomyces sp. NRRL S-455 JOCT01*
*Streptomyces sp. NRRL S-474 JODD01*
*Streptomyces sp. NRRL S-475 JOJB01*
*Streptomyces sp. NRRL S-481 JOCV01*
*Streptomyces sp. NRRL S-515 JODE01*
*Streptomyces sp. NRRL S-575 JODJ01*
*Streptomyces sp. NRRL S-623 JOJC01*
*Streptomyces sp. NRRL S-646 JODC01*
*Streptomyces sp. NRRL S-813 JODG01*
*Streptomyces sp. NRRL S-87 JOGB01*
*Streptomyces sp. NRRL S-920 JODF01*
*Streptomyces sp. NRRL S-98 JOJD01*
*Streptomyces sp. NRRL WC-3540 JOCW01*
*Streptomyces sp. NRRL WC-3549 JOCS01*
*Streptomyces sp. NRRL WC-3626 JOCA01*
*Streptomyces sp. NRRL WC-3703 JOCH01*
*Streptomyces sp. NRRL WC-3744 JOCJ01*
*Streptomyces sp. NRRL WC-3773 JOJI01*
*Streptomyces sp. NRRL WC-3774 JOBZ01*
*Streptomyces sp. NTK 937 JJOB01*
*Streptomyces sp. PCS3-D2 JDUZ01*
*Streptomyces sp. PRh5 JABQ01*
*Streptomyces sp. PsTaAH-124 AREJ01*
*Streptomyces sp. PVA 94-07 ASHE01*
*Streptomyces sp. S4 CADY01*
*Streptomyces sp. SA3_actG ADXA01*
*Streptomyces sp. SPB74 ABJG02*
*Streptomyces sp. SPB78 ACEU01*
*Streptomyces sp. TAA040 AUKR01*
*Streptomyces sp. TAA204 AUKW01*
*Streptomyces sp. TOR3209 AGNH01*
*Streptomyces sp. Tu 6176 JFJQ01*
*Streptomyces sp. Tu6071 AFHJ01*
*Streptomyces sp. URHA0041 JNIH01*
*Streptomyces sulphureus DSM 40104 ARLC01*
*Streptomyces sviceus ATCC 29083 ABJJ02*
*Streptomyces tsukubaensis NRRL18488 AJSZ01*
*Streptomyces varsoviensis JOBF01*
*Streptomyces varsoviensis JOFN01*
*Streptomyces venezuelae ATCC 10712*
*Streptomyces vinaceus JNYP01*

***Streptomyces violaceoruber JOCE01***
***Streptomyces violens JOBH01***
***Streptomyces virginiae JOAK01***
***Streptomyces viridochromogenes DSM 40736 ACEZ01***
***Streptomyces viridochromogenes Tue57 AMLP01***
***Streptomyces viridosporus T7A AJFD01***
***Streptomyces vitaminophilus DSM 41686 ARCK01***
***Streptomyces wadayamensis JHDU01***
***Streptomyces xanthophaeus JNZH01***
***Streptomyces zinciresistens K42 AGBF01***
***Streptosporangium amethystogenes JOEQ01***
***Streptosporangium roseum JOEP01***
***Thermobifida fusca TM51 AOSG01***
***Thermocrispum agreste DSM 44070 ATYX01***
***Thermocrispum municipale DSM 44069 ATYT01***
***Tsukamurella paurometabola NC_014158.1***
***Tsukamurella sp. 1534 CAJY01***

## 10.4 Expansion Analysis for Central Carbon and Amino Acid Metabolism

On the following pages the complete expansion analysis table is shown in readable partitions. Highlighted in yellow are the expansions, a functional role was marked as expanded if the mean value in the genus was higher as the overall mean for all analysed genera. Explanations are as following:

**Part 1** **A** Expansion analysis for glycolysis and gluconeogenesis (GNG) for the genera of the suborders Streptomycineae, Catenulisporineae, Streptosporangineae, Frankineae and Pseudonocardineae

**B** Expansion analysis for glycolysis and gluconeogenesis (GNG) for the genera of the suborders Corynebacterineae, Micromonosporineae and Glycomycineae

**C** Expansion analysis for glycolysis and gluconeogenesis (GNG) for the genera of the suborders Micrococcineae, Bifidobacteriales, Actinomycineae, Kineosporiineae and Propionibacterineae

**Part 2** **A** Expansion analysis for the tricarboxylic acid cycle (TCA) for the genera of the suborders Streptomycineae, Catenulisporineae, Streptosporangineae, Frankineae and Pseudonocardineae

**B** Expansion analysis for the tricarboxylic acid cycle (TCA) for the genera of the suborders Corynebacterineae, Micromonosporineae and Glycomycineae

**C** Expansion analysis for the tricarboxylic acid cycle (TCA) for the genera of the suborders Micrococcineae, Bifidobacteriales, Actinomycineae, Kineosporiineae and Propionibacterineae

**Part 3** **A** Expansion analysis for amino acids derived from Pyruvate and Threonine and the pentose phosphate pathway (PPP) for the genera of the suborders Streptomycineae, Catenulisporineae, Streptosporangineae, Frankineae and Pseudonocardineae

**B** Expansion for amino acids derived from Pyruvate and Threonine and the pentose phosphate pathway (PPP) for the genera of the suborders Corynebacterineae, Micromonosporineae and Glycomycineae

**C** Expansion analysis for amino acids derived from Pyruvate and Threonine and the pentose phosphate pathway (PPP) for the genera of the suborders Micrococcineae, Bifidobacteriales, Actinomycineae, Kineosporiineae and Propionibacterineae

**Part 4** **A** Expansion analysis for amino acids derived from α-ketoglutarate (AKG) for the genera of the suborders Streptomycineae, Catenulisporineae, Streptosporangineae, Frankineae and Pseudonocardineae

**B** Expansion analysis for amino acids derived from α-ketoglutarate (AKG) for the genera of the suborders Corynebacterineae, Micromonosporineae and Glycomycineae

C Expansion analysis for amino acids derived from α-ketoglutarate (AKG) for the genera of the suborders Micrococcineae, Bifidobacteriales, Actinomycineae, Kineosporiineae and Propionibacterineae

Part 5 A Expansion analysis for amino acids derived from oxaloacetate (OAA) for the genera of the suborders Streptomycineae, Catenulisporineae, Streptosporangineae, Frankineae and Pseudonocardineae

B Expansion analysis for amino acids derived from oxaloacetate (OAA) for the genera of the suborders Corynebacterineae, Micromonosporineae and Glycomycineae

C Expansion analysis for amino acids derived from oxaloacetate (OAA) for the genera of the suborders Micrococcineae, Bifidobacteriales, Actinomycineae, Kineosporiineae and Propionibacterineae

Part 6 A Expansion analysis for amino acids derived from 3-Phosphoglycerate (3PGA) and Ribose-5-Phosphate (R5P) for the genera of the suborders Streptomycineae, Catenulisporineae, Streptosporangineae, Frankineae and Pseudonocardineae

B Expansion analysis for amino acids derived from 3-Phosphoglycerate (3PGA) and Ribose-5-Phosphate (R5P) for the genera of the suborders Corynebacterineae, Micromonosporineae and Glycomycineae

C Expansion analysis for amino acids derived from 3-Phosphoglycerate (3PGA) and Ribose-5-Phosphate (R5P) for the genera of the suborders Micrococcineae, Bifidobacteriales, Actinomycineae, Kineosporiineae and Propionibacterineae

Part 7 A Expansion analysis for amino acids derived from Erythrose-4-Phosphate (E4P) or Phosphoenolpyruvate (PEP) for the genera of the suborders Streptomycineae, Catenulisporineae, Streptosporangineae, Frankineae and Pseudonocardineae

B Expansion analysis for amino acids derived from Erythrose-4-Phosphate (E4P) or Phosphoenolpyruvate (PEP) for the genera of the suborders Corynebacterineae, Micromonosporineae and Glycomycineae

C Expansion analysis for amino acids derived from Erythrose-4-Phosphate (E4P) or Phosphoenolpyruvate (PEP) for the genera of the suborders Micrococcineae, Bifidobacteriales, Actinomycineae, Kineosporiineae and Propionibacterineae

Part 1 A

| | | | Glycolysis | | | | | | | | | | GNG | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Glucokinase (EC 2.7.1.2) | Enolase (EC 4.2.1.11) | FBP aldolase (EC 4.1.2.13) | NAD-dep GAPDH (EC 1.2.1.12) | PFK (EC 2.7.1.56) | G6P isomerase (EC 5.3.1.9) | PG kinase (EC 2.7.2.3) | PGM (EC 5.4.2.1) | Pyr kinase (EC 2.7.1.40) | Triosephosphate isomerase (EC 5.3.1.1) | FBPase (EC 3.1.3.11) | PEP carboxykinase [GTP] (EC 4.1.1.32) | PEP synthase (EC 2.7.9.2) | PPDK (EC 2.7.9.1) | Pyr carboxyl transferase (EC 6.4.1.1) |
| Streptomycineae | *Kitasatospora* | 5 | 1.80 | 1.20 | 0.80 | 1.80 | 0.80 | 1.00 | 1.00 | 3.60 | 1.80 | 1.00 | 1.20 | 1.00 | 0.60 | 1.00 | 0.00 |
| | *Streptacidiphilus* | 9 | 2.56 | 1.33 | 1.00 | 1.89 | 0.89 | 1.22 | 1.00 | 4.22 | 2.00 | 1.11 | 1.00 | 1.00 | 0.89 | 1.33 | 0.00 |
| | *Streptomyces* | 289 | 0.37 | 1.50 | 1.05 | 1.81 | 1.76 | 1.29 | 1.06 | 2.16 | 2.00 | 1.18 | 1.02 | 1.01 | 0.96 | 1.37 | 0.00 |
| Catenulisporineae | *Actinospica* | 2 | 0.00 | 1.50 | 1.50 | 1.50 | 2.00 | 1.00 | 1.00 | 3.50 | 1.50 | 1.00 | 1.00 | 1.00 | 0.00 | 2.00 | 0.00 |
| | *Catenulispora* | 1 | 3.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 |
| Streptosporangineae | *Nocardiopsis* | 9 | 0.22 | 1.00 | 1.00 | 1.00 | 0.89 | 1.00 | 1.00 | 3.44 | 1.00 | 1.22 | 1.00 | 1.11 | 1.44 | 1.22 | 0.11 |
| | *Thermobifida* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | *Herbidospora* | 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 |
| | *Streptomonospora* | 1 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 |
| | *Microbispora* | 4 | 0.50 | 0.75 | 0.75 | 2.00 | 0.50 | 1.25 | 0.75 | 2.00 | 0.75 | 0.75 | 1.00 | 0.75 | 1.75 | 0.75 | 0.00 |
| | *Microtetraspora* | 1 | 0.00 | 2.00 | 1.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | *Streptosporangium* | 2 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.50 | 1.00 | 3.00 | 1.50 | 0.00 |
| | *Actinomadura* | 5 | 1.00 | 1.20 | 1.00 | 1.00 | 0.60 | 3.60 | 1.00 | 4.00 | 1.00 | 1.20 | 1.00 | 1.20 | 2.00 | 1.20 | 0.00 |
| | *Actinopolyspora* | 4 | 0.25 | 1.00 | 1.25 | 1.00 | 1.00 | 1.00 | 1.00 | 3.25 | 1.00 | 1.25 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| Frankineae | *Frankia* | 9 | 0.00 | 1.00 | 1.00 | 1.22 | 0.11 | 1.00 | 1.00 | 3.44 | 1.00 | 1.00 | 1.00 | 1.11 | 1.67 | 0.67 | 0.56 |
| | *Modestobacter* | 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 5.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 |
| | *Blastococcus* | 1 | 0.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 |
| Pseudonocardineae | *Actinoalloteichus* | 3 | 0.00 | 1.00 | 1.00 | 1.00 | 0.67 | 1.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.67 | 0.00 | 0.00 |
| | *Actinokineospora* | 3 | 0.00 | 1.00 | 1.00 | 1.00 | 0.67 | 1.00 | 1.00 | 3.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.67 | 0.33 | 0.00 |
| | *Actinomycetospora* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 2.00 | 0.00 |
| | *Allokutzneria* | 1 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 6.00 | 0.00 | 0.00 |
| | *Amycolatopsis* | 20 | 0.95 | 1.50 | 1.25 | 1.20 | 1.10 | 1.05 | 0.95 | 3.75 | 0.95 | 0.95 | 1.15 | 0.95 | 1.35 | 0.50 | 0.00 |
| | *Kibdelosporangium* | 1 | 2.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 6.00 | 1.00 | 0.00 |
| | *Kutzneria* | 2 | 1.00 | 0.50 | 1.50 | 2.50 | 1.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 4.50 | 1.00 | 0.00 |
| | *Lechevalieria* | 1 | 0.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 6.00 | 2.00 | 0.00 |
| | *Lentzea* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 1.00 | 0.00 |
| | *Prauserella* | 2 | 0.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.50 | 0.50 | 0.00 |
| | *Pseudonocardia* | 5 | 0.80 | 1.40 | 2.20 | 1.60 | 0.80 | 1.40 | 1.00 | 4.40 | 1.20 | 1.00 | 1.40 | 1.20 | 5.00 | 1.20 | 0.00 |
| | *Sciscionella* | 2 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.50 | 0.50 | 0.50 | 0.00 |
| | *Thermocrispum* | 2 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 0.50 | 0.00 |
| | *Saccharomonospora* | 7 | 0.00 | 0.86 | 1.29 | 2.00 | 0.57 | 0.86 | 0.86 | 2.86 | 1.00 | 1.14 | 1.29 | 1.00 | 0.29 | 0.00 | 0.14 |
| | *Saccharopolyspora* | 4 | 0.00 | 1.25 | 1.00 | 1.00 | 0.25 | 1.00 | 1.25 | 4.50 | 1.00 | 1.00 | 1.50 | 1.25 | 3.25 | 0.00 | 0.50 |
| | *Saccharothrix* | 2 | 0.50 | 1.50 | 1.00 | 1.00 | 0.00 | 1.00 | 1.50 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 |

AA

Part 1 B

| | | | Glycolysis | | | | | | | | | | GNG | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Glucokinase (EC 2.7.1.2) | Enolase (EC 4.2.1.11) | FBP aldolase (EC 4.1.2.13) | NAD-dep GAPDH (EC 1.2.1.12) | PFK (EC 2.7.1.56) | G6P isomerase (EC 5.3.1.9) | PG kinase (EC 2.7.2.3) | PGM (EC 5.4.2.1) | Pyr kinase (EC 2.7.1.40) | Triosephosphate isomerase (EC 5.3.1.1) | FBPase (EC 3.1.3.11) | PEP carboxykinase [GTP] (EC 4.1.1.32) | PEP synthase (EC 2.7.9.2) | PPDK (EC 2.7.9.1) | Pyr carboxyl transferase (EC 6.4.1.1) |
| Corynebacterineae | *Corynebacterium* | 7 | 0.00 | 1.00 | 1.00 | 1.00 | 0.86 | 1.00 | 1.00 | 1.86 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | *Dietzia* | 2 | 0.00 | 1.00 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 0.00 |
| | *Gordonia* | 3 | 0.00 | 1.33 | 1.00 | 1.00 | 0.33 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.33 | 0.00 | 0.00 |
| | *Mycobacterium* | 4 | 0.00 | 1.00 | 1.50 | 1.00 | 2.50 | 1.00 | 1.00 | 3.25 | 1.50 | 1.50 | 1.00 | 1.00 | 2.25 | 0.50 | 0.25 |
| | *Nocardia* | 28 | 0.11 | 1.00 | 1.14 | 1.04 | 1.75 | 1.14 | 1.04 | 3.43 | 0.96 | 1.04 | 1.64 | 1.75 | 4.21 | 1.11 | 0.00 |
| | *Rhodococcus* | 15 | 0.07 | 1.07 | 0.93 | 1.00 | 1.33 | 1.20 | 0.93 | 2.53 | 1.40 | 1.00 | 0.93 | 1.07 | 1.47 | 0.40 | 0.00 |
| | *Tsukamurella* | 2 | 0.00 | 1.00 | 0.50 | 1.50 | 1.00 | 1.00 | 1.00 | 2.00 | 1.50 | 1.00 | 1.00 | 1.00 | 0.50 | 0.00 | 0.50 |
| | *Segniliparus* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| Micromonosporineae | *Actinoplanes* | 3 | 0.33 | 1.33 | 2.33 | 3.67 | 1.00 | 1.00 | 1.00 | 5.00 | 1.00 | 1.00 | 1.33 | 1.00 | 1.33 | 2.00 | 0.00 |
| | *Catenuloplanes* | 1 | 0.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | 5.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 |
| | *Actinocatenispora* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 0.00 | 3.00 | 1.00 | 0.00 |
| | *Dactylosporangium* | 1 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 6.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 0.00 |
| | *Hamadaea* | 1 | 0.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | *Longispora* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | *Micromonospora* | 7 | 0.00 | 1.29 | 1.00 | 1.00 | 0.14 | 1.57 | 1.00 | 3.29 | 1.14 | 1.00 | 1.00 | 1.00 | 1.71 | 1.71 | 0.00 |
| | *Pilimelia* | 1 | 0.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 0.00 | 0.00 |
| | *Salinispora* | 3 | 0.00 | 1.00 | 1.33 | 1.00 | 0.67 | 1.00 | 1.00 | 3.00 | 1.00 | 1.33 | 1.00 | 1.00 | 0.33 | 1.00 | 0.00 |
| Glycomycineae | *Glycomyces* | 3 | 0.33 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.67 | 1.00 | 1.00 | 0.00 | 1.00 | 0.33 | 0.00 | 0.00 |

BB

Part 1 C

| | | | Glycolysis | | | | | | | | | | GNG | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Glucokinase (EC 2.7.1.2) | Enolase (EC 4.2.1.11) | FBP aldolase (EC 4.1.2.13) | NAD-dep GAPDH (EC 1.2.1.12) | PFK (EC 2.7.1.56) | G6P isomerase (EC 5.3.1.9) | PG kinase (EC 2.7.2.3) | PGM (EC 5.4.2.1) | Pyr kinase (EC 2.7.1.40) | Triosephosphate isomerase (EC 5.3.1.1) | FBPase (EC 3.1.3.11) | PEP carboxykinase [GTP] (EC 4.1.1.32) | PEP synthase (EC 2.7.9.2) | PPDK (EC 2.7.9.1) | Pyr carboxyl transferase (EC 6.4.1.1) |
| Micrococcineae | *Brevibacterium* | 6 | 0.00 | 1.00 | 1.00 | 1.50 | 0.67 | 1.00 | 1.00 | 3.67 | 1.00 | 1.17 | 1.00 | 1.33 | 0.33 | 0.00 | 0.00 |
| | *Cellulomonas* | 5 | 1.80 | 1.40 | 2.40 | 1.20 | 1.20 | 1.00 | 1.00 | 3.60 | 1.00 | 1.00 | 0.00 | 0.80 | 0.20 | 0.60 | 0.00 |
| | *Actinotalea* | 2 | 0.50 | 1.00 | 1.50 | 1.50 | 0.50 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | *Dermabacter* | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| | *Intrasporangium* | 2 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.50 |
| | *Serinicoccus* | 2 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 0.50 | 0.50 | 1.00 | 0.00 |
| | *Knoellia* | 4 | 0.00 | 1.00 | 1.25 | 1.00 | 0.25 | 1.00 | 1.00 | 4.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 0.00 |
| | *Agromyces* | 2 | 1.00 | 1.00 | 1.00 | 2.00 | 0.50 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.50 | 0.00 | 0.00 |
| | *Clavibacter* | 2 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.50 | 1.00 | 2.00 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 |
| | *Leifsonia* | 3 | 0.67 | 2.00 | 1.00 | 1.67 | 1.67 | 1.00 | 1.00 | 3.67 | 1.00 | 1.33 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | *Leucobacter* | 4 | 0.00 | 1.00 | 1.00 | 1.00 | 0.25 | 1.00 | 1.00 | 1.25 | 1.00 | 1.00 | 1.00 | 1.25 | 0.00 | 0.00 | 0.00 |
| | *Microbacterium* | 18 | 0.39 | 1.11 | 1.28 | 1.22 | 1.06 | 1.33 | 1.06 | 2.61 | 1.00 | 1.44 | 1.00 | 1.00 | 0.39 | 0.06 | 0.06 |
| | *Mycetocola* | 1 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | *Arthrobacter* | 10 | 0.40 | 1.00 | 1.70 | 0.90 | 1.20 | 0.90 | 1.00 | 3.40 | 1.00 | 1.70 | 0.90 | 1.10 | 3.00 | 0.00 | 0.00 |
| | *Citricoccus* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 0.00 | 0.00 |
| | *Kocuria* | 6 | 0.17 | 1.00 | 1.00 | 1.00 | 1.00 | 0.83 | 1.00 | 2.17 | 0.83 | 1.17 | 1.00 | 1.00 | 0.83 | 0.00 | 0.00 |
| | *Rothia* | 4 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 0.00 | 0.00 |
| Bifidobacteriales | *Bifidobacterium* | 3 | 0.67 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.67 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Actinomycineae | *Actinobaculum* | 5 | 0.80 | 1.00 | 1.00 | 1.00 | 0.20 | 1.00 | 1.00 | 2.00 | 1.00 | 1.40 | 0.00 | 0.20 | 0.00 | 1.00 | 0.00 |
| | *Actinomyces* | 21 | 0.33 | 1.00 | 1.67 | 1.24 | 0.48 | 1.14 | 1.00 | 2.33 | 1.05 | 1.10 | 0.00 | 1.33 | 0.10 | 1.00 | 0.00 |
| Kineosporiineae | *Kineococcus* | 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 6.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 0.00 | 0.00 |
| Propionibacterineae | *Propionibacterium* | 9 | 0.22 | 1.33 | 2.67 | 1.44 | 1.22 | 1.00 | 1.00 | 2.78 | 1.00 | 1.56 | 0.00 | 0.11 | 0.22 | 1.00 | 0.00 |
| | *Actinopolymorpha* | 1 | 0.00 | 2.00 | 3.00 | 2.00 | 1.00 | 2.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 1.00 | 0.00 |
| | *Aeromicrobium* | 2 | 0.00 | 1.00 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 | 3.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | *Pimelobacter* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | *Kribbella* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 5.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 2.00 | 0.00 |
| | *Marmoricola* | 2 | 0.00 | 1.00 | 1.50 | 1.00 | 0.50 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 0.00 |
| | *Nocardioides* | 7 | 0.14 | 1.00 | 1.14 | 1.14 | 0.86 | 1.00 | 1.14 | 3.43 | 1.29 | 1.00 | 1.14 | 1.57 | 0.57 | 0.43 | 0.00 |
| | *Propionicicella* | 1 | 0.00 | 3.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 3.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| | Mean + Std | 614 | 1.05 | 1.58 | 1.72 | 1.86 | 1.39 | 1.46 | 1.15 | 4.27 | 1.34 | 1.43 | 1.52 | 1.36 | 2.98 | 1.36 | 0.15 |

CC

Part 2 A

| Suborder | Genera | # of species | TCA | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Citrate synthase (si) (EC 2.3.3.1) | Aconitate hydratase (EC 4.2.1.3) | Isocitrate DH [NADP] (EC 1.1.1.42) | Isocitrate lyase (EC 4.1.3.1) | Suc DH flavoprotein su (EC 1.3.99.1) | 2-oxoglutarate DH E1 (EC 1.2.4.2) | Succinyl-CoA ligase, α chain (EC 6.2.1.5) | Fumarate hydratase (EC 4.2.1.2) | Malate dehydrogenase (EC 1.1.1.37) | D-malic enzyme (EC 1.1.1.83) | NAD-de malic enzyme (EC 1.1.1.38) | NADP-dep malic enzyme (EC 1.1.1.40) | Malate synthase (EC 2.3.3.9) | Pyruvate dehydrogenase E1 (EC 1.2.4.1) | Pyruvate DH E1 alpha subunit (EC 1.2.4.1) | Pyruvate DH E1 beta subunit (EC 1.2.4.1) | PEP carboxylase (EC 4.1.1.31) |
| Streptomycineae | *Kitasatospora* | 5 | 2.80 | 1.20 | 1.00 | 0.60 | 2.60 | 1.00 | 1.00 | 1.80 | 1.40 | 0.00 | 1.00 | 1.00 | 1.20 | 1.20 | 1.20 | 1.20 | 1.00 |
| | *Streptacidiphilus* | 9 | 2.67 | 1.56 | 1.00 | 0.44 | 2.22 | 1.00 | 1.11 | 1.89 | 1.33 | 0.00 | 1.44 | 1.11 | 1.00 | 1.56 | 1.00 | 1.22 | 1.00 |
| | *Streptomyces* | 289 | 3.42 | 1.14 | 1.06 | 0.36 | 2.83 | 1.01 | 1.29 | 2.05 | 1.14 | 0.24 | 1.15 | 1.06 | 1.35 | 2.18 | 0.72 | 0.77 | 1.00 |
| Catenulisporineae | *Actinospica* | 2 | 2.00 | 1.50 | 1.00 | 0.00 | 4.00 | 1.00 | 1.50 | 2.50 | 1.00 | 0.00 | 1.50 | 1.00 | 0.50 | 2.00 | 0.50 | 0.50 | 1.00 |
| | *Catenulispora* | 1 | 3.00 | 2.00 | 1.00 | 0.00 | 4.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 2.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| Streptosporangineae | *Nocardiopsis* | 9 | 1.56 | 1.33 | 1.22 | 1.11 | 1.56 | 1.00 | 0.89 | 0.67 | 1.44 | 0.67 | 1.00 | 1.00 | 1.00 | 0.00 | 2.11 | 2.00 | 1.00 |
| | *Thermobifida* | 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 2.00 | 2.00 | 1.00 |
| | *Herbidospora* | 1 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Streptomonospora* | 1 | 3.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 2.00 | 2.00 | 1.00 |
| | *Microbispora* | 4 | 2.50 | 1.00 | 0.75 | 0.75 | 1.75 | 0.75 | 0.75 | 1.50 | 0.50 | 0.00 | 0.75 | 0.75 | 0.75 | 1.50 | 1.00 | 1.00 | 0.75 |
| | *Microtetraspora* | 1 | 2.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Streptosporangium* | 2 | 2.00 | 1.00 | 1.00 | 1.00 | 2.50 | 1.00 | 1.00 | 2.00 | 2.00 | 0.50 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.50 | 1.00 |
| | *Actinomadura* | 5 | 3.00 | 1.60 | 1.00 | 1.40 | 3.00 | 1.00 | 1.00 | 2.20 | 1.60 | 0.20 | 1.00 | 1.00 | 1.00 | 1.40 | 1.80 | 1.80 | 1.00 |
| | *Actinopolyspora* | 4 | 2.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.25 | 1.00 | 0.25 | 0.00 | 2.25 | 1.25 | 1.25 | 1.75 | 1.75 | 0.00 |
| Frankineae | *Frankia* | 9 | 2.22 | 1.00 | 1.00 | 0.33 | 1.89 | 1.00 | 1.00 | 1.00 | 1.22 | 0.00 | 1.11 | 0.78 | 1.11 | 2.00 | 1.56 | 1.11 | 0.33 |
| | *Modestobacter* | 1 | 3.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Blastococcus* | 1 | 3.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Pseudonocardineae | *Actinoalloteichus* | 3 | 4.00 | 1.00 | 1.00 | 0.67 | 2.00 | 1.00 | 0.67 | 0.67 | 1.00 | 0.00 | 1.00 | 2.00 | 0.67 | 1.00 | 1.67 | 2.00 | 0.00 |
| | *Actinokineospora* | 3 | 2.00 | 1.00 | 1.00 | 0.00 | 2.67 | 1.33 | 1.00 | 2.00 | 2.00 | 0.00 | 0.00 | 2.00 | 0.33 | 1.00 | 2.33 | 1.33 | 0.00 |
| | *Actinomycetospora* | 1 | 4.00 | 2.00 | 1.00 | 0.00 | 3.00 | 1.00 | 1.00 | 3.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 5.00 | 1.00 | 1.00 | 1.00 |
| | *Allokutzneria* | 1 | 2.00 | 1.00 | 1.00 | 0.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 |
| | *Amycolatopsis* | 20 | 2.60 | 1.60 | 0.95 | 0.95 | 2.60 | 0.90 | 1.40 | 2.20 | 2.15 | 1.20 | 0.05 | 2.10 | 1.90 | 1.45 | 1.15 | 1.35 | 0.00 |
| | *Kibdelosporangium* | 1 | 4.00 | 2.00 | 1.00 | 0.00 | 2.00 | 1.00 | 3.00 | 3.00 | 3.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 2.00 | 2.00 | 0.00 |
| | *Kutzneria* | 2 | 3.50 | 2.50 | 1.00 | 1.00 | 3.00 | 1.00 | 1.50 | 2.50 | 2.00 | 0.50 | 0.50 | 4.50 | 2.00 | 2.00 | 1.50 | 2.00 | 0.00 |
| | *Lechevalieria* | 1 | 4.00 | 1.00 | 1.00 | 0.00 | 3.00 | 1.00 | 2.00 | 2.00 | 1.00 | 0.00 | 0.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 |
| | *Lentzea* | 1 | 4.00 | 2.00 | 1.00 | 0.00 | 3.00 | 1.00 | 2.00 | 3.00 | 2.00 | 0.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 |
| | *Prauserella* | 2 | 3.00 | 1.00 | 1.00 | 2.00 | 3.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 0.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 |
| | *Pseudonocardia* | 5 | 4.00 | 1.60 | 1.00 | 1.00 | 3.00 | 0.60 | 2.00 | 2.20 | 1.80 | 1.00 | 0.00 | 2.00 | 1.00 | 1.60 | 1.40 | 2.00 | 0.80 |
| | *Sciscionella* | 2 | 5.00 | 2.00 | 1.00 | 2.50 | 2.00 | 1.00 | 2.00 | 2.00 | 1.50 | 1.00 | 0.00 | 2.00 | 2.00 | 2.50 | 1.50 | 1.50 | 0.00 |
| | *Thermocrispum* | 2 | 2.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 0.00 | 0.00 | 2.00 | 0.50 | 1.00 | 1.00 | 1.00 | 0.00 |
| | *Saccharomonospora* | 7 | 2.14 | 1.71 | 1.14 | 1.00 | 2.00 | 1.00 | 1.14 | 2.00 | 1.00 | 1.00 | 0.00 | 2.29 | 2.00 | 1.71 | 1.43 | 1.71 | 0.00 |
| | *Saccharopolyspora* | 4 | 2.00 | 2.00 | 1.25 | 1.00 | 2.75 | 0.75 | 1.00 | 2.00 | 1.75 | 1.25 | 0.00 | 2.00 | 1.00 | 1.50 | 2.25 | 3.00 | 0.00 |
| | *Saccharothrix* | 2 | 2.00 | 1.00 | 1.00 | 0.00 | 4.00 | 1.00 | 1.00 | 2.00 | 2.00 | 0.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 |

DD

Part 2 B

| Suborder | Genera | # of species | TCA | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Citrate synthase (si) (EC 2.3.3.1) | Aconitate hydratase (EC 4.2.1.3) | Isocitrate DH [NADP] (EC 1.1.1.42) | Isocitrate lyase (EC 4.1.3.1) | Suc DH flavoprotein su (EC 1.3.99.1) | 2-oxoglutarate DH E1 (EC 1.2.4.2) | Succinyl-CoA ligase, α chain (EC 6.2.1.5) | Fumarate hydratase (EC 4.2.1.2) | Malate dehydrogenase (EC 1.1.1.37) | D-malic enzyme (EC 1.1.1.83) | NAD-de malic enzyme (EC 1.1.1.38) | NADP-dep malic enzyme (EC 1.1.1.40) | Malate synthase (EC 2.3.3.9) | Pyruvate dehydrogenase E1 (EC 1.2.4.1) | Pyruvate DH E1 alpha subunit (EC 1.2.4.1) | Pyruvate DH E1 beta subunit (EC 1.2.4.1) | PEP carboxylase (EC 4.1.1.31) |
| Corynebacterineae | *Corynebacterium* | 7 | 1.00 | 1.14 | 1.00 | 0.43 | 1.00 | 0.86 | 0.29 | 1.00 | 1.00 | 0.00 | 0.00 | 0.29 | 0.43 | 1.00 | 0.00 | 0.00 | 0.71 |
| | *Dietzia* | 2 | 1.00 | 1.50 | 1.50 | 1.50 | 2.50 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 |
| | *Gordonia* | 3 | 2.33 | 1.00 | 1.33 | 1.00 | 2.00 | 0.67 | 1.00 | 2.00 | 0.00 | 0.00 | 0.00 | 1.33 | 1.00 | 1.00 | 0.67 | 0.67 | 1.00 |
| | *Mycobacterium* | 4 | 2.75 | 1.50 | 1.50 | 3.00 | 3.50 | 1.25 | 1.00 | 1.50 | 0.50 | 0.50 | 0.50 | 0.50 | 1.00 | 1.75 | 0.00 | 0.00 | 0.75 |
| | *Nocardia* | 28 | 2.61 | 1.46 | 1.00 | 1.04 | 3.57 | 1.07 | 1.00 | 2.07 | 0.75 | 1.07 | 0.00 | 1.07 | 1.00 | 1.21 | 1.50 | 1.82 | 1.04 |
| | *Rhodococcus* | 15 | 2.53 | 1.47 | 1.27 | 0.87 | 3.07 | 0.93 | 1.07 | 1.87 | 0.93 | 0.47 | 0.07 | 1.53 | 0.87 | 2.20 | 1.07 | 1.13 | 0.87 |
| | *Tsukamurella* | 2 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 |
| | *Segniliparus* | 1 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 |
| Micromonosporineae | *Actinoplanes* | 3 | 5.67 | 1.00 | 1.00 | 1.33 | 3.00 | 1.00 | 1.00 | 2.00 | 1.33 | 0.00 | 0.00 | 2.67 | 2.00 | 1.33 | 1.00 | 1.00 | 1.00 |
| | *Catenuloplanes* | 1 | 5.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 |
| | *Actinocatenispora* | 1 | 3.00 | 2.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 |
| | *Dactylosporangium* | 1 | 3.00 | 1.00 | 2.00 | 0.00 | 2.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 3.00 | 1.00 |
| | *Hamadaea* | 1 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Longispora* | 1 | 2.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 0.00 | 2.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Micromonospora* | 7 | 2.29 | 1.00 | 1.00 | 0.29 | 2.00 | 0.57 | 0.86 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.29 | 1.86 | 1.29 | 1.29 | 1.00 |
| | *Pilimelia* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 4.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Salinispora* | 3 | 2.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 0.67 | 1.00 |
| Glycomycineae | *Glycomyces* | 3 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.33 | 0.33 | 0.00 | 2.00 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 |

EE

Part 2 C

| | | | TCA | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Citrate synthase (si) (EC 2.3.3.1) | Aconitate hydratase (EC 4.2.1.3) | Isocitrate DH [NADP] (EC 1.1.1.42) | Isocitrate lyase (EC 4.1.3.1) | Suc DH flavoprotein su (EC 1.3.99.1) | 2-oxoglutarate DH E1 (EC 1.2.4.2) | Succinyl-CoA ligase, α chain (EC 6.2.1.5) | Fumarate hydratase (EC 4.2.1.2) | Malate dehydrogenase (EC 1.1.1.37) | D-malic enzyme (EC 1.1.1.83) | NAD-de malic enzyme (EC 1.1.1.38) | NADP-dep malic enzyme (EC 1.1.1.40) | Malate synthase (EC 2.3.3.9) | Pyruvate dehydrogenase E1 (EC 1.2.4.1) | Pyruvate DH E1 alpha subunit (EC 1.2.4.1) | Pyruvate DH E1 beta subunit (EC 1.2.4.1) | PEP carboxylase (EC 4.1.1.31) |
| Micrococcineae | *Brevibacterium* | 6 | 1.00 | 1.67 | 1.00 | 1.00 | 1.67 | 1.00 | 1.00 | 1.33 | 0.67 | 0.67 | 0.00 | 1.00 | 2.00 | 2.33 | 1.17 | 1.17 | 0.00 |
| | *Cellulomonas* | 5 | 1.00 | 1.00 | 1.00 | 0.60 | 1.20 | 1.20 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.60 | 1.00 | 1.40 | 1.40 | 0.80 |
| | *Actinotalea* | 2 | 1.00 | 1.50 | 1.50 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.50 | 1.00 | 1.00 | 1.50 | 1.50 | 0.00 |
| | *Dermabacter* | 2 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Intrasporangium* | 2 | 3.50 | 1.50 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.50 | 0.50 | 2.50 | 0.50 | 1.00 | 1.00 | 1.50 | 0.50 |
| | *Serinicoccus* | 2 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 0.00 | 1.50 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Knoellia* | 4 | 2.00 | 1.25 | 1.00 | 0.00 | 1.25 | 1.00 | 1.00 | 1.75 | 1.25 | 0.00 | 1.00 | 1.00 | 0.25 | 1.00 | 1.00 | 1.00 | 0.00 |
| | *Agromyces* | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.50 | 2.00 | 1.00 | 2.00 | 2.00 | 0.00 |
| | *Clavibacter* | 2 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.50 | 0.50 | 1.00 | 1.00 |
| | *Leifsonia* | 3 | 2.33 | 1.67 | 1.33 | 1.00 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.33 | 2.33 | 2.33 | 0.00 |
| | *Leucobacter* | 4 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.25 | 0.00 | 0.00 | 0.00 | 1.50 | 1.00 | 1.00 | 1.00 | 0.00 |
| | *Microbacterium* | 18 | 1.11 | 1.06 | 1.00 | 0.89 | 1.44 | 1.17 | 1.00 | 1.00 | 0.78 | 0.06 | 0.00 | 0.11 | 1.06 | 1.06 | 1.94 | 1.83 | 0.83 |
| | *Mycetocola* | 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 |
| | *Arthrobacter* | 10 | 1.30 | 1.40 | 1.10 | 1.10 | 1.30 | 1.00 | 1.00 | 1.20 | 0.30 | 0.40 | 0.00 | 1.60 | 1.90 | 1.50 | 1.40 | 2.90 | 0.50 |
| | *Citricoccus* | 1 | 2.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 2.00 | 3.00 | 0.00 |
| | *Kocuria* | 6 | 1.00 | 1.00 | 1.00 | 0.83 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 2.17 | 0.00 | 0.83 | 1.17 | 1.83 | 1.00 | 1.17 | 0.33 |
| | *Rothia* | 4 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 |
| Bifidobacteriales | *Bifidobacterium* | 3 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Actinomycineae | *Actinobaculum* | 5 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.20 | 0.80 | 0.20 | 0.00 | 0.20 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.80 |
| | *Actinomyces* | 21 | 0.71 | 0.71 | 0.71 | 0.00 | 1.43 | 0.67 | 0.67 | 0.71 | 1.05 | 0.00 | 0.14 | 0.95 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| Kineosporiineae | *Kineococcus* | 1 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Propionibacterineae | *Propionibacterium* | 9 | 1.89 | 1.00 | 1.00 | 0.00 | 2.22 | 0.67 | 0.78 | 1.00 | 1.11 | 0.00 | 1.78 | 0.00 | 0.00 | 0.89 | 1.00 | 1.00 | 0.00 |
| | *Actinopolymorpha* | 1 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Aeromicrobium* | 2 | 2.00 | 1.50 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.50 |
| | *Pimelobacter* | 1 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 2.00 | 0.00 |
| | *Kribbella* | 1 | 5.00 | 1.00 | 2.00 | 0.00 | 4.00 | 1.00 | 2.00 | 2.00 | 3.00 | 0.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 |
| | *Marmoricola* | 2 | 3.00 | 1.00 | 1.00 | 0.00 | 1.50 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.50 | 1.50 | 0.50 |
| | *Nocardioides* | 7 | 2.29 | 1.29 | 1.00 | 0.00 | 1.71 | 1.14 | 1.00 | 2.00 | 0.86 | 0.00 | 1.00 | 1.14 | 0.57 | 1.29 | 1.14 | 1.14 | 0.00 |
| | *Propionicicella* | 1 | 2.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 2.00 | 1.00 | 1.00 | 0.00 |
| | Mean + Std | 614 | 3.51 | 1.70 | 1.32 | 1.23 | 2.95 | 1.20 | 1.51 | 2.31 | 1.76 | 0.74 | 1.12 | 2.03 | 1.53 | 2.04 | 1.73 | 1.93 | 1.04 |

FF

Part 3 A

| | | | AA from PYR and THR | | | | | | | | | | PPP | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Argininosuccinate lyase (EC 4.3.2.1) | Alanine dehydrogenase (EC 1.4.1.1) | Threonine dehydratase (EC 4.3.1.19) | Acetolactate synthase (EC 2.2.1.6) | Ketol-acid reductoisomerase (EC 1.1.1.86) | Dihydroxy-acid dehydratase (EC 4.2.1.9) | Branched-chain AA ATF (EC 2.6.1.42) | 2-isopropylmalate synthase (EC 2.3.3.13) | 3-isopropylmalate DHT (EC 4.2.1.33) | 3-isopropylmalate DH (EC 1.1.1.85) | G6P1 dehydrogenase (EC 1.1.1.49) | 6-phosphogluconate DH (EC 1.1.1.44) | 6-phosphogluconolactonase (EC 3.1.1.31) | RSP isomerase (EC 5.3.1.6) | Ribulose-P 3-epimerase (EC 5.1.3.1) | Transaldolase (EC 2.2.1.2) | Transketolase (EC 2.2.1.1) |
| Streptomycineae | *Kitasatospora* | 5 | 3.00 | 1.40 | 2.80 | 2.40 | 1.00 | 1.20 | 1.80 | 1.20 | 2.40 | 1.20 | 1.40 | 1.40 | 1.00 | 1.20 | 1.20 | 1.20 | 3.00 |
| | *Streptacidiphilus* | 9 | 1.11 | 1.22 | 3.11 | 3.56 | 1.00 | 1.67 | 1.00 | 1.22 | 3.33 | 1.00 | 2.22 | 1.89 | 1.00 | 1.33 | 1.00 | 2.11 | 4.33 |
| | *Streptomyces* | 289 | 1.97 | 1.13 | 2.64 | 4.09 | 1.10 | 1.97 | 1.38 | 1.46 | 2.42 | 1.20 | 1.76 | 2.18 | 1.33 | 1.35 | 1.04 | 1.49 | 3.74 |
| Catenulisporineae | *Actinospica* | 2 | 2.00 | 1.00 | 3.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.50 | 2.00 | 1.00 | 2.50 | 2.50 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 |
| | *Catenulispora* | 1 | 1.00 | 2.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 3.00 | 3.00 | 3.00 | 1.00 | 1.00 | 3.00 | 4.00 |
| Streptosporangineae | *Nocardiopsis* | 9 | 1.33 | 1.00 | 2.33 | 4.44 | 1.00 | 1.44 | 1.22 | 1.89 | 2.22 | 1.00 | 2.56 | 1.00 | 1.33 | 1.44 | 1.00 | 1.00 | 3.78 |
| | *Thermobifida* | 1 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Herbidospora* | 1 | 1.00 | 1.00 | 3.00 | 6.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 |
| | *Streptomonospora* | 1 | 1.00 | 1.00 | 3.00 | 5.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 4.00 |
| | *Microbispora* | 4 | 1.25 | 1.50 | 2.00 | 3.50 | 0.75 | 1.50 | 1.00 | 1.50 | 1.50 | 1.00 | 2.00 | 1.75 | 1.25 | 1.75 | 0.75 | 1.00 | 1.50 |
| | *Microtetraspora* | 1 | 2.00 | 1.00 | 2.00 | 3.00 | 1.00 | 3.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 5.00 |
| | *Streptosporangium* | 2 | 1.50 | 2.00 | 4.50 | 3.50 | 1.00 | 1.50 | 1.50 | 2.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Actinomadura* | 5 | 2.20 | 1.40 | 4.40 | 5.00 | 1.00 | 1.80 | 1.40 | 1.80 | 3.60 | 1.20 | 1.80 | 2.00 | 1.20 | 1.00 | 1.00 | 1.20 | 2.60 |
| | *Actinopolyspora* | 4 | 1.25 | 1.00 | 2.25 | 2.75 | 1.00 | 1.50 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 0.75 | 1.75 | 3.50 |
| Frankineae | *Frankia* | 9 | 1.00 | 0.22 | 1.22 | 2.78 | 1.00 | 1.00 | 1.00 | 1.11 | 2.00 | 1.00 | 1.11 | 1.11 | 1.00 | 1.00 | 1.44 | 1.00 | 1.11 |
| | *Modestobacter* | 1 | 1.00 | 1.00 | 4.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 4.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Blastococcus* | 1 | 1.00 | 0.00 | 2.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Pseudonocardineae | *Actinoalloteichus* | 3 | 3.00 | 2.00 | 1.67 | 2.00 | 1.00 | 1.67 | 1.67 | 1.00 | 2.00 | 3.00 | 1.00 | 1.67 | 2.00 | 3.00 | 1.00 | 1.00 | 6.00 |
| | *Actinokineospora* | 3 | 1.00 | 1.33 | 2.00 | 3.67 | 1.00 | 1.00 | 2.00 | 1.67 | 2.67 | 1.00 | 1.67 | 1.00 | 1.33 | 1.00 | 1.00 | 1.00 | 2.33 |
| | *Actinomycetospora* | 1 | 1.00 | 1.00 | 3.00 | 4.00 | 1.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 5.00 |
| | *Allokutzneria* | 1 | 2.00 | 1.00 | 5.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 3.00 |
| | *Amycolatopsis* | 20 | 2.65 | 1.70 | 5.20 | 8.15 | 1.00 | 3.40 | 1.45 | 1.30 | 2.80 | 1.55 | 1.10 | 1.85 | 2.45 | 2.35 | 1.05 | 1.10 | 4.70 |
| | *Kibdelosporangium* | 1 | 2.00 | 1.00 | 7.00 | 4.00 | 0.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | 1.00 | 3.00 |
| | *Kutzneria* | 2 | 2.50 | 1.00 | 5.50 | 5.00 | 1.50 | 2.00 | 3.00 | 1.00 | 2.00 | 1.50 | 1.00 | 2.00 | 2.50 | 1.00 | 1.00 | 2.50 | 2.50 |
| | *Lechevalieria* | 1 | 3.00 | 2.00 | 5.00 | 8.00 | 1.00 | 4.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 3.00 | 1.00 | 1.00 | 2.00 | 7.00 |
| | *Lentzea* | 1 | 1.00 | 3.00 | 6.00 | 8.00 | 1.00 | 4.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 3.00 |
| | *Prauserella* | 2 | 2.00 | 1.00 | 3.00 | 4.00 | 1.00 | 5.00 | 1.00 | 1.00 | 4.00 | 3.00 | 1.00 | 2.00 | 1.00 | 4.00 | 2.00 | 1.00 | 4.00 |
| | *Pseudonocardia* | 5 | 1.80 | 1.00 | 4.20 | 8.60 | 2.20 | 2.40 | 2.00 | 1.60 | 3.20 | 2.00 | 2.20 | 2.20 | 1.00 | 1.60 | 1.40 | 1.00 | 2.80 |
| | *Sciscionella* | 2 | 2.00 | 1.00 | 4.50 | 7.50 | 1.00 | 1.00 | 3.00 | 1.50 | 6.00 | 2.00 | 1.00 | 2.00 | 1.50 | 2.00 | 1.00 | 2.00 | 5.00 |
| | *Thermocrispum* | 2 | 1.50 | 1.00 | 3.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.50 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Saccharomonospora* | 7 | 1.57 | 0.57 | 3.57 | 4.14 | 1.00 | 2.14 | 0.86 | 1.14 | 2.29 | 1.57 | 1.00 | 2.00 | 0.86 | 2.00 | 1.14 | 1.00 | 2.86 |
| | *Saccharopolyspora* | 4 | 1.25 | 1.50 | 3.50 | 6.50 | 1.25 | 2.50 | 1.50 | 1.50 | 2.00 | 2.00 | 1.00 | 2.25 | 1.25 | 2.50 | 1.00 | 1.50 | 6.00 |
| | *Saccharothrix* | 2 | 1.50 | 1.50 | 3.00 | 2.50 | 1.50 | 3.00 | 2.50 | 1.50 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 4.00 |

GG

Part 3 B

| | | | AA from PYR and THR | | | | | | | | | | PPP | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Argininosuccinate lyase (EC 4.3.2.1) | Alanine dehydrogenase (EC 1.4.1.1) | Threonine dehydratase (EC 4.3.1.19) | Acetolactate synthase (EC 2.2.1.6) | Ketol-acid reductoisomerase (EC 1.1.1.86) | Dihydroxy-acid dehydratase (EC 4.2.1.9) | Branched-chain AA ATF (EC 2.6.1.42) | 2-isopropylmalate synthase (EC 2.3.3.13) | 3-isopropylmalate DHT (EC 4.2.1.33) | 3-isopropylmalate DH (EC 1.1.1.85) | G6P1 dehydrogenase (EC 1.1.1.49) | 6-phosphogluconate DH (EC 1.1.1.44) | 6-phosphogluconolactonase (EC 3.1.1.31) | R5P isomerase (EC 5.3.1.6) | Ribulose-P 3-epimerase (EC 5.1.3.1) | Transaldolase (EC 2.2.1.2) | Transketolase (EC 2.2.1.1) |
| Corynebacterineae | *Corynebacterium* | 7 | 1.00 | 0.43 | 1.00 | 2.14 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.86 | 1.00 |
| | *Dietzia* | 2 | 1.00 | 0.50 | 0.50 | 2.50 | 1.00 | 1.50 | 1.50 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Gordonia* | 3 | 1.00 | 0.33 | 1.33 | 3.00 | 1.00 | 1.33 | 1.00 | 1.33 | 2.33 | 1.67 | 1.00 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Mycobacterium* | 4 | 1.00 | 1.25 | 1.50 | 5.50 | 1.00 | 2.00 | 1.00 | 1.00 | 4.50 | 1.50 | 2.00 | 2.00 | 1.00 | 2.25 | 1.00 | 1.25 | 2.25 |
| | *Nocardia* | 28 | 1.14 | 1.68 | 3.43 | 3.93 | 1.00 | 2.04 | 1.32 | 2.11 | 2.54 | 1.96 | 1.36 | 1.89 | 1.39 | 1.21 | 1.04 | 1.14 | 1.25 |
| | *Rhodococcus* | 15 | 1.07 | 1.40 | 2.73 | 4.87 | 0.93 | 2.07 | 1.47 | 1.73 | 2.53 | 1.93 | 1.53 | 1.93 | 1.07 | 1.20 | 1.00 | 1.33 | 2.07 |
| | *Tsukamurella* | 2 | 1.00 | 1.00 | 1.00 | 4.50 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.50 | 2.00 | 1.00 | 1.50 | 1.00 | 1.00 | 2.00 |
| | *Segniliparus* | 1 | 1.00 | 0.00 | 4.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 |
| Micromonosporineae | *Actinoplanes* | 3 | 1.00 | 2.00 | 4.00 | 4.67 | 1.00 | 1.33 | 2.33 | 1.33 | 2.67 | 1.00 | 3.33 | 2.00 | 2.00 | 2.33 | 1.00 | 1.33 | 7.00 |
| | *Catenuloplanes* | 1 | 2.00 | 1.00 | 5.00 | 4.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 3.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Actinocatenispora* | 1 | 1.00 | 1.00 | 4.00 | 4.00 | 2.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 4.00 |
| | *Dactylosporangium* | 1 | 1.00 | 1.00 | 5.00 | 5.00 | 1.00 | 3.00 | 3.00 | 1.00 | 2.00 | 1.00 | 3.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 3.00 |
| | *Hamadaea* | 1 | 1.00 | 1.00 | 3.00 | 5.00 | 1.00 | 2.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Longispora* | 1 | 2.00 | 1.00 | 3.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 3.00 |
| | *Micromonospora* | 7 | 1.86 | 1.00 | 2.57 | 3.43 | 1.00 | 1.14 | 1.57 | 1.00 | 2.00 | 1.43 | 1.57 | 2.00 | 1.00 | 1.00 | 1.00 | 1.14 | 3.14 |
| | *Pilimelia* | 1 | 1.00 | 1.00 | 4.00 | 4.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 |
| | *Salinispora* | 3 | 1.00 | 1.00 | 3.00 | 5.33 | 1.00 | 1.33 | 1.67 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.67 |
| Glycomycineae | *Glycomyces* | 3 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 3.00 |

Part 3 C

| | | | AA from PYR and THR | | | | | | | | | | PPP | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Argininosuccinate lyase (EC 4.3.2.1) | Alanine dehydrogenase (EC 1.4.1.1) | Threonine dehydratase (EC 4.3.1.19) | Acetolactate synthase (EC 2.2.1.6) | Ketol-acid reductoisomerase (EC 1.1.1.86) | Dihydroxy-acid dehydratase (EC 4.2.1.9) | Branched-chain AA ATF (EC 2.6.1.42) | 2-isopropylmalate synthase (EC 2.3.3.13) | 3-isopropylmalate DHT (EC 4.2.1.33) | 3-isopropylmalate DH (EC 1.1.1.85) | G6P1 dehydrogenase (EC 1.1.1.49) | 6-phosphogluconate DH (EC 1.1.1.44) | 6-phosphogluconolactonase (EC 3.1.1.31) | R5P isomerase (EC 5.3.1.6) | Ribulose-P 3-epimerase (EC 5.1.3.1) | Transaldolase (EC 2.2.1.2) | Transketolase (EC 2.2.1.1) |
| Micrococcineae | *Brevibacterium* | 6 | 1.00 | 0.83 | 1.83 | 6.67 | 1.00 | 1.67 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 0.50 | 0.00 | 1.50 | 1.00 | 1.00 | 1.67 |
| | *Cellulomonas* | 5 | 1.00 | 1.20 | 1.20 | 2.20 | 1.00 | 2.00 | 1.20 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.80 | 1.20 | 1.20 | 1.20 | 2.00 |
| | *Actinotalea* | 2 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.50 | 1.50 | 1.00 | 2.00 | 1.00 | 1.50 | 1.50 | 1.00 | 1.50 | 1.50 | 1.00 | 1.00 |
| | *Dermabacter* | 2 | 0.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Intrasporangium* | 2 | 1.00 | 1.50 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.50 | 0.50 | 1.50 | 1.00 | 1.00 | 1.00 |
| | *Serinicoccus* | 2 | 0.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 1.50 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Knoellia* | 4 | 1.00 | 1.75 | 1.50 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.25 | 1.25 | 1.25 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Agromyces* | 2 | 1.00 | 1.50 | 1.00 | 2.50 | 1.00 | 1.00 | 1.00 | 1.50 | 2.00 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 | 1.00 | 1.50 | 1.00 |
| | *Clavibacter* | 2 | 1.00 | 1.00 | 1.50 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Leifsonia* | 3 | 1.00 | 1.00 | 1.33 | 2.33 | 1.00 | 1.67 | 1.00 | 1.33 | 2.00 | 1.00 | 2.00 | 2.33 | 1.00 | 1.33 | 1.00 | 1.00 | 1.33 |
| | *Leucobacter* | 4 | 0.75 | 1.00 | 1.00 | 4.50 | 1.00 | 2.00 | 1.00 | 1.50 | 2.00 | 2.00 | 0.75 | 1.00 | 0.75 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Microbacterium* | 18 | 1.00 | 1.22 | 1.06 | 2.83 | 1.00 | 1.33 | 1.00 | 1.44 | 2.00 | 1.00 | 1.72 | 1.33 | 1.06 | 1.61 | 1.11 | 1.50 | 1.44 |
| | *Mycetocola* | 1 | 1.00 | 0.00 | 1.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Arthrobacter* | 10 | 1.00 | 1.70 | 2.00 | 4.50 | 1.10 | 2.00 | 1.00 | 1.00 | 2.00 | 1.40 | 2.00 | 1.80 | 1.00 | 2.50 | 1.20 | 1.30 | 3.00 |
| | *Citricoccus* | 1 | 1.00 | 2.00 | 1.00 | 5.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Kocuria* | 6 | 1.00 | 1.50 | 0.83 | 3.83 | 1.00 | 1.33 | 1.00 | 1.00 | 2.00 | 1.00 | 2.33 | 1.33 | 1.00 | 1.33 | 1.00 | 1.00 | 1.33 |
| | *Rothia* | 4 | 0.00 | 0.00 | 0.00 | 2.00 | 1.00 | 1.25 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Bifidobacteriales | *Bifidobacterium* | 3 | 1.00 | 0.00 | 1.00 | 2.33 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.33 | 1.00 | 1.00 | 1.67 | 1.00 | 1.00 |
| Actinomycineae | *Actinobaculum* | 5 | 0.80 | 0.60 | 0.80 | 2.00 | 1.00 | 1.40 | 1.00 | 1.20 | 1.60 | 0.80 | 1.00 | 1.00 | 0.00 | 1.60 | 1.00 | 1.00 | 1.00 |
| | *Actinomyces* | 21 | 0.76 | 0.29 | 1.10 | 1.90 | 0.95 | 0.90 | 1.00 | 1.29 | 1.67 | 0.86 | 0.86 | 0.90 | 0.95 | 2.19 | 1.00 | 1.43 | 2.00 |
| Kineosporiineae | *Kineococcus* | 1 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 3.00 | 2.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 |
| Propionibacterineae | *Propionibacterium* | 9 | 1.00 | 1.11 | 0.78 | 2.33 | 1.00 | 0.78 | 1.11 | 0.89 | 1.89 | 0.89 | 0.89 | 1.11 | 1.00 | 2.00 | 1.00 | 1.11 | 1.78 |
| | *Actinopolymorpha* | 1 | 1.00 | 2.00 | 3.00 | 4.00 | 2.00 | 4.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 |
| | *Aeromicrobium* | 2 | 1.00 | 0.00 | 1.50 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Pimelobacter* | 1 | 1.00 | 1.00 | 2.00 | 4.00 | 1.00 | 1.00 | 2.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Kribbella* | 1 | 1.00 | 3.00 | 8.00 | 4.00 | 1.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 2.00 | 1.00 |
| | *Marmoricola* | 2 | 1.00 | 1.50 | 2.00 | 3.50 | 1.00 | 1.50 | 2.50 | 1.00 | 2.00 | 1.00 | 2.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 |
| | *Nocardioides* | 7 | 1.00 | 1.43 | 2.00 | 3.00 | 1.00 | 1.57 | 1.29 | 1.14 | 2.00 | 1.00 | 1.86 | 1.14 | 1.00 | 0.86 | 1.14 | 1.14 | 1.43 |
| | *Propionicicella* | 1 | 1.00 | 2.00 | 2.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 3.00 |
| | Mean + Std | 614 | 1.85 | 1.75 | 4.18 | 5.39 | 1.39 | 2.51 | 1.95 | 1.60 | 2.99 | 1.73 | 2.27 | 2.27 | 1.78 | 2.05 | 1.26 | 1.56 | 3.97 |

Part 4 A

| | | | AA from AKG | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | NADP-specific glutamate DH (EC 1.4.1.4) | Glutamate synthase [NADPH] (EC 1.4.1.13) | Glutamine synthetase (EC 6.3.1.2) | Glutamate 5-kinase (EC 2.7.2.11) | γ-glutamyl P reductase (EC 1.2.1.41) | Pyrroline-5-carboxylate reduc (EC 1.5.1.2) | N-acetylglutamate synthase (EC 2.3.1.1) | Acetylglutamate kinase (EC 2.7.2.8) | N-acetyl-γ-glutamyl-P reduc (EC 1.2.1.38) | Acetylornithine AT (EC 2.6.1.11) | Glut N-acetyltransferase (EC 2.3.1.35) | Ornithine carbamoylTF (EC 2.1.3.3) | Argininosuccinate synthase (EC 6.3.4.5) |
| Streptomycineae | *Kitasatospora* | 5 | 0.80 | 3.40 | 4.20 | 1.00 | 1.20 | 1.00 | 2.20 | 1.20 | 1.60 | 1.00 | 1.20 | 1.20 | 1.00 |
| | *Streptacidiphilus* | 9 | 0.11 | 3.00 | 4.22 | 1.00 | 1.00 | 1.00 | 2.33 | 1.22 | 1.44 | 1.11 | 1.22 | 1.33 | 1.11 |
| | *Streptomyces* | 289 | 1.10 | 2.81 | 4.00 | 1.00 | 1.01 | 1.14 | 2.10 | 1.31 | 1.25 | 1.66 | 1.21 | 1.34 | 1.26 |
| Catenulisporineae | *Actinospica* | 2 | 0.50 | 2.50 | 3.50 | 1.00 | 1.00 | 1.50 | 2.00 | 1.00 | 1.00 | 1.50 | 1.00 | 2.00 | 1.00 |
| | *Catenulispora* | 1 | 0.00 | 2.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| Streptosporangineae | *Nocardiopsis* | 9 | 0.89 | 2.22 | 2.44 | 1.00 | 1.00 | 1.22 | 2.56 | 1.00 | 1.33 | 1.11 | 1.78 | 1.11 | 1.00 |
| | *Thermobifida* | 1 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Herbidospora* | 1 | 2.00 | 2.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 |
| | *Streptomonospora* | 1 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Microbispora* | 4 | 1.00 | 2.50 | 2.75 | 0.75 | 0.75 | 0.75 | 1.25 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 | 1.00 |
| | *Microtetraspora* | 1 | 1.00 | 3.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Streptosporangium* | 2 | 2.00 | 2.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 2.50 |
| | *Actinomadura* | 5 | 0.80 | 2.60 | 3.20 | 1.00 | 1.00 | 1.20 | 2.00 | 1.40 | 1.40 | 1.60 | 1.00 | 1.80 | 1.40 |
| | *Actinopolyspora* | 4 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.25 | 1.25 | 1.00 | 1.00 | 1.00 |
| Frankineae | *Frankia* | 9 | 0.89 | 3.56 | 3.11 | 1.00 | 1.00 | 1.00 | 1.11 | 1.00 | 1.00 | 1.78 | 1.00 | 1.00 | 1.00 |
| | *Modestobacter* | 1 | 1.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Blastococcus* | 1 | 0.00 | 5.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| Pseudonocardineae | *Actinoalloteichus* | 3 | 0.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Actinokineospora* | 3 | 0.00 | 4.00 | 4.00 | 1.33 | 1.00 | 1.00 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.33 |
| | *Actinomycetospora* | 1 | 0.00 | 6.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Allokutzneria* | 1 | 1.00 | 2.00 | 3.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 3.00 | 1.00 | 2.00 | 2.00 |
| | *Amycolatopsis* | 20 | 1.40 | 4.10 | 3.30 | 1.00 | 0.20 | 1.65 | 1.00 | 1.05 | 1.10 | 1.15 | 0.95 | 1.50 | 1.55 |
| | *Kibdelosporangium* | 1 | 0.00 | 4.00 | 4.00 | 1.00 | 0.00 | 2.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Kutzneria* | 2 | 2.00 | 4.50 | 3.00 | 1.00 | 1.50 | 1.50 | 2.50 | 1.00 | 1.50 | 2.00 | 1.00 | 1.50 | 2.00 |
| | *Lechevalieria* | 1 | 2.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 2.00 |
| | *Lentzea* | 1 | 2.00 | 2.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| | *Prauserella* | 2 | 2.00 | 4.00 | 2.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Pseudonocardia* | 5 | 1.60 | 6.60 | 4.60 | 1.00 | 1.40 | 1.00 | 1.80 | 1.00 | 1.00 | 1.40 | 1.20 | 1.00 | 1.20 |
| | *Sciscionella* | 2 | 3.00 | 2.00 | 3.50 | 1.00 | 1.00 | 2.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Thermocrispum* | 2 | 2.50 | 3.00 | 2.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Saccharomonospora* | 7 | 2.00 | 3.43 | 2.43 | 1.00 | 0.43 | 1.00 | 2.14 | 1.00 | 1.00 | 1.00 | 1.14 | 1.29 | 1.00 |
| | *Saccharopolyspora* | 4 | 3.00 | 4.75 | 2.75 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.25 | 1.25 | 1.00 | 1.50 | 1.25 |
| | *Saccharothrix* | 2 | 0.50 | 4.50 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.50 | 2.00 | 3.00 | 1.00 | 1.50 | 2.00 |

JJ

Part 4 B

| Suborder | Genera | # of species | AA from AKG | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | NADP-specific glutamate DH (EC 1.4.1.4) | Glutamate synthase [NADPH] (EC 1.4.1.13) | Glutamine synthetase (EC 6.3.1.2) | Glutamate 5-kinase (EC 2.7.2.11) | γ-glutamyl P reductase (EC 1.2.1.41) | Pyrroline-5-carboxylate reduc (EC 1.5.1.2) | N-acetylglutamate synthase (EC 2.3.1.1) | Acetylglutamate kinase (EC 2.7.2.8) | N-acetyl-γ-glutamyl-P reduc (EC 1.2.1.38) | Acetylornithine AT (EC 2.6.1.11) | Glut N-acetyltransferase (EC 2.3.1.35) | Ornithine carbamoylTF (EC 2.1.3.3) | Argininosuccinate synthase (EC 6.3.4.5) |
| Corynebacterineae | *Corynebacterium* | 7 | 1.00 | 0.57 | 2.29 | 1.00 | 1.00 | 1.00 | 0.71 | 1.00 | 1.00 | 1.14 | 1.00 | 1.00 | 1.00 |
| | *Dietzia* | 2 | 2.00 | 3.50 | 3.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Gordonia* | 3 | 2.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Mycobacterium* | 4 | 0.50 | 5.50 | 4.25 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Nocardia* | 28 | 0.82 | 2.79 | 3.14 | 1.00 | 1.00 | 1.04 | 1.29 | 1.00 | 1.07 | 1.39 | 1.04 | 1.04 | 1.04 |
| | *Rhodococcus* | 15 | 1.60 | 5.27 | 3.33 | 0.93 | 0.87 | 1.27 | 1.53 | 0.93 | 1.00 | 1.07 | 1.07 | 0.93 | 1.07 |
| | *Tsukamurella* | 2 | 0.50 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Segniliparus* | 1 | 0.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 |
| Micromonosporineae | *Actinoplanes* | 3 | 0.00 | 3.00 | 3.67 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.33 | 1.00 | 2.33 | 1.00 |
| | *Catenuloplanes* | 1 | 0.00 | 5.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Actinocatenispora* | 1 | 1.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Dactylosporangium* | 1 | 2.00 | 4.00 | 4.00 | 1.00 | 1.00 | 1.00 | 3.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Hamadaea* | 1 | 1.00 | 5.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 |
| | *Longispora* | 1 | 1.00 | 2.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Micromonospora* | 7 | 0.29 | 3.86 | 3.29 | 1.00 | 1.00 | 1.14 | 2.29 | 1.14 | 1.29 | 1.00 | 1.14 | 1.29 | 1.71 |
| | *Pilimelia* | 1 | 0.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Salinispora* | 3 | 0.33 | 4.33 | 2.00 | 1.00 | 1.33 | 1.00 | 2.33 | 1.00 | 1.00 | 1.33 | 1.33 | 1.67 | 1.33 |
| Glycomycineae | *Glycomyces* | 3 | 1.00 | 2.33 | 3.00 | 0.00 | 0.00 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.33 |

KK

Part 4 C

| | | | AA from AKG | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | NADP-specific glutamate DH (EC 1.4.1.4) | Glutamate synthase [NADPH] (EC 1.4.1.13) | Glutamine synthetase (EC 6.3.1.2) | Glutamate 5-kinase (EC 2.7.2.11) | γ-glutamyl P reductase (EC 1.2.1.41) | Pyrroline-5-carboxylate reduc (EC 1.5.1.2) | N-acetylglutamate synthase (EC 2.3.1.1) | Acetylglutamate kinase (EC 2.7.2.8) | N-acetyl-γ-glutamyl-P reduc (EC 1.2.1.38) | Acetylornithine AT (EC 2.6.1.11) | Glut N-acetyltransferase (EC 2.3.1.35) | Omithine carbamoylTF (EC 2.1.3.3) | Argininosuccinate synthase (EC 6.3.4.5) |
| Micrococcineae | *Brevibacterium* | 6 | 1.50 | 2.17 | 2.50 | 1.00 | 1.00 | 1.67 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Cellulomonas* | 5 | 1.60 | 2.60 | 2.20 | 1.00 | 1.00 | 1.00 | 1.80 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Actinotalea* | 2 | 2.00 | 2.50 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Dermabacter* | 2 | 1.00 | 0.00 | 3.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Intrasporangium* | 2 | 1.00 | 4.50 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Serinicoccus* | 2 | 0.00 | 2.00 | 3.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| | *Knoellia* | 4 | 0.50 | 3.00 | 3.00 | 1.00 | 1.00 | 1.00 | 3.25 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Agromyces* | 2 | 2.00 | 3.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Clavibacter* | 2 | 0.00 | 2.50 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 |
| | *Leifsonia* | 3 | 1.00 | 3.67 | 1.67 | 1.00 | 1.00 | 1.00 | 3.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Leucobacter* | 4 | 1.00 | 3.00 | 3.50 | 1.00 | 1.00 | 1.00 | 2.25 | 0.75 | 0.75 | 0.50 | 0.75 | 1.00 | 0.75 |
| | *Microbacterium* | 18 | 0.78 | 2.06 | 2.61 | 1.00 | 1.00 | 1.00 | 2.78 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Mycetocola* | 1 | 1.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Arthrobacter* | 10 | 1.10 | 2.40 | 2.60 | 1.00 | 1.00 | 1.10 | 2.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Citricoccus* | 1 | 2.00 | 3.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Kocuria* | 6 | 1.17 | 2.67 | 2.50 | 0.83 | 0.83 | 0.83 | 2.33 | 1.00 | 1.00 | 1.17 | 1.00 | 1.00 | 0.83 |
| | *Rothia* | 4 | 0.00 | 0.25 | 2.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Bifidobacteriales | *Bifidobacterium* | 3 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.67 | 1.00 | 1.00 | 1.33 | 1.00 | 1.00 | 1.00 |
| Actinomycineae | *Actinobaculum* | 5 | 1.00 | 0.60 | 2.00 | 0.80 | 0.80 | 1.00 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 1.20 | 0.80 |
| | *Actinomyces* | 21 | 1.14 | 0.67 | 2.19 | 0.81 | 0.90 | 0.95 | 1.14 | 0.62 | 0.62 | 0.62 | 0.62 | 1.29 | 0.76 |
| Kineosporiineae | *Kineococcus* | 1 | 0.00 | 2.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Propionibacterineae | *Propionibacterium* | 9 | 0.89 | 2.22 | 2.11 | 1.00 | 1.00 | 1.33 | 0.89 | 0.89 | 0.89 | 1.00 | 0.89 | 1.00 | 1.00 |
| | *Actinopolymorpha* | 1 | 1.00 | 3.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Aeromicrobium* | 2 | 1.50 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 |
| | *Pimelobacter* | 1 | 4.00 | 2.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Kribbella* | 1 | 1.00 | 2.00 | 4.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 |
| | *Marmoricola* | 2 | 1.00 | 2.00 | 3.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Nocardioides* | 7 | 0.14 | 2.43 | 3.14 | 1.00 | 1.00 | 1.14 | 1.14 | 1.00 | 1.00 | 1.00 | 1.00 | 1.43 | 1.00 |
| | *Propionicicella* | 1 | 1.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 |
| | Mean + Std | 614 | 1.86 | 4.12 | 3.69 | 1.14 | 1.28 | 1.36 | 2.47 | 1.31 | 1.52 | 1.69 | 1.25 | 1.61 | 1.46 |

LL

Part 5 A

| | | | AA from OOA | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Aspartate aminotransferase (EC 2.6.1.1) | Asparagine synthetase (EC 6.3.5.4) | Aspartokinase (EC 2.7.2.4) | Aspartate-semialdehyde DH(EC 1.2.1.11) | Homoserine dehydrogenase (EC 1.1.1.3) | Homoserine kinase (EC 2.7.1.39) | Threonine synthase (EC 4.2.3.1) | Dipicolinate synthase SU A (EC 4.2.1.52) | Tetrahydropyridine-diCT STF (EC 2.3.1.117) | N-succinyl-L,L-DAP ATF (EC 2.6.1.17) | N-succ-DAP desuccinylase (EC 3.5.1.18) | Diaminopimelate epimerase (EC 5.1.1.7) | Diaminopimelate deCx (EC 4.1.1.20) | Homoserine O-succinylTF (EC 2.3.1.46) | Cystathionine γ-synthase (EC 2.5.1.48) | O-succinylHS sulfhydrylase (EC 2.5.1.48) | Cystathionine beta-lyase (EC 4.4.1.8) | 5-MTHtriglut-HCys methylTF (EC 2.1.1.14) |
| Streptomycineae | *Kitasatospora* | 5 | 10.20 | 2.40 | 1.00 | 1.40 | 1.20 | 1.80 | 1.80 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.40 | 0.00 | 0.00 | 1.20 | 0.20 | 0.20 |
| | *Streptacidiphilus* | 9 | 9.44 | 2.78 | 1.00 | 1.78 | 1.00 | 2.00 | 2.22 | 0.00 | 1.00 | 0.89 | 1.00 | 1.22 | 4.11 | 0.00 | 0.00 | 0.78 | 1.00 | 0.11 |
| | *Streptomyces* | 289 | 10.93 | 2.67 | 1.17 | 1.92 | 1.33 | 1.67 | 2.76 | 0.00 | 1.01 | 1.09 | 1.07 | 1.08 | 3.57 | 0.06 | 0.02 | 0.48 | 0.45 | 0.80 |
| Catenulisporineae | *Actinospica* | 2 | 9.00 | 1.50 | 1.00 | 1.50 | 1.00 | 1.50 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.50 | 0.00 | 0.00 | 0.50 | 0.50 | 0.50 |
| | *Catenulispora* | 1 | 8.00 | 3.00 | 2.00 | 2.00 | 1.00 | 1.00 | 4.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Streptosporangineae | *Nocardiopsis* | 9 | 6.56 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.78 | 1.00 | 1.00 | 1.44 | 0.00 | 0.00 | 1.33 | 0.22 | 1.00 |
| | *Thermobifida* | 1 | 3.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 |
| | *Herbidospora* | 1 | 13.00 | 1.00 | 2.00 | 1.00 | 2.00 | 1.00 | 2.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 | 3.00 | 1.00 | 1.00 |
| | *Streptomonospora* | 1 | 7.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 2.00 |
| | *Microbispora* | 4 | 6.25 | 1.75 | 0.75 | 0.75 | 0.75 | 1.00 | 1.25 | 0.00 | 0.75 | 1.25 | 0.75 | 0.75 | 1.75 | 0.00 | 0.00 | 1.00 | 0.00 | 0.75 |
| | *Microtetraspora* | 1 | 10.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 3.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 |
| | *Streptosporangium* | 2 | 12.50 | 4.00 | 1.00 | 1.00 | 1.00 | 1.50 | 2.50 | 0.00 | 2.00 | 2.00 | 1.00 | 1.50 | 5.50 | 0.00 | 0.00 | 3.00 | 0.00 | 1.00 |
| | *Actinomadura* | 5 | 13.60 | 1.80 | 1.20 | 1.20 | 1.00 | 1.20 | 2.40 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.60 | 0.00 | 0.00 | 0.60 | 0.00 | 0.80 |
| | *Actinopolyspora* | 4 | 3.50 | 3.25 | 1.25 | 1.00 | 1.00 | 1.00 | 2.75 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 0.00 | 0.00 | 0.25 | 0.00 | 0.75 |
| Frankineae | *Frankia* | 9 | 4.11 | 2.33 | 1.44 | 1.00 | 1.44 | 1.33 | 2.11 | 0.00 | 1.00 | 1.00 | 1.11 | 1.00 | 2.11 | 0.56 | 0.00 | 1.22 | 0.00 | 0.00 |
| | *Modestobacter* | 1 | 10.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| | *Blastococcus* | 1 | 4.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Pseudonocardineae | *Actinoalloteichus* | 3 | 4.00 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.67 |
| | *Actinokineospora* | 3 | 7.33 | 3.67 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.33 | 3.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.33 |
| | *Actinomycetospora* | 1 | 6.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 2.00 | 1.00 | 1.00 |
| | *Allokutzneria* | 1 | 15.00 | 4.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 0.00 | 0.00 | 2.00 | 0.00 | 1.00 |
| | *Amycolatopsis* | 20 | 11.30 | 4.35 | 1.30 | 0.95 | 1.00 | 1.20 | 2.70 | 0.05 | 1.75 | 1.05 | 0.95 | 0.80 | 3.05 | 0.00 | 0.95 | 1.90 | 0.00 | 1.25 |
| | *Kibdelosporangium* | 1 | 22.00 | 8.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 2.00 | 1.00 | 2.00 | 0.00 | 6.00 | 0.00 | 0.00 | 6.00 | 0.00 | 2.00 |
| | *Kutzneria* | 2 | 17.00 | 5.00 | 2.00 | 1.00 | 1.00 | 1.50 | 5.00 | 0.00 | 1.50 | 0.50 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 | 3.00 | 0.50 | 1.00 |
| | *Lechevalieria* | 1 | 10.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 3.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 | 0.00 | 2.00 | 0.00 |
| | *Lentzea* | 1 | 9.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 0.00 | 3.00 | 0.00 |
| | *Prauserella* | 2 | 4.00 | 2.00 | 1.00 | 1.00 | 1.50 | 1.00 | 3.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Pseudonocardia* | 5 | 10.80 | 4.00 | 1.00 | 1.00 | 1.80 | 1.20 | 2.40 | 0.20 | 1.60 | 1.00 | 1.20 | 1.20 | 2.40 | 0.00 | 0.00 | 2.20 | 0.20 | 1.40 |
| | *Sciscionella* | 2 | 10.50 | 3.50 | 1.00 | 1.00 | 1.50 | 1.00 | 3.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 5.00 | 0.00 | 0.00 | 2.00 | 0.50 | 1.50 |
| | *Thermocrispum* | 2 | 3.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.50 | 0.00 | 1.00 |
| | *Saccharomonospora* | 7 | 5.14 | 2.29 | 1.14 | 1.00 | 1.00 | 0.86 | 1.71 | 0.00 | 1.43 | 1.00 | 1.00 | 0.86 | 1.14 | 0.00 | 0.71 | 1.00 | 0.00 | 0.86 |
| | *Saccharopolyspora* | 4 | 7.50 | 2.75 | 1.25 | 1.00 | 1.75 | 1.00 | 3.25 | 0.00 | 1.75 | 1.00 | 1.00 | 1.00 | 1.50 | 0.00 | 0.00 | 0.75 | 0.25 | 1.75 |
| | *Saccharothrix* | 2 | 10.00 | 3.50 | 2.00 | 1.50 | 1.00 | 1.00 | 2.50 | 0.00 | 2.00 | 1.00 | 1.00 | 1.50 | 4.00 | 0.00 | 0.00 | 2.00 | 1.50 | 1.00 |

MM

Part 5 B

| Suborder | Genera | # of species | AA from OOA | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Aspartate aminotransferase (EC 2.6.1.1) | Asparagine synthetase (EC 6.3.5.4) | Aspartokinase (EC 2.7.2.4) | Aspartate-semialdehyde DH(EC 1.2.1.11) | Homoserine dehydrogenase (EC 1.1.1.3) | Homoserine kinase (EC 2.7.1.39) | Threonine synthase (EC 4.2.3.1) | Dipicolinate synthase SU A (EC 4.2.1.52) | Tetrahydropyridine-diCT STF (EC 2.3.1.117) | N-succinyl-L,L-DAP ATF (EC 2.6.1.17) | N-succ-DAP desuccinylase (EC 3.5.1.18) | Diaminopimelate epimerase (EC 5.1.1.7) | Diaminopimelate deCx (EC 4.1.1.20) | Homoserine O-succinylTF (EC 2.3.1.46) | Cystathionine γ-synthase (EC 2.5.1.48) | O-succinylHS sulfhydrylase (EC 2.5.1.48) | Cystathionine beta-lyase (EC 4.4.1.8) | 5-MTHtriglut-HCys methylTF (EC 2.1.1.14) |
| Corynebacterineae | *Corynebacterium* | 7 | 1.86 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.43 | 1.00 | 0.86 | 0.86 | 1.14 | 0.00 | 0.43 | 0.71 | 0.57 | 0.57 |
| | *Dietzia* | 2 | 5.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.50 | 0.00 | 1.50 |
| | *Gordonia* | 3 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.33 | 0.67 | 0.00 | 1.33 | 0.00 | 1.67 |
| | *Mycobacterium* | 4 | 4.75 | 1.75 | 1.00 | 1.00 | 1.50 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 2.00 | 0.50 | 1.00 |
| | *Nocardia* | 28 | 6.64 | 3.57 | 1.00 | 1.00 | 1.21 | 1.00 | 1.18 | 0.00 | 1.25 | 1.07 | 1.00 | 1.00 | 4.36 | 0.07 | 0.00 | 1.21 | 0.00 | 2.25 |
| | *Rhodococcus* | 15 | 6.93 | 1.80 | 1.07 | 0.93 | 1.00 | 0.93 | 1.27 | 0.00 | 1.13 | 0.93 | 1.00 | 0.93 | 1.40 | 0.07 | 0.07 | 2.40 | 0.20 | 1.80 |
| | *Tsukamurella* | 2 | 6.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 |
| | *Segniliparus* | 1 | 4.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 0.00 | 0.00 | 2.00 | 0.00 | 2.00 |
| Micromonosporineae | *Actinoplanes* | 3 | 13.33 | 3.67 | 1.00 | 2.00 | 1.00 | 1.67 | 2.00 | 0.00 | 1.00 | 1.00 | 1.33 | 1.00 | 3.33 | 0.00 | 0.00 | 0.33 | 0.00 | 0.00 |
| | *Catenuloplanes* | 1 | 8.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 3.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| | *Actinocatenispora* | 1 | 11.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Dactylosporangium* | 1 | 10.00 | 6.00 | 1.00 | 2.00 | 1.00 | 3.00 | 2.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Hamadaea* | 1 | 8.00 | 6.00 | 1.00 | 2.00 | 1.00 | 2.00 | 3.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 3.00 | 0.00 | 0.00 | 2.00 | 2.00 | 1.00 |
| | *Longispora* | 1 | 11.00 | 5.00 | 1.00 | 2.00 | 1.00 | 3.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 2.00 | 3.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| | *Micromonospora* | 7 | 5.29 | 4.00 | 1.14 | 2.00 | 1.57 | 1.14 | 3.29 | 0.00 | 1.14 | 1.14 | 1.00 | 1.14 | 3.43 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Pilimelia* | 1 | 6.00 | 2.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 | 1.00 | 0.00 | 2.00 |
| | *Salinispora* | 3 | 3.67 | 4.67 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.33 | 1.00 | 1.00 | 1.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Glycomycineae | *Glycomyces* | 3 | 5.33 | 1.33 | 1.00 | 1.00 | 1.00 | 2.33 | 2.00 | 0.00 | 1.00 | 1.33 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |

NN

Part 5 C

| | | | AA from OOA | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | Aspartate aminotransferase (EC 2.6.1.1) | Asparagine synthetase (EC 6.3.5.4) | Aspartokinase (EC 2.7.2.4) | Aspartate-semialdehyde DH(EC 1.2.1.11) | Homoserine dehydrogenase (EC 1.1.1.3) | Homoserine kinase (EC 2.7.1.39) | Threonine synthase (EC 4.2.3.1) | Dipicolinate synthase SU A (EC 4.2.1.52) | Tetrahydropyridine-diCT STF (EC 2.3.1.117) | N-succinyl-L,L-DAP ATF (EC 2.6.1.17) | N-succ-DAP desuccinylase (EC 3.5.1.18) | Diaminopimelate epimerase (EC 5.1.1.7) | Diaminopimelate deCx (EC 4.1.1.20) | Homoserine O-succinylTF (EC 2.3.1.46) | Cystathionine γ-synthase (EC 2.5.1.48) | O-succinylHS sulfhydrylase (EC 2.5.1.48) | Cystathionine beta-lyase (EC 4.4.1.8) | 5-MTHtriglut-HCys methylTF (EC 2.1.1.14) |
| **Micrococcineae** | ***Brevibacterium*** | 6 | 5.83 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.83 | 1.00 | 0.83 | 1.00 | 1.83 | 0.00 | 0.83 | 2.17 | 0.83 | 0.83 |
| | ***Cellulomonas*** | 5 | 5.00 | 0.40 | 1.00 | 1.00 | 1.00 | 1.40 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.80 | 1.60 | 0.80 | 0.80 |
| | ***Actinotalea*** | 2 | 2.50 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.50 | 1.00 | 0.00 |
| | ***Dermabacter*** | 2 | 3.50 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| | ***Intrasporangium*** | 2 | 4.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.50 | 0.00 |
| | ***Serinicoccus*** | 2 | 8.50 | 0.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 3.00 | 0.50 | 0.00 |
| | ***Knoellia*** | 4 | 3.50 | 1.75 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.50 | 0.00 | 0.00 |
| | ***Agromyces*** | 2 | 8.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.50 | 4.00 | 0.00 | 1.00 |
| | ***Clavibacter*** | 2 | 3.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.50 | 2.00 | 0.50 | 0.00 |
| | ***Leifsonia*** | 3 | 5.00 | 0.33 | 1.00 | 1.00 | 1.00 | 1.00 | 1.33 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.67 | 0.00 | 0.00 | 3.33 | 1.33 | 0.67 |
| | ***Leucobacter*** | 4 | 5.25 | 0.25 | 1.00 | 1.00 | 1.25 | 2.00 | 1.00 | 0.00 | 0.75 | 1.00 | 0.75 | 1.00 | 3.50 | 0.00 | 0.25 | 2.00 | 1.50 | 2.00 |
| | ***Microbacterium*** | 18 | 5.39 | 0.83 | 1.00 | 1.00 | 1.06 | 1.00 | 1.06 | 0.00 | 1.00 | 1.00 | 1.83 | 0.94 | 1.22 | 0.00 | 0.11 | 4.44 | 1.61 | 0.89 |
| | ***Mycetocola*** | 1 | 6.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 3.00 | 1.00 | 1.00 |
| | ***Arthrobacter*** | 10 | 7.10 | 0.30 | 1.10 | 1.00 | 1.70 | 1.30 | 1.10 | 0.00 | 0.90 | 1.00 | 0.90 | 1.00 | 1.30 | 0.00 | 0.90 | 2.30 | 0.30 | 0.80 |
| | ***Citricoccus*** | 1 | 4.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.00 | 0.00 |
| | ***Kocuria*** | 6 | 5.50 | 0.33 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 0.00 | 1.00 | 2.83 | 0.17 | 1.17 |
| | ***Rothia*** | 4 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.00 | 1.00 |
| **Bifidobacteriales** | ***Bifidobacterium*** | 3 | 7.67 | 0.33 | 2.00 | 1.00 | 1.00 | 1.33 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.33 | 2.33 | 1.33 | 1.00 |
| **Actinomycineae** | ***Actinobaculum*** | 5 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.20 | 0.40 | 0.20 | 0.20 | 1.00 | 0.00 | 0.40 | 0.80 | 1.00 | 0.40 |
| | ***Actinomyces*** | 21 | 2.29 | 0.52 | 0.95 | 0.90 | 0.95 | 0.90 | 0.95 | 0.00 | 0.76 | 0.86 | 0.52 | 0.76 | 0.86 | 0.00 | 0.38 | 0.76 | 1.05 | 0.81 |
| **Kineosporiineae** | ***Kineococcus*** | 1 | 6.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 4.00 | 0.00 | 1.00 |
| **Propionibacterineae** | ***Propionibacterium*** | 9 | 2.33 | 0.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.11 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.11 | 0.00 | 0.78 | 0.11 | 0.67 | 0.11 |
| | ***Actinopolymorpha*** | 1 | 9.00 | 2.00 | 2.00 | 2.00 | 2.00 | 7.00 | 4.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 |
| | ***Aeromicrobium*** | 2 | 3.50 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.50 | 0.50 | 0.00 | 0.00 |
| | ***Pimelobacter*** | 1 | 7.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| | ***Kribbella*** | 1 | 14.00 | 1.00 | 1.00 | 2.00 | 1.00 | 3.00 | 3.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | ***Marmoricola*** | 2 | 4.50 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 3.50 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.50 |
| | ***Nocardioides*** | 7 | 5.14 | 0.29 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.14 | 0.00 | 0.29 | 0.29 | 0.00 | 0.00 |
| | ***Propionicicella*** | 1 | 5.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 4.00 | 1.00 | 0.00 |
| | **Mean + Std** | 614 | 10.88 | 3.68 | 1.52 | 1.57 | 1.52 | 2.11 | 2.80 | 0.03 | 1.55 | 1.50 | 1.25 | 1.23 | 3.35 | 0.18 | 0.67 | 2.70 | 1.07 | 0.00 |

OO

Part 6 A

| Suborder | Genera | # of species | AAs from 3PGA. | | | | | | | Amino acids from R5P | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 5-MtetraHF-HCys methylTF (EC 2.1.1.13) | D-3-PG dehydrogenase (EC 1.1.1.95) | Phosphoserine ATF (EC 2.6.1.52) | Phosphoserine phosphatase (EC 3.1.3.3) | Serine HTF (EC 2.1.2.1) | Serine acetyltransferase (EC 2.3.1.30) | Cysteine synthase (EC 2.5.1.47) | Ribose-P pyroPK (EC 2.7.6.1) | ATP phosphoribosylTF (EC 2.4.2.17) | Pribosyl-ATP pyroPase (EC 3.6.1.31) | Pribosyl-AMP cyclohydrolase (EC 3.5.4.19) | PRF-5-Aimid CA ribot isom (EC 5.3.1.16) | Glutamine amidotransferase (EC 2.4.2.-) | Imidazoleglycerol-P DH (EC 4.2.1.19) | Histidinol-phosphate ATF (EC 2.6.1.9) | Histidinol-phosphatase (EC 3.1.3.15) | Histidinol dehydrogenase (EC 1.1.1.23) |
| Streptomycineae | *Kitasatospora* | 5 | 1.20 | 4.40 | 1.00 | 4.00 | 2.60 | 1.00 | 2.80 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.80 | 1.00 |
| | *Streptacidiphilus* | 9 | 0.89 | 5.56 | 1.00 | 3.78 | 2.33 | 0.22 | 2.56 | 1.78 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.33 | 1.33 | 1.22 |
| | *Streptomyces* | 289 | 1.70 | 6.40 | 1.02 | 3.50 | 2.94 | 0.19 | 3.96 | 1.25 | 1.03 | 1.00 | 0.90 | 1.00 | 0.01 | 1.00 | 1.11 | 1.47 | 1.16 |
| Catenulisporineae | *Actinospica* | 2 | 1.00 | 5.50 | 1.00 | 3.50 | 2.50 | 0.00 | 4.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.50 | 1.00 |
| | *Catenulispora* | 1 | 0.00 | 10.00 | 1.00 | 5.00 | 1.00 | 0.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 |
| Streptosporangineae | *Nocardiopsis* | 9 | 1.22 | 4.11 | 1.11 | 2.11 | 1.33 | 0.33 | 3.89 | 1.11 | 1.00 | 1.00 | 0.89 | 1.00 | 0.00 | 1.00 | 1.22 | 1.11 | 1.00 |
| | *Thermobifida* | 1 | 1.00 | 3.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Herbidospora* | 1 | 2.00 | 6.00 | 1.00 | 4.00 | 1.00 | 0.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 2.00 | 1.00 |
| | *Streptomonospora* | 1 | 1.00 | 6.00 | 1.00 | 3.00 | 1.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Microbispora* | 4 | 1.00 | 4.50 | 0.75 | 2.00 | 1.25 | 0.50 | 2.50 | 1.50 | 0.75 | 0.75 | 0.75 | 0.75 | 0.00 | 0.75 | 0.75 | 1.50 | 0.75 |
| | *Microtetraspora* | 1 | 2.00 | 5.00 | 1.00 | 4.00 | 3.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Streptosporangium* | 2 | 2.50 | 6.00 | 1.00 | 2.00 | 3.50 | 0.00 | 5.50 | 2.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Actinomadura* | 5 | 1.20 | 5.80 | 1.00 | 2.60 | 2.40 | 0.00 | 3.00 | 1.00 | 1.00 | 0.20 | 1.00 | 1.00 | 0.40 | 1.00 | 1.00 | 1.40 | 1.60 |
| | *Actinopolyspora* | 4 | 0.75 | 6.00 | 1.00 | 3.00 | 1.00 | 0.00 | 4.25 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Frankineae | *Frankia* | 9 | 1.11 | 2.00 | 1.00 | 2.56 | 2.11 | 1.67 | 2.22 | 1.78 | 1.00 | 1.00 | 1.00 | 1.00 | 0.11 | 1.00 | 1.56 | 1.56 | 1.00 |
| | *Modestobacter* | 1 | 1.00 | 2.00 | 1.00 | 3.00 | 3.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 |
| | *Blastococcus* | 1 | 0.00 | 2.00 | 1.00 | 3.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 |
| Pseudonocardineae | *Actinoalloteichus* | 3 | 1.00 | 8.67 | 1.00 | 3.00 | 2.00 | 0.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Actinokineospora* | 3 | 1.00 | 4.00 | 1.00 | 4.33 | 2.67 | 1.00 | 3.67 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Actinomycetospora* | 1 | 2.00 | 5.00 | 1.00 | 4.00 | 2.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Allokutzneria* | 1 | 2.00 | 5.00 | 1.00 | 6.00 | 2.00 | 0.00 | 4.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 2.00 | 1.00 |
| | *Amycolatopsis* | 20 | 2.15 | 7.35 | 0.95 | 3.20 | 2.15 | 0.15 | 3.75 | 1.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.05 | 0.95 | 1.10 | 1.35 | 1.20 |
| | *Kibdelosporangium* | 1 | 1.00 | 6.00 | 1.00 | 3.00 | 3.00 | 0.00 | 8.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Kutzneria* | 2 | 1.50 | 7.50 | 1.00 | 3.50 | 2.50 | 0.00 | 5.50 | 2.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Lechevalieria* | 1 | 2.00 | 6.00 | 1.00 | 4.00 | 3.00 | 0.00 | 4.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 |
| | *Lentzea* | 1 | 2.00 | 4.00 | 1.00 | 3.00 | 2.00 | 1.00 | 5.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 |
| | *Prauserella* | 2 | 0.00 | 8.50 | 1.00 | 3.00 | 2.00 | 0.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 3.00 |
| | *Pseudonocardia* | 5 | 2.00 | 6.60 | 1.20 | 4.60 | 2.80 | 0.40 | 2.20 | 2.00 | 1.20 | 1.20 | 1.20 | 1.20 | 1.40 | 1.20 | 1.60 | 1.20 | 1.60 |
| | *Sciscionella* | 2 | 1.00 | 7.50 | 1.00 | 3.00 | 1.50 | 0.00 | 4.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.50 |
| | *Thermocrispum* | 2 | 0.50 | 4.50 | 1.00 | 3.00 | 1.00 | 0.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 | 1.00 | 2.00 |
| | *Saccharomonospora* | 7 | 1.00 | 6.00 | 1.00 | 2.71 | 1.86 | 0.00 | 3.00 | 1.00 | 0.86 | 0.86 | 0.86 | 0.86 | 0.00 | 0.86 | 1.00 | 1.00 | 1.14 |
| | *Saccharopolyspora* | 4 | 1.25 | 8.50 | 1.00 | 3.25 | 0.75 | 0.75 | 5.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.75 | 1.00 | 1.50 |
| | *Saccharothrix* | 2 | 1.50 | 4.50 | 1.50 | 4.00 | 3.00 | 0.00 | 4.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 3.00 | 1.00 |

PP

Part 6 B

| | | | AAs from 3PGA. | | | | | | | Amino acids from R5P | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | 5-MtetraHF-HCys methylTF (EC 2.1.1.13) | D-3-PG dehydrogenase (EC 1.1.1.95) | Phosphoserine ATF (EC 2.6.1.52) | Phosphoserine phosphatase (EC 3.1.3.3) | Serine HTF (EC 2.1.2.1) | Serine acetyltransferase (EC 2.3.1.30) | Cysteine synthase (EC 2.5.1.47) | Ribose-P pyroPK (EC 2.7.6.1) | ATP phosphoribosylTF (EC 2.4.2.17) | Pribosyl-ATP pyroPase (EC 3.6.1.31) | Pribosyl-AMP cyclohydrolase (EC 3.5.4.19) | PRF-5-Aimid CA ribot isom (EC 5.3.1.16) | Glutamine amidotransferase (EC 2.4.2.-) | Imidazoleglycerol-P DH (EC 4.2.1.19) | Histidinol-phosphate ATF (EC 2.6.1.9) | Histidinol-phosphatase (EC 3.1.3.15) | Histidinol dehydrogenase (EC 1.1.1.23) |
| Corynebacterineae | *Corynebacterium* | 7 | 0.57 | 2.00 | 1.00 | 3.14 | 1.00 | 1.00 | 1.43 | 1.00 | 0.86 | 0.86 | 0.86 | 0.86 | 0.00 | 0.86 | 0.86 | 1.71 | 0.86 |
| | *Dietzia* | 2 | 1.00 | 3.00 | 1.00 | 3.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Gordonia* | 3 | 1.00 | 2.33 | 1.00 | 3.33 | 1.67 | 0.67 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.33 | 1.00 |
| | *Mycobacterium* | 4 | 1.50 | 4.50 | 1.00 | 4.25 | 1.50 | 0.50 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 | 2.00 | 2.25 |
| | *Nocardia* | 28 | 1.18 | 4.18 | 1.00 | 5.21 | 1.32 | 0.11 | 6.00 | 1.29 | 1.04 | 1.00 | 1.00 | 1.00 | 0.25 | 1.00 | 1.21 | 3.29 | 1.36 |
| | *Rhodococcus* | 15 | 1.33 | 5.33 | 0.93 | 3.73 | 1.87 | 0.73 | 2.27 | 1.13 | 0.93 | 0.93 | 0.93 | 0.93 | 0.53 | 0.93 | 0.93 | 2.60 | 1.27 |
| | *Tsukamurella* | 2 | 1.00 | 3.50 | 1.00 | 3.00 | 1.00 | 0.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Segniliparus* | 1 | 0.00 | 2.00 | 1.00 | 3.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Micromonosporineae | *Actinoplanes* | 3 | 1.00 | 4.33 | 1.00 | 3.00 | 2.00 | 0.33 | 2.67 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.67 | 1.00 |
| | *Catenuloplanes* | 1 | 2.00 | 8.00 | 1.00 | 2.00 | 3.00 | 0.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 |
| | *Actinocatenispora* | 1 | 0.00 | 9.00 | 1.00 | 3.00 | 2.00 | 0.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Dactylosporangium* | 1 | 0.00 | 6.00 | 1.00 | 4.00 | 3.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 4.00 | 1.00 |
| | *Hamadaea* | 1 | 2.00 | 4.00 | 1.00 | 4.00 | 1.00 | 0.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Longispora* | 1 | 1.00 | 3.00 | 1.00 | 2.00 | 1.00 | 0.00 | 4.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Micromonospora* | 7 | 0.00 | 3.14 | 1.00 | 2.00 | 2.29 | 0.00 | 3.00 | 2.00 | 1.00 | 1.14 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.14 | 1.00 |
| | *Pilimelia* | 1 | 1.00 | 3.00 | 1.00 | 6.00 | 1.00 | 0.00 | 8.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 3.00 | 1.00 |
| | *Salinispora* | 3 | 0.00 | 2.67 | 1.00 | 2.00 | 1.33 | 0.00 | 3.67 | 2.00 | 1.33 | 1.33 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 |
| Glycomycineae | *Glycomyces* | 3 | 1.00 | 5.00 | 1.00 | 2.33 | 1.00 | 0.00 | 2.33 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |

QQ

Part 6 C

| | | | AAs from 3PGA. | | | | | | | Amino acids from R5P | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | 5-MtetraHF-HCys methyITF (EC 2.1.1.13) | D-3-PG dehydrogenase (EC 1.1.1.95) | Phosphoserine ATF (EC 2.6.1.52) | Phosphoserine phosphatase (EC 3.1.3.3) | Serine HTF (EC 2.1.2.1) | Serine acetyltransferase (EC 2.3.1.30) | Cysteine synthase (EC 2.5.1.47) | Ribose-P pyroPK (EC 2.7.6.1) | ATP phosphoribosyITF (EC 2.4.2.17) | Pribosyl-ATP pyroPase (EC 3.6.1.31) | Pribosyl-AMP cyclohydrolase (EC 3.5.4.19) | PRF-5-Aimid CA ribot isom (EC 5.3.1.16) | Glutamine amidotransferase (EC 2.4.2.-) | Imidazoleglycerol-P DH (EC 4.2.1.19) | Histidinol-phosphate ATF (EC 2.6.1.9) | Histidinol-phosphatase (EC 3.1.3.15) | Histidinol dehydrogenase (EC 1.1.1.23) |
| Micrococcineae | *Brevibacterium* | 6 | 0.00 | 4.83 | 1.00 | 1.33 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.83 | 1.00 | 0.00 | 1.00 | 2.33 | 1.00 | 1.33 |
| | *Cellulomonas* | 5 | 0.20 | 2.80 | 1.00 | 2.60 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.40 | 1.20 |
| | *Actinotalea* | 2 | 0.50 | 2.50 | 1.00 | 2.00 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Dermabacter* | 2 | 0.00 | 1.50 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Intrasporangium* | 2 | 1.00 | 3.50 | 1.00 | 2.00 | 1.50 | 1.00 | 1.50 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.50 | 1.00 |
| | *Serinicoccus* | 2 | 2.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 1.00 | 1.00 | 2.00 | 1.00 |
| | *Knoellia* | 4 | 1.00 | 2.75 | 1.00 | 2.00 | 1.00 | 1.50 | 2.25 | 2.00 | 1.00 | 0.75 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Agromyces* | 2 | 0.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.50 | 1.00 | 1.00 |
| | *Clavibacter* | 2 | 0.00 | 2.50 | 1.00 | 2.00 | 1.50 | 2.00 | 0.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Leifsonia* | 3 | 0.00 | 4.33 | 1.00 | 1.33 | 1.00 | 1.00 | 1.33 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Leucobacter* | 4 | 0.00 | 4.75 | 1.00 | 1.25 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Microbacterium* | 18 | 0.11 | 4.28 | 1.06 | 1.06 | 1.00 | 1.22 | 1.61 | 1.06 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.11 | 1.06 | 1.00 |
| | *Mycetocola* | 1 | 0.00 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 |
| | *Arthrobacter* | 10 | 0.30 | 6.40 | 1.00 | 1.90 | 2.60 | 1.00 | 2.30 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.20 | 1.00 | 1.00 | 1.30 | 1.10 |
| | *Citricoccus* | 1 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Kocuria* | 6 | 0.00 | 4.67 | 1.00 | 1.67 | 1.50 | 2.00 | 1.33 | 1.00 | 1.00 | 1.00 | 0.83 | 1.00 | 0.00 | 1.00 | 1.17 | 1.50 | 1.00 |
| | *Rothia* | 4 | 0.00 | 0.75 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Bifidobacteriales | *Bifidobacterium* | 3 | 0.00 | 2.33 | 1.00 | 1.00 | 0.33 | 0.00 | 0.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 1.00 |
| Actinomycineae | *Actinobaculum* | 5 | 0.00 | 1.80 | 0.20 | 0.20 | 1.00 | 1.00 | 1.00 | 1.00 | 0.40 | 0.40 | 0.40 | 1.00 | 0.00 | 0.40 | 0.40 | 0.40 | 0.40 |
| | *Actinomyces* | 21 | 0.00 | 1.43 | 0.86 | 1.57 | 0.95 | 1.00 | 1.05 | 1.00 | 0.62 | 0.67 | 0.67 | 0.95 | 0.00 | 0.67 | 0.67 | 0.67 | 0.67 |
| Kineosporiineae | *Kineococcus* | 1 | 0.00 | 6.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 2.00 |
| Propionibacterineae | *Propionibacterium* | 9 | 1.00 | 3.56 | 1.00 | 1.22 | 1.00 | 1.22 | 1.44 | 1.22 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.11 | 1.00 |
| | *Actinopolymorpha* | 1 | 2.00 | 14.00 | 1.00 | 4.00 | 1.00 | 0.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 6.00 | 1.00 |
| | *Aeromicrobium* | 2 | 0.50 | 2.00 | 1.00 | 3.00 | 1.00 | 0.50 | 2.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.50 | 1.00 | 1.00 |
| | *Pimelobacter* | 1 | 1.00 | 5.00 | 1.00 | 3.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 1.00 | 1.00 |
| | *Kribbella* | 1 | 0.00 | 12.00 | 1.00 | 4.00 | 2.00 | 0.00 | 5.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Marmoricola* | 2 | 4.00 | 4.00 | 1.00 | 3.00 | 1.50 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | *Nocardioides* | 7 | 0.29 | 3.57 | 1.00 | 3.86 | 1.00 | 0.29 | 1.43 | 1.14 | 1.00 | 1.00 | 1.00 | 1.00 | 0.14 | 1.00 | 1.29 | 1.43 | 1.00 |
| | *Propionicicella* | 1 | 1.00 | 4.00 | 1.00 | 2.00 | 1.00 | 3.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Mean + Std | 614 | 1.70 | 7.10 | 1.16 | 4.02 | 2.37 | 1.21 | 4.27 | 1.94 | 1.19 | 1.15 | 1.15 | 1.13 | 0.42 | 1.13 | 1.58 | 2.52 | 1.46 |

RR

Part 7 A

| Suborder | Genera | # of species | Amino acids from E4P and PEP | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | deoxy-D-arabino 7-P synth (EC 2.5.1.54) | 3-dehydroquinate synthase (EC 4.2.3.4) | 3-dehydroquinate dehydr (EC 4.2.1.10) | Shikimate 5-dehydrogenase (EC 1.1.1.25) | Shikimate kinase I (EC 2.7.1.71) | 5-pyruvylshikimate-3-P synt (EC 2.5.1.19) | Chorismate synthase (EC 4.2.3.5) | Chorismate mutase (EC 5.4.99.5) | Anthranilate PribosylTF (EC 2.4.2.18) | Pribosylanthranilate isom (EC 5.3.1.24) | Pribosylanthranilate isom (EC 5.3.1.24) | Indole-3-glycerol P synth (EC 4.1.1.48) | Diaminopimelate DCx (EC 4.1.1.20) | Chexadienyl DH (EC 1.3.1.12)(EC 1.3.1.43) | Preph/arog DH (EC 1.3.1.12/1.3.1.43) | Aspartate aminotransferase (EC 2.6.1.1) | Chexadienyl dehydr (EC 4.2.1.51/4.2.1.91) | Prephenate dehydratase (EC 4.2.1.51) |
| Streptomycineae | *Kitasatospora* | 5 | 2.20 | 1.40 | 1.20 | 1.40 | 1.00 | 2.00 | 1.00 | 1.00 | 0.80 | 0.80 | 0.40 | 1.00 | 3.40 | 0.00 | 0.20 | 10.20 | 0.00 | 1.00 |
| | *Streptacidiphilus* | 9 | 1.22 | 1.22 | 1.00 | 1.33 | 1.00 | 1.89 | 1.11 | 2.00 | 1.11 | 1.00 | 0.00 | 1.00 | 4.11 | 0.00 | 0.00 | 9.44 | 0.00 | 1.00 |
| | *Streptomyces* | 289 | 1.95 | 1.37 | 1.17 | 1.33 | 1.01 | 1.49 | 1.06 | 2.66 | 1.28 | 0.98 | 0.24 | 1.17 | 3.57 | 0.00 | 0.03 | 10.93 | 0.10 | 1.11 |
| Catenulisporineae | *Actinospica* | 2 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 | 3.00 | 1.00 | 1.00 | 0.00 | 1.50 | 3.50 | 0.00 | 0.00 | 9.00 | 0.00 | 1.50 |
| | *Catenulispora* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 0.00 | 2.00 | 4.00 | 0.00 | 0.00 | 8.00 | 0.00 | 1.00 |
| Streptosporangineae | *Nocardiopsis* | 9 | 3.44 | 1.00 | 1.00 | 1.22 | 1.00 | 1.11 | 1.00 | 1.00 | 1.11 | 1.00 | 0.22 | 1.11 | 1.44 | 0.00 | 0.00 | 6.56 | 0.33 | 1.00 |
| | *Thermobifida* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 3.00 | 0.00 | 1.00 |
| | *Herbidospora* | 1 | 3.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 4.00 | 0.00 | 0.00 | 13.00 | 0.00 | 1.00 |
| | *Streptomonospora* | 1 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 7.00 | 0.00 | 1.00 |
| | *Microbispora* | 4 | 2.75 | 0.75 | 0.75 | 1.00 | 1.25 | 1.00 | 0.75 | 0.75 | 1.00 | 0.75 | 0.75 | 0.75 | 1.75 | 0.00 | 0.25 | 6.25 | 0.00 | 1.25 |
| | *Microtetraspora* | 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 0.00 | 0.00 | 10.00 | 0.00 | 1.00 |
| | *Streptosporangium* | 2 | 5.00 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 | 1.50 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 5.50 | 0.00 | 0.00 | 12.50 | 0.00 | 1.00 |
| | *Actinomadura* | 5 | 3.40 | 1.40 | 1.60 | 1.40 | 1.00 | 1.00 | 1.00 | 1.20 | 1.00 | 1.00 | 0.00 | 1.00 | 2.60 | 0.00 | 0.00 | 13.60 | 0.00 | 1.00 |
| | *Actinopolyspora* | 4 | 2.75 | 1.00 | 1.00 | 1.25 | 1.00 | 0.75 | 1.00 | 1.00 | 1.00 | 0.75 | 0.00 | 1.25 | 1.50 | 0.00 | 0.00 | 3.50 | 0.00 | 0.75 |
| Frankineae | *Frankia* | 9 | 1.11 | 1.89 | 1.00 | 1.78 | 1.00 | 1.00 | 1.00 | 1.78 | 1.11 | 1.00 | 0.00 | 1.11 | 2.11 | 0.00 | 0.00 | 4.11 | 0.00 | 2.00 |
| | *Modestobacter* | 1 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 10.00 | 0.00 | 1.00 |
| | *Blastococcus* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 4.00 | 0.00 | 0.00 |
| Pseudonocardineae | *Actinoalloteichus* | 3 | 3.00 | 2.00 | 2.00 | 1.00 | 1.00 | 0.67 | 1.00 | 0.33 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 4.00 | 0.00 | 1.00 |
| | *Actinokineospora* | 3 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.33 | 1.33 | 1.00 | 1.33 | 1.33 | 3.00 | 0.00 | 0.00 | 7.33 | 0.00 | 1.00 |
| | *Actinomycetospora* | 1 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 6.00 | 0.00 | 1.00 |
| | *Allokutzneria* | 1 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 0.00 | 0.00 | 15.00 | 1.00 | 1.00 |
| | *Amycolatopsis* | 20 | 3.75 | 1.55 | 1.25 | 1.70 | 0.95 | 1.40 | 0.95 | 0.95 | 0.95 | 0.95 | 0.15 | 1.00 | 3.05 | 0.00 | 0.00 | 11.30 | 0.35 | 0.95 |
| | *Kibdelosporangium* | 1 | 6.00 | 2.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 1.00 | 3.00 | 4.00 | 6.00 | 0.00 | 0.00 | 22.00 | 0.00 | 1.00 |
| | *Kutzneria* | 2 | 4.00 | 1.50 | 1.50 | 1.00 | 1.00 | 1.50 | 1.00 | 3.00 | 0.00 | 0.00 | 1.00 | 1.00 | 4.00 | 0.50 | 0.00 | 17.00 | 0.50 | 1.00 |
| | *Lechevalieria* | 1 | 5.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 4.00 | 0.00 | 0.00 | 10.00 | 0.00 | 1.00 |
| | *Lentzea* | 1 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 2.00 | 2.00 | 2.00 | 0.00 | 0.00 | 9.00 | 0.00 | 1.00 |
| | *Prauserella* | 2 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.00 | 0.00 | 4.00 | 0.00 | 1.00 |
| | *Pseudonocardia* | 5 | 2.60 | 1.00 | 1.40 | 1.80 | 1.00 | 1.00 | 1.00 | 2.00 | 1.40 | 1.20 | 0.20 | 1.80 | 2.40 | 0.00 | 0.00 | 10.80 | 0.60 | 1.20 |
| | *Sciscionella* | 2 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 5.00 | 0.00 | 0.00 | 10.50 | 1.00 | 1.00 |
| | *Thermocrispum* | 2 | 1.00 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 3.00 | 0.00 | 1.00 |
| | *Saccharomonospora* | 7 | 2.14 | 0.86 | 1.14 | 1.00 | 0.86 | 0.86 | 0.86 | 1.00 | 0.86 | 0.86 | 0.00 | 1.00 | 1.14 | 0.00 | 0.00 | 5.14 | 0.00 | 0.86 |
| | *Saccharopolyspora* | 4 | 3.25 | 1.50 | 1.50 | 1.75 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 1.00 | 0.75 | 1.25 | 1.50 | 0.00 | 0.00 | 7.50 | 1.00 | 1.00 |
| | *Saccharothrix* | 2 | 5.00 | 1.00 | 1.00 | 1.50 | 1.00 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 0.00 | 1.50 | 10.00 | 0.00 | 1.00 |

SS

Part 7 B

| | | | Amino acids from E4P and PEP | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suborder | Genera | # of species | deoxy-D-arabino 7-P synth (EC 2.5.1.54) | 3-dehydroquinate synthase (EC 4.2.3.4) | 3-dehydroquinate dehydr (EC 4.2.1.10) | Shikimate 5-dehydrogenase (EC 1.1.1.25) | Shikimate kinase I (EC 2.7.1.71) | 5-pyruvylshikimate-3-P synt (EC 2.5.1.19) | Chorismate synthase (EC 4.2.3.5) | Chorismate mutase (EC 5.4.99.5) | Anthranilate PribosylTF (EC 2.4.2.18) | Pribosylanthranilate isom (EC 5.3.1.24) | Pribosylanthranilate isom (EC 5.3.1.24) | Indole-3-glycerol P synth (EC 4.1.1.48) | Diaminopimelate DCx (EC 4.1.1.20) | Chexadienyl DH (EC 1.3.1.12)(EC 1.3.1.43) | Preph/arog DH (EC 1.3.1.12/1.3.1.43) | Aspartate aminotransferase (EC 2.6.1.1) | Chexadienyl dehydr (EC 4.2.1.51/4.2.1.91) | Prephenate dehydratase (EC 4.2.1.51) |
| Corynebacterineae | *Corynebacterium* | 7 | 1.14 | 1.00 | 1.00 | 1.86 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.14 | 0.86 | 2.00 | 1.14 | 0.00 | 0.00 | 1.86 | 0.00 | 1.00 |
| | *Dietzia* | 2 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 | 0.00 | 0.00 | 5.00 | 0.00 | 1.00 |
| | *Gordonia* | 3 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.67 | 1.00 | 1.33 | 0.00 | 0.00 | 4.00 | 0.00 | 1.00 |
| | *Mycobacterium* | 4 | 1.50 | 1.00 | 1.25 | 1.75 | 1.00 | 1.00 | 1.00 | 1.00 | 0.25 | 0.25 | 0.00 | 1.00 | 2.00 | 0.00 | 0.00 | 4.75 | 0.50 | 1.00 |
| | *Nocardia* | 28 | 2.68 | 1.25 | 1.00 | 1.07 | 0.89 | 1.00 | 1.00 | 1.04 | 1.00 | 1.00 | 0.36 | 1.00 | 4.36 | 0.00 | 0.00 | 6.64 | 0.46 | 1.00 |
| | *Rhodococcus* | 15 | 1.93 | 1.20 | 1.33 | 3.20 | 0.93 | 0.87 | 0.93 | 1.47 | 1.00 | 0.93 | 0.80 | 1.00 | 1.40 | 0.07 | 0.27 | 6.93 | 0.00 | 0.93 |
| | *Tsukamurella* | 2 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.50 | 0.00 | 0.00 | 6.00 | 0.00 | 1.00 |
| | *Segniliparus* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 3.00 | 0.00 | 0.00 | 4.00 | 0.00 | 1.00 |
| Micromonosporineae | *Actinoplanes* | 3 | 3.00 | 1.67 | 1.00 | 2.33 | 1.00 | 1.00 | 1.00 | 1.33 | 1.00 | 1.00 | 0.00 | 1.67 | 3.33 | 0.00 | 0.00 | 13.33 | 0.00 | 1.00 |
| | *Catenuloplanes* | 1 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.00 | 0.00 | 8.00 | 0.00 | 1.00 |
| | *Actinocatenispora* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 11.00 | 0.00 | 1.00 |
| | *Dactylosporangium* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.00 | 0.00 | 10.00 | 0.00 | 1.00 |
| | *Hamadaea* | 1 | 2.00 | 3.00 | 2.00 | 3.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 0.00 | 1.00 | 3.00 | 0.00 | 0.00 | 8.00 | 0.00 | 1.00 |
| | *Longispora* | 1 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 3.00 | 0.00 | 0.00 | 11.00 | 0.00 | 1.00 |
| | *Micromonospora* | 7 | 2.29 | 1.57 | 1.43 | 0.86 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 3.43 | 0.00 | 0.00 | 5.29 | 0.00 | 1.00 |
| | *Pilimelia* | 1 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 4.00 | 0.00 | 0.00 | 6.00 | 0.00 | 1.00 |
| | *Salinispora* | 3 | 3.00 | 2.33 | 1.33 | 1.33 | 1.00 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.33 | 0.00 | 0.00 | 3.67 | 0.00 | 1.00 |
| Glycomycineae | *Glycomyces* | 3 | 1.67 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.00 | 0.00 | 5.33 | 0.00 | 1.00 |

Part 7 C

| Suborder | Genera | # of species | Amino acids from E4P and PEP | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | deoxy-D-arabino 7-P synth (EC 2.5.1.54) | 3-dehydroquinate synthase (EC 4.2.3.4) | 3-dehydroquinate dehydr (EC 4.2.1.10) | Shikimate 5-dehydrogenase (EC 1.1.1.25) | Shikimate kinase I (EC 2.7.1.71) | 5-pyruvylshikimate-3-P synt (EC 2.5.1.19) | Chorismate synthase (EC 4.2.3.5) | Chorismate mutase (EC 5.4.99.5) | Anthranilate PribosylTF (EC 2.4.2.18) | Pribosylanthranilate isom (EC 5.3.1.24) | Pribosylanthranilate isom (EC 5.3.1.24) | Indole-3-glycerol P synth (EC 4.1.1.48) | Diaminopimelate DCx (EC 4.1.1.20) | Chexadienyl DH (EC 1.3.1.12)(EC 1.3.1.43) | Preph/arog DH (EC 1.3.1.12/1.3.1.43) | Aspartate aminotransferase (EC 2.6.1.1) | Chexadienyl dehydr (EC 4.2.1.51/4.2.1.91) | Prephenate dehydratase (EC 4.2.1.51) |
| Micrococcineae | *Brevibacterium* | 6 | 1.00 | 1.00 | 1.33 | 1.17 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.17 | 1.00 | 1.83 | 0.00 | 0.00 | 5.83 | 0.00 | 1.00 |
| | *Cellulomonas* | 5 | 1.00 | 1.00 | 1.00 | 1.20 | 1.00 | 1.20 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 5.00 | 0.00 | 1.00 |
| | *Actinotalea* | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 2.50 | 0.00 | 1.00 |
| | *Dermabacter* | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.50 | 0.00 | 0.00 | 3.50 | 0.00 | 1.00 |
| | *Intrasporangium* | 2 | 1.00 | 1.00 | 1.50 | 1.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 4.50 | 0.00 | 1.00 |
| | *Serinicoccus* | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 8.50 | 0.50 | 1.00 |
| | *Knoellia* | 4 | 1.00 | 1.00 | 1.00 | 1.75 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 3.50 | 0.00 | 1.00 |
| | *Agromyces* | 2 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 2.00 | 0.00 | 0.00 | 8.00 | 0.00 | 1.00 |
| | *Clavibacter* | 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 3.00 | 0.00 | 1.00 |
| | *Leifsonia* | 3 | 1.00 | 1.00 | 1.00 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.33 | 1.00 | 1.67 | 0.00 | 0.00 | 5.00 | 0.00 | 1.00 |
| | *Leucobacter* | 4 | 1.00 | 1.00 | 0.50 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.50 | 0.00 | 0.00 | 5.25 | 0.00 | 1.00 |
| | *Microbacterium* | 18 | 1.11 | 1.00 | 1.00 | 1.67 | 1.00 | 1.00 | 1.06 | 1.00 | 1.00 | 0.89 | 0.00 | 1.00 | 1.22 | 0.00 | 0.00 | 5.39 | 0.00 | 1.06 |
| | *Mycetocola* | 1 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 6.00 | 0.00 | 1.00 |
| | *Arthrobacter* | 10 | 1.60 | 1.00 | 1.00 | 3.40 | 1.00 | 1.10 | 1.00 | 1.10 | 1.00 | 1.00 | 0.90 | 1.00 | 1.30 | 0.00 | 0.00 | 7.10 | 0.30 | 1.70 |
| | *Citricoccus* | 1 | 1.00 | 1.00 | 1.00 | 3.00 | 1.00 | 3.00 | 2.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 4.00 | 1.00 | 1.00 |
| | *Kocuria* | 6 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.50 | 0.00 | 0.00 | 5.50 | 0.00 | 1.00 |
| | *Rothia* | 4 | 1.00 | 1.00 | 1.00 | 2.25 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 2.00 | 0.00 | 1.00 |
| Bifidobacteriales | *Bifidobacterium* | 3 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.33 | 1.00 | 1.00 | 0.00 | 0.00 | 7.67 | 0.00 | 1.00 |
| Actinomycineae | *Actinobaculum* | 5 | 1.00 | 1.00 | 1.20 | 1.20 | 1.20 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.00 | 1.40 | 1.00 | 0.00 | 0.00 | 2.00 | 0.00 | 1.00 |
| | *Actinomyces* | 21 | 1.10 | 1.00 | 0.95 | 1.19 | 1.52 | 0.95 | 1.05 | 0.95 | 0.81 | 0.76 | 0.00 | 1.05 | 0.86 | 0.00 | 0.00 | 2.29 | 0.00 | 0.90 |
| Kineosporiineae | *Kineococcus* | 1 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 6.00 | 0.00 | 1.00 |
| Propionibacterineae | *Propionibacterium* | 9 | 1.00 | 1.33 | 1.11 | 1.11 | 1.11 | 1.11 | 1.00 | 0.89 | 0.78 | 0.78 | 0.00 | 0.89 | 1.11 | 0.00 | 0.00 | 2.33 | 0.00 | 0.67 |
| | *Actinopolymorpha* | 1 | 4.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 9.00 | 0.00 | 1.00 |
| | *Aeromicrobium* | 2 | 1.00 | 1.00 | 1.50 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 3.50 | 0.00 | 1.00 |
| | *Pimelobacter* | 1 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 7.00 | 0.00 | 1.00 |
| | *Kribbella* | 1 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 14.00 | 0.00 | 1.00 |
| | *Marmoricola* | 2 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 4.50 | 0.00 | 1.00 |
| | *Nocardioides* | 7 | 1.29 | 1.00 | 1.14 | 1.29 | 1.00 | 1.14 | 1.00 | 1.14 | 1.00 | 1.00 | 0.14 | 1.00 | 1.14 | 0.00 | 0.00 | 5.14 | 0.00 | 1.00 |
| | *Propionicicella* | 1 | 1.00 | 1.00 | 1.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 5.00 | 0.00 | 1.00 |
| | Mean + Std | 614 | 3.12 | 1.58 | 1.36 | 2.04 | 1.20 | 1.67 | 1.18 | 1.86 | 1.52 | 1.16 | 0.87 | 1.55 | 3.35 | 0.14 | 0.20 | 10.88 | 0.35 | 1.21 |

UU

CHAPTER FOUR

# Tailoring Specialized Metabolite Production in *Streptomyces*

**Jana K. Hiltner*, Iain S. Hunter[§] and Paul A. Hoskisson*,[1]**
*Strathclyde Institute of Pharmacy and Biological Science, University of Strathclyde, Glasgow, UK
[§]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK
[1]Corresponding author: E-mail: paul.hoskisson@strath.ac.uk

## Contents



## Abstract

Streptomycetes are prolific producers of a plethora of medically useful metabolites. These compounds are made by complex secondary (specialized) metabolic pathways, which utilize primary metabolic intermediates as building blocks. In this review we discuss the evolution of specialized metabolites and how expansion of gene families in primary metabolism has lead to the evolution of diversity in these specialized metabolic pathways and how developing a better understanding of expanded primary metabolic pathways can help enhance synthetic biology approaches to industrial pathway engineering.

*You can know the name of a bird in all the languages of the world, but when you're finished, you'll know absolutely nothing whatever about the bird… So let's look at the bird and see what it's doing—that's what counts.*

***Richard P. Feynman***

## 1. INTRODUCTION

The widespread use of antibiotics is a relatively new addition to human health care that emerged from the discovery of penicillin in 1928 by Alexander Fleming and its subsequent development to industrial scale production by

*Advances in Applied Microbiology*, Volume 91
ISSN 0065-2164
http://dx.doi.org/10.1016/bs.aambs.2015.02.002

Howard Florey and coworkers. This discovery of bioactive metabolites produced by microorganisms and useful in human health, prompted the so-called golden era of antibiotics. Between the 1950s and 1970s the search for new metabolites resulted in the delivery of many chemical classes of antibiotic, antifungal, antihelminthics, anticancer agents, and immunosuppressive drugs to market (Davies & Davies, 2010). During this intensive period of research the actinobacteria came to the fore as prolific producers of bioactive metabolites, particularly antibiotics from the genera *Streptomyces* and *Micromonospora* (Davies & Davies, 2010; Hoskisson, Hobbs, & Sharples, 2000). This initiated the widespread public perception that antibiotics were "wonder-drugs" that signaled the end of life-threatening bacterial infections, yet this was premature. Rapidly after the introduction of these drugs to the clinic, antibiotic resistance was observed (Abraham & Chain, 1940; Davies & Davies, 2010). This began a race to discover and bring to the clinic new antibiotics to help combat the emergence of resistance. Yet with the introduction of each new drug, rapid resistance was observed. This inevitable resistance, coupled to the rising costs of development, tightening of regulatory rules, diminishing discovery rates (and rediscovery of known compounds), and the lack of financial returns due to short-treatment durations resulted in the withdrawal of large pharmaceutical companies from large-scale antibiotic discovery in the 1990s (Projan, 2003). These problems coupled with profligate use of antibiotics in medicine and agriculture lead to the emergence of extensive antibiotic resistance on a global scale. In 2009, the World Health Organization raised concerns that the rise of antibiotic resistance was becoming one of the major threats to human health and that researchers, clinicians, industry, and policy makers needed to work together to address this multifactorial problem. These concerns lead to the release of a global action plan on antimicrobial resistance (http://www.who.int/drugresistance/en/).

Synthetic biology is a new methodology that can enable development of novel, clinically useful antibiotics. Recent advances in the technology of DNA synthesis, the ability to assemble longer tracts of DNA and at much reduced cost, provide the opportunity to design and improve biosynthetic gene clusters. Moreover this synthetic biology revolution has great potential to transform the more traditional disciplines of metabolic engineering, through creation of novel heterologous hosts with engineered precursor supply or deletion of native biosynthetic clusters to reduce wastage of precursors (Gomez-Escribano & Bibb, 2011, 2012) that may help overcome metabolic limitations.

In this review we will focus on primary metabolism in streptomycetes and how a genome-level understanding of the evolution of metabolic nodes

can aid the rational development of metabolic engineering targets for improved metabolite production.

## 2. INTERPLAY OF PRIMARY AND SECONDARY METABOLISM IN STREPTOMYCETES

The distinction of primary and secondary metabolism probably arose through the studies of Albrecht Kossel, who proposed that plants show two different metabolisms: "primary" and "secondary." Primary metabolism is common among all organisms and is composed of all essential reactions, whereas secondary metabolism is thought to be specialized, distinct, and comprised of species-specific pathways (Firn & Jones, 2009; Hartmann, 2008). This view has been widely adopted and adapted through many fields of biology and while the designations imply secondary metabolism is less important than primary metabolism, this view has been modified over the years, and it is implicit that secondary metabolism is dependent on supply of precursors from primary metabolism.

Disconnecting primary and secondary metabolism was challenged by Firn and Jones (2009) as misleading. This accords with the term "specialized metabolites" which may be a more useful term to replace the bias implied by the term "secondary"—indicating that this aspect has less importance (Davies, 2013; van Keulen & Dyson, 2014).

## 3. STREPTOMYCETES AS SPECIALIZED METABOLITE PRODUCERS

The actinobacterial phylum represents a large lineage of physiologically and morphologically diverse bacteria that includes the industrially, agriculturally, and medically important genera *Bifidobacterium*, *Corynebacterium*, *Mycobacterium*, *Nocardia*, *Leifsonia*, *Frankia*, and *Streptomyces* (Ventura et al., 2007). The most speciate genus is *Streptomyces* which are sporulating soil bacteria with a filamentous growth habit, characterized by large (>7 Mbp) high G + C-content, linear genomes. The streptomycete life cycle begins when a unigenomic spore germinates and grows through apical extension to form a mat of vegetative mycelia that, in response to nutrient-limitation produce aerial hyphae, which subsequently form septa and eventually form mature spores (Flärdh & Buttner, 2009). Specialized metabolites are produced at the onset of the developmental process and molecular studies have identified common regulators of both processes (Chandra & Chater, 2013). The

plethora of specialized metabolites produced by this group of organisms have a high degree of chemical diversity including classes such as polyketides, terpenes, lactams, aminoglycosides, and nonribosomal peptides with many species able to produce multiple examples of the same class of specialized metabolite (Bérdy, 2005). To emphasize this diversity, readers are directed to the StreptomeDB database (Lucas et al., 2013; http://www.pharmaceutical-bioinformatics.de/streptomedb/), where the structures of more than 2400 compounds from more than 1900 species are held. These estimates are increasing and will continue to increase exponentially along with the expansion of whole genome sequencing projects and specialized metabolite predictions software such as antiSMASH (Medema et al., 2011) and EvoMining (Cruz-Morales & Barona-Gómez Personal Communication). In addition, the sequences of about 7758 actinobacteria genomes are currently available in the Genomes Online Database (GOLD; Reddy et al., 2015) with the number continually increasing. This represents a vast resource for identification of biotechnologically useful genes and gene clusters. One particularly revealing actinobacterial genome feature that the next-generation sequencing revolution has opened our eyes to is the vast array of antibiotic (and other specialized metabolite) biosynthetic gene clusters present in the genomes of actinobacterial strains. This was first observed in well-studied species such as *Streptomyces coelicolor* for which, prior to the availability of the whole genome sequences, we knew of only four bioactive metabolites produced by the strain, yet the whole genome sequence revealed more than 20 specialized metabolite gene clusters in its repertoire, that are often referred to as cryptic or silent biosynthetic clusters (Bentley et al., 2002). This trend has continued with the release of each streptomycete genome. Significantly, the biosynthetic clusters that are cryptic or poorly expressed in their natural hosts offer great potential for the discovery of novel, clinically useful compounds. Moreover these gene clusters represent a significant resource of genes for synthetic biology to create novel metabolites through synthetic biology or semisynthesis, where existing compounds may be biosynthesized as a chemical backbone and then modified further through synthetic chemistry. Great advances have been made in this area in recent years (Wu, 2000). However semisynthetic derivatives of natural products are often limited by availability of starting material. Semisynthetic derivatives of erythromycin such as azithromycin require increasing amounts of the starter molecule yet the natural producing strains have been difficult to engineer to high production levels when compared to related species (Wu et al., 2011). Given that specialized metabolites, such as

antibiotics, are derived from primary metabolic starter units, a deeper understanding of how these precursors are synthesized and channeled into the biosynthesis of these specialized metabolites offers great potential in metabolic engineering and manipulation through synthetic biology approaches for the development of novel compounds. Many specialized metabolites have acetyl-CoA, malonyl-CoA (Olano, Méndez, & Salas, 2010), or amino acids as direct precursors (Stirrett, Denoya, & Westpheling, 2009). Competition for the same precursors between central metabolism and specialized metabolism may represent key conflicts in supply of precursors such as in fatty acid metabolism and polyketide synthesis (Rodriguez, Navone, Casati, & Gramajo, 2012). Given these issues, a better understanding of precursor supply may enable increased production of the poorly understood cryptic/silent biosynthetic clusters.

Therefore there are several challenges facing researchers in this field, such as how to awaken or enhance production of the array of cryptic gene clusters emerging from whole genome sequencing projects and how we can understand better the links between primary and secondary metabolism so that we can increase metabolite flow to maximize industrial yield of new and existing medically useful compounds.

## 4. EVOLUTION OF PRIMARY AND SPECIALIZED METABOLISM

Primary metabolism refers to the core pathways of central metabolism that provides building blocks for all the cellular macromolecules including DNA, RNA, proteins, lipids/fatty acids, etc., and also provides the precursors for specialized metabolites. Surprisingly many primary metabolic genes are nonessential for survival due to genetic redundancy providing isoenzymes or alternative reactions that allow the cell to adapt to changing environmental conditions providing an adaptive robustness to metabolism (Kim & Copley, 2007). The use of the term "redundancy" may be misleading as it suggests nonessentially. However, there may be multiple routes to a metabolic intermediate under a given set of environmental conditions. Therefore "contingency," "metabolic flexibility," or "enzyme expansion" may be more suitable terms to reflect this phenomenon (Challis & Hopwood, 2003; Noda-García & Barona-Gómez, 2013; Treangen & Rocha, 2011). Such terms would also account for the so-called "moonlighting" enzyme functions and catalytic promiscuity where the main catalytic function is supplemented by the catalysis of additional reactions (Copley, 2003, 2012,

2014, 2015). The main reason for these promiscuous activities is that evolution of the "perfect" catalytic site is difficult and natural selection acts upon those that provide a "good enough" functionality (Copley, 2015). This indicates that such accidental catalysis may shift fitness effects to other cellular functions and can aid in the generation of new pathways and chemistry within cells, such as specialized metabolism.

An interesting feature of streptomycete genomes is that multiple genes are often annotated to code for the same biochemical function in central carbon metabolism (Bentley et al., 2002; Hiltner & Hoskisson, unpublished; Figure 1). Understanding the relationship between function and evolution is a key to elucidating metabolic plasticity and exploiting these traits in biotechnology. In metabolic models these homologous functions are often combined into one flux pathway, yet this does not reflect the nuances of regulation and allostery for each gene product. However, knowledge of regulation and allostery is necessary to understand the roles of these gene expansions to fully appreciate functionality at the biochemical level. These redundant functions are hypothesized to allow cellular flexibility and adaptation in dynamic environments. Interestingly, given that streptomycetes exhibit so many of these gene expansions it may also, in part, reflect their ability to produce such a vast variety of specialized metabolites, as these pathways are usually connected to core metabolism and use common intermediates. There are well-studied examples of such gene family expansions in streptomycete developmental genes and DNA-binding protein families (Clark & Hoskisson, 2011; Girard et al., 2013). However, little attention has been paid to central metabolism. An example of gene expansion and evolution relating to specialized metabolites is that of polyketide biosynthesis for which the multimodular and iterative polyketide synthases that perform the stepwise condensation of activated carboxylic acid subunits provide the carbon backbones for the polyketides. It appears that these pathways have evolved from fatty acid synthases (Jenke-Kodama, 2005). This premise is based on the presence of highly conserved modules such as the ketoacyl synthase domains and acyl carrier proteins which through duplications, deletion, and horizontal gene transfer (HGT) have lead to the diverse chemistry array of polyketide chemistry observed today (Jenke-Kodama, 2005; Ridley, Lee, & Khosla, 2008).

The nature of these gene expansions needs to be carefully considered as it is generally assumed that these expansions arise through gene duplications. However, little consideration has been paid to the role of HGT in metabolic gene expansion (Noda-García & Barona-Gómez, 2013; Noda-García et al.,

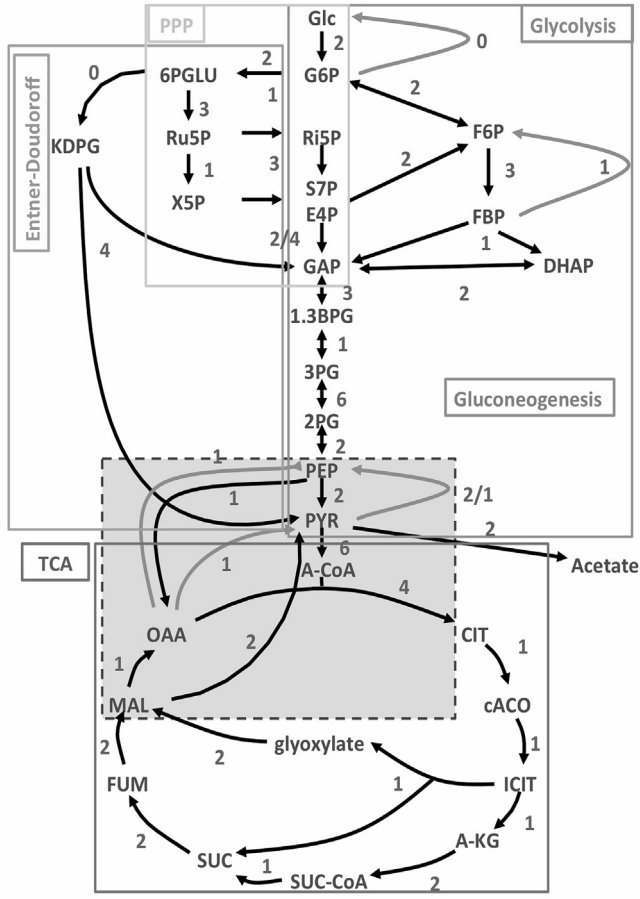



**Figure 1** ***Schematic overview with main metabolites of the central carbon metabolism grouped according to pathway.*** Numbers indicate number of genes predicted in the genome of *Streptomyces coelicolor*. Glc, glucose; G6P, glucose-6-phosphate; 6-PGLU, 6-phosphogluconate; Ru5P, ribulose-5-phosphate; X5P, xylose-5-phosphate; KDPG, 2-keto-3-deoxy-6-phosphogluconate; F6P, fructose-6-phosphate; FBP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; Ri5P, ribose-5-phosphate; S7P, seduheptulose-7-phosphate; E4P, erythrose-4-phosphate; GAP, glyceraldehyd-3-phosphate; 1.3BGP, 1.3-bisphosphoglycerate; 3 PG, 3-phosphoglycerate; 2 PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; ACoA, acetyl-CoA; Cit, citrate; cAco cisAconitate; ICit, isocitrate; A-KG, alpha-ketoglutarate (2-oxoglutarate); SucCoA, succinyl-CoA; Suc, succinate; Fum, fumarate; Mal, malate; OAA, oxaloacetate.

2013). This reflects the idea that gene duplication is an important source of biological innovation, where orthologous genes exhibit conserved functionality, and paralogs tend toward diverged function. However, integrating the role of HGT in this process requires sophisticated tools to identify HGT

events in metabolic genes, coupled with thorough studies of enzyme function to tease out detailed mechanisms. Recently, Noda-Garcia (Noda-García et al., 2013) showed that horizontally acquired metabolic genes could drive evolution of existing metabolic function through altered substrate specificity, and this may reflect a common, but under appreciated mechanism for enzyme expansion in prokaryotic genomes.

This is a developing area of interest that requires integration of a range of techniques (genomics, molecular genetics, biochemistry, X-ray crystallography, molecular dynamics simulations, and evolutionary modeling), but offers great potential for deep understanding of evolution of enzyme function and how this has contributed to metabolic plasticity. Ultimately studies such as these can be valuable to inform metabolic engineering for biotechnology.

## 5. THE PEP-PYR-OAA NODE AS TARGET FOR METABOLIC ENGINEERING

The phosphoenolpyruvate-pyruvate-oxaloacetate node is a major branch point within central carbon metabolism of all organisms that acts as a connection point for glycolysis, gluconeogenesis, and the TCA cycle (Figure 2). Many key precursors for specialized metabolites are derived from the node or pathways are limited by flux through it. Yet it is surprisingly diverse among bacteria. Sauer and Eikmanns (2005) examined this node in *Escherichia coli*, *Corynebacterium glutamicum*, and *Bacillus subtilis* in their roles as major workhorses for the industrial production of bulk chemicals such as amino acids, organic acids, or proteins. These authors concluded that this node of metabolism is a key target for metabolic engineering in bacteria.

This node represents a major flux distribution point for carbon skeletons in the cell with the key metabolites being phosphoenolpyruvate (PEP), pyruvate (PYR), and oxaloacetate (OAA). The reactions that interconvert each of these substrates are listed in Table 1. Pyruvate can be metabolized further into either OAA or acetyl-CoA, the former being a precursor for amino acids such as aspartate, lysine, methionine, threonine, and isoleucine and the latter for fatty acids and or polyketides. Across a range of species the architecture of this node can be highly variable and likely reflects the ecology of individual organisms. Here we will focus mainly on *Streptomyces* and the related actinobacterium, *Corynebacterium*, which with only one copy of each gene per reaction has a much reduced gene expansion at this node when compared to *Streptomyces* (Table 2).

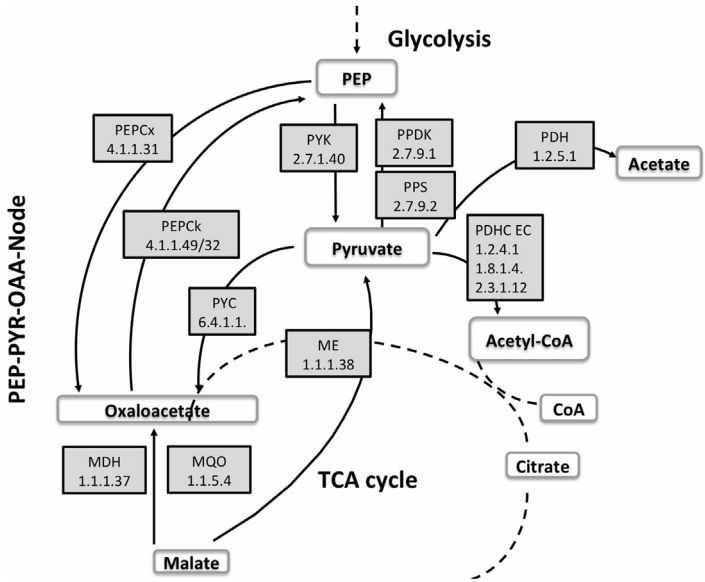



**Figure 2** General overview of the reactions that form the phosphoenolpyruvate-pyruvate-oxaloacetate node of central carbon metabolism including the EC numbers of the enzyme responsible for each reaction.

It is also known that manipulating primary metabolic pathways in the PEP-PYR-OAA node can lead to higher product yields in a range of bacterial species. In *C. glutamicum*, a well-studied amino acid producing organism, it is well established that this node is crucial for amino acid production. Much can be learned from studying the PEP-PYR-OAA node in this organism in terms of how the basal node functions. PEP carboxylase (PEPC) is not essential for lysine production and has no effect on the growth rate when deleted. However, a mutant lacking both PEPC and pyruvate kinase (Pyk) has decreased growth rates as well as reduced rates of lysine production rates (Koffas & Stephanopoulos, 2005). Moreover, inducing low expression levels of *aceE* (E1 subunit of the pyruvate dehydrogenase complex (PDHC)) through promoter exchange and deletion of *pqo* (pyruvate:quinone oxidoreductase) and *ppc* (PEP carboxylase) leads to an increase in L-valine production (Buchholz et al., 2013). It was also shown that L-lysine production could be increased by reducing PDHC levels indicating that this enzyme activity levels has an important influence on the carbon flux and can increase pyruvate-derived molecules (Blombach, Arndt, Auchter, & Eikmanns, 2009; Buchholz et al., 2013; Eikmanns & Blombach, 2014). The PEPCk gene in

**Table 1** Reactions of the PEP-PYR-OAA node including their name, abbreviations in the text, EC numbers, and the reactions they catalyze

| Enzyme | Code | EC number | Reaction catalyzed | Part of minimal set of genes |
|---|---|---|---|---|
| PEP carboxylase | PEPCx | 4.1.1.31 | $H_2O$ + PEP + $CO_2$ <-> $PO_4^{3-}$ + OAA | No |
| PEP carboxykinase | PEPCk | 4.1.1.32 | GTP + OAA <-> GDP + PEP + $CO_2$ | Yes |
| Malic enzyme | ME | 1.1.1.38 | (S)-MAL + $NAD^+$ <-> PYR + $CO_2$ + NADH + $H^+$<br>OAA <-> PYR + $CO_2$ | Yes |
| Malate dehydrogenase (quinone) | MQO | 1.1.5.4 | (S)-MAL + a quinone <-> OAA + reduced quinone | No |
| Malate dehydrogenase | MDH | 1.1.1.37 | (S)-MAL + $NAD^+$ <-> OAA + NADH + $H^+$ | Yes |
| Pyruvate carboxylase | PYC | 6.4.1.1 | ATP + PYR + $HCO^{3-}$ <-> ADP + $PO_4^{3-}$ + OAA | No |
| Pyruvate kinase (pyk) | PYK | 2.7.1.40 | ADP + PEP <-> PYR + ATP | Yes |
| Pyruvate phosphate dikinase (PPDK) | PPDK | 2.7.9.1 | ATP + PYR + P <-> AMP + PEP + diP | No |
| PEP synthase (PPS) | PPS | 2.7.9.2 | ATP + PYR + P + $H_2O$ <-> AMP + PEP + diP | No |
| Pyruvate dehydrogenase | PDH | 1.2.5.1 | PYR + ubiquinone + $H_2O$ <-> ACE + ubiquinol + $CO_2$ | No |
| Pyruvate dehydrogenase complex | PDHC E1 | 1.2.4.1 | PYR + ThdP <-> HeThdP + $CO_2$<br>HeThdP + lipoamide-E <-> S-acetyldihydrolipoamide-E + ThdP | Yes |
| Pyruvate dehydrogenase complex | PDHC E2 | 2.3.1.12 | CoA + S-acetyldihydrolipoamide-E <-> acetyl-CoA + lipoamide-E | |
| Pyruvate dehydrogenase complex | PDHC E3 | 1.8.1.4 | PYR + CoA + $NAD^+$ <-> acetyl-CoA + $CO_2$ + NADH + $H^+$ | |

ACE, acetate; ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; GDP, guanosine diphosphate; GTP, guanosine triphosphate; HeThdP, 2-(alpha-hydroxyethyl)thiamine diphosphate; MAL, malate; NAD, nicotinamide adenine dinucleotide; NADH, reduced nicotinamide adenine dinucleotide; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; ThdP, thiamin diphosphate.

**Table 2** Comparison of PEP-PYR-OAA node expansion in a range of bacteria, indicating the number of homologues indicating the level of gene expansion at each point. The nucleotide cofactor is indicated where appropriate

| | ***Streptomyces coelicolor*** | ***Streptomyces rimosus*** | ***Saccharopolyspora erythraea*** | ***Streptomyces tsukubaensis*** | ***Corynebacterium glutamicum*** | ***Escherichia coli*** | ***Bacillus subtilis*** |
|---|---|---|---|---|---|---|---|
| Pyk | 2 | 2 | 1 | 2 | 1 | 2 | 1 |
| PEPCx | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| PEPCk | 1 GTP | 1 GTP | 1 GTP | 1 GTP | 1 GTP | 1 ATP | 1 ATP |
| Pyc | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| ME | 2 | 1 | 2 | 2 | 1 | 1 | 4 |
| mdh | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| mqo | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| PEP synthase | 1 | 1 | 3 | 0 | 1 | 1 | 2 |
| PPDK | 2 | 2 | 0 | 1 | 1 | 0 | 0 |
| PDH | 2 | 1 | 2 | 1 | 1 | 1 | 0 |
| PDHC | ≥6 E1 | ≥3 E1 | ≥6 E1 | ≥3 E1 | 1 | 1 | 4 |

Mdh, malate dehydrogenase; ME, malic enzyme; Mqo, malate dehydrogenase (quinone); PDH, pyruvate dehydrogenase; PDHC, pyruvate dehydrogenase complex; PEPCk, PEP carboxykinase; PEPCx, PEP carboxylase; PPDK, pyruvate phosphate dikinase; Pyc, pyruvate carboxylase; Pyk, pyruvate kinase.

*C. glutamicum* also influences the production of glutamate and lysine; inactivation leads to an increase in production whereas overexpression decreases production (Riedel et al., 2001). This indicates that blocking certain routes within metabolism can increase flux in diverse pathways. Disrupting these processes can also be mediated through mutating the transcriptional regulators. For example, deletion of *pckR* (Cg0196), a negative repressor of PEPCk during growth on glucose results in cellular PEPCk activity even in the presence of glucose (Hyeon, Kang, Kim, You, & Han, 2012).

Overexpression of pyruvate carboxylase (*pyc*) in *C. glutamicum* results in increased PCx activity and glutamate production in an optimized lysine-producing strain. However, inactivation of *pyc* resulted in no PCx activity and lower levels of glutamate production levels (Peters-Wendisch et al., 1998). A similar effect was observed for threonine production and its precursor homoserine, pointing toward the importance of Pyc and PCx in amino acid metabolism and industrial production (Peters-Wendisch et al., 2001). In a lysine overproducing strain, deletion of pyruvate kinase (*pyk*) resulted in similar growth rates but higher rates of overflow metabolites such as dihydroxyacetone and glycerol as well as a shift from pyruvate carboxylase to phosphoenolpyruvate carboxylase flux during glucose utilization (Becker, Klopprogge, & Wittmann, 2008) which enabled a metabolic bypass via malic enzyme to account for deletion of *pyk*. This highlights the level of metabolic flexibility that this node provides in central carbon metabolism.

The Gram-negative bacterium *E. coli* has two genes encoding pyruvate kinase; *pykA* and *pykE* (Muñoz & Ponce, 2003). Disruption of a single pyruvate kinase (*pykA*) in a phenylalanine overproducing strain in combination with inactivation of a Phosphotransferase system (PTS) sugar transporter, leads to redistribution of cellular carbon flux when glucose was the substrate. Inactivation of PTS and PykA (which is allosterically regulated by AMP) resulted in decreased PCx, PEPCk, and TCA cycle activities. Inactivation of PTS and PykF (which is allosterically regulated by fructose 1,6 biphosphate) resulted in increased OAA formation from PEP and flux through the TCA cycle. Interestingly, both strains showed increased production of phenylalanine compared to the parental strain (Meza, Becker, Bolivar, Gosset, & Wittmann, 2012). The enlargement of this part of the PEP-PYR-OAA node in *E. coli* compared to *C. glutamicum* and the results reported by Meza (Meza et al., 2012) indicate that dissection of these aspects of enzyme expansion in metabolism is difficult. In the streptomycetes, this is especially true when up to six copies of some enzymes within the PEP-PYR-OAA node are present (Figure 2 and Table 2).

The PEP-PYR-OAA node is poorly understood in streptomycetes. However, some primary metabolic enzymes that have undergone gene expansion in *Streptomyces* have been studied in more detail. One such example is the presence of three copies of the glycolytic enzyme phosphofructokinase (SCO1214—*pfkA*3; SCO2119—*pfkA*1; and SCO5426—*PfkA*2) catalyzing the addition of a second phosphate to fructose-6-phosphate at the C1 position. Deletion of *pfkA*2 leads to an increase of undecylprodigiosin and actinorhodin production on certain media, whereas deletion of either *pfkA*1 or *pfkA*3 does not show the same phenotype. The *pfkA*2 mutant also had higher intracellular pools of glucose-6-phosphate and fructose-6-phosphate and radiolabeling experiments indicated increased flux through the pentose phosphate pathway (PPP) and concomitant increased levels of reduced nicotinamide adenine dinucleotide phosphate (NADPH) showing that despite having three genes encoding the same function they have different physiological roles (Borodina et al., 2008). These data also correlate with higher production of methylenomycin and increased PPP flux during slow growth (Obanye, Hobbs, Gardner, & Oliver, 1996). In *Streptomyces clavuligerus*, two glyceraldehyde-3-phosphate dehydrogenase (GAPDH) genes are present in the genome, *gap*1 and *gap*2. Disruption of *gap*1 leads to an increase in clavulanic acid production, but not when *gap*2 is deleted. Since clavulanic acid biosynthesis starts with the condensation of L-arginine and glyceraldehyde-3-phosphate (G3P), this downstream block of *gap*1 appears to redirect flux toward clavulanic acid biosynthesis rather than glycolysis. Furthermore it demonstrates the different physiological roles played by the different isoforms of GAPDH (Li & Townsend, 2006). In the model streptomycete *S. coelicolor* disruption of either of the genes encoding the two isoforms of glucose-6-phosphate dehydrogenase *zwf*1 (SCO6661) and *zwf*2 (SCO1937) leads to changes in the production of the polyketide actinorhodin. Again, two genes encoding for the same function seem to play different roles. It appears that *zwf*2 plays a more important role than *zwf*1 for directing the flux toward actinorhodin production (Butler et al., 2002; Ryu, Butler, Chater, & Lee, 2006). Similarly it has also been shown that disruption of *zwf* in *Streptomyces lividans*, which also has two copies results in higher production of actinorhodin and undecylprodigiosin production (Butler et al., 2002). Further studies in chemostat culture of *S. lividans* indicate that the flux of carbon was dependent on growth rate and the carbon source (Avignone Rossa et al., 2002). When gluconate was utilized as carbon source a higher flux through PPP was observed than on glucose

and a decline in secondary metabolite production resulted when growth rate was increased (Avignone Rossa et al., 2002).

Within the PEP-PYR-OAA node of *Streptomyces* only two genes have been studied in detail—PEPCx and the malic enzymes from *Streptomyces coelicolor*. The activity of this enzyme increased during biosynthesis of actinorhodin and overexpression of PEPCx in *Streptomyces lividans* reduced the growth rate of the strain, delaying actinorhodin biosynthesis (Bramwell, Nimmo, Hunter, & Coggins, 1993). The study of the two malic enzymes SCO2951 and SCO5261 revealed that the mutant in SCO2951 and the double mutant show decreased actinorhodin production, due to a decrease of expression of *actIIORF4* and the double mutant also has decreased triacylglycerol storage levels during exponential growth (Rodriguez et al., 2012). These data indicate the importance of this node to specialized metabolite production, especially polyketides, and highlight the potential for further investigation.

Recent global metabolomics studies in *Streptomyces* have also revealed key findings about the metabolic flexibility at the PEP-PYR-OAA node at onset of specialized metabolite production. Studies conducted during phosphate and L-glutamate depletion indicate that most changes in the global metabolite pool (metabolome) were in amino acid and organic acid levels (Wentzel, Sletta, Consortium, Ellingsen, & Bruheim, 2012). During phosphate depletion the amino acid pools of histidine, phenylalanine, tyrosine, alanine, valine, leucine, glycine, proline, isoleucine, and lysine were increased in addition to the intracellular pools of succinate and ornithine pools. Intracellular pools of glutamate and aspartate were both reduced as was pyruvate, citrate, 2-oxoglutarate, fumarate, and malate. Under glutamate depletion the pools of histidine, phenylalanine, tyrosine, alanine, valine, leucine, glycine, proline, and lysine decreased initially before recovering at around 40 h. This was also observed for citrate and succinate. Pyruvate, 2-oxoglutarate, fumarate, malate, and ornithine were all reduced in addition to the amino acids glutamate, glutamine, and aspartate (Wentzel et al., 2012). These data confirm the central role for pyruvate in balancing central carbon metabolism during growth and reinforces our understanding of key branch points in specialized metabolite production. An additional study by this group to develop cultivation media for studying metabolic shifts tested a wide range of carbon sources (arabinose, alanine, aspartate, glucose/glutamate, glucose, glutamate, proline, Tween 20, Tween 40, Tween 60, Tween 80, and xylose) and examined expression of the PDHC genes. Only one of the genes (SCO2183) showed decreased

expression on Tween and alanine. SCO2181, SCO2180, and SCO4919 had increased expression, yet under all other conditions no expression differences were observed, which may suggest that the gene products of the PDHC may act as a metabolic bottleneck during glucose growth (Wentzel et al., 2012). Furthermore, studies of carbon preferences using $^{13}$C-glucose to label metabolites during culture on glucose and glutamate as a mixed carbon source, indicate that during rapid growth glycolysis and PPP are enriched for $^{13}$C compounds, but upon cessation of growth $^{13}$C-labeled PPP intermediates decreased. The TCA cycle is generally low in $^{13}$C-labeled intermediates indicating that glutamate is a preferred carbon source, being catabolized via 2-oxoglutarate following deamination and release of ammonium ions. Interestingly 2-oxoglutarate can be decarboxylated in the TCA cycle to form malate and further decarboxylated to pyruvate which can be converted to acetyl-CoA by PDHC. Acetyl-CoA is an important precursor for fatty acids and polyketides. Clearly glutamate is the preferred carbon source providing the main cellular carbon and with glucose playing an ancillary role (Wentzel et al., 2012). Interestingly, it is known that *Streptomyces* secrete pyruvate and 2-oxoglutarate during growth under certain conditions, prior to specialized metabolite production (Hobbs et al., 1992), which may reflect the inefficiency of formation of acetyl-CoA from pyruvate under some physiological conditions. PDHC expansion in streptomycetes may be an evolutionary solution to this phenomenon. These natural examples may provide a framework for metabolic engineering strategies.

## 6. CONCLUDING REMARKS

Detailed insight into the evolution, regulation, and biochemistry of primary metabolism and how this feeds into specialized metabolism will help us understand better complex biological systems and will allow targeting key points in metabolism for metabolic engineering. While global "omics" studies and modeling can help with general phenomena, a full understanding of gene function in a classical reductionist manner is the only way to gain true biological insight. These approaches guide synthetic biology strategies for strain and pathway construction and increased production of medically useful metabolites. There are key points within central metabolism that represent excellent metabolic engineering targets, such as the PEP-PYR-OAA node, which are particularly well suited

for engineering metabolite production and will make it more efficient and easier to enhance industrial production processes.

## ACKNOWLEDGMENTS

We would like to thank Dr Paco Barona-Gómez, Dr Pablo Cruz-Morales, Dr Lorena Fernández-Martínez, and Professor David Hodgson for helpful discussions over many years relating to actinobacteria metabolism. We would like to thank the Scottish Universities Life Science Alliance (SULSA) and Acies Bio (Slovenia) for funding a BioSkape PhD Scholarship to JKH. PAH would like to thank The Leverhulme Trust (RPG248), NERC (NE/M001415/1), and Medical Research Scotland (422 FRG) for funding research in his laboratory.

## REFERENCES

Abraham, E. P., & Chain, E. (1940). An enzyme from bacteria able to destroy penicillin. *Reviews of Infectious Diseases, 10*, 677–678.

Avignone Rossa, C., White, J., Kuiper, A., Postma, P. W., Bibb, M., & Teixeira de Mattos, M. J. (2002). Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*. *Metabolic Engineering, 4*, 138–150.

Becker, J., Klopprogge, C., & Wittmann, C. (2008). Metabolic responses to pyruvate kinase deletion in lysine producing *Corynebacterium glutamicum*. *Microbial Cell Factories, 7*, 8.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature, 417*, 141–147.

Bérdy, J. (2005). Bioactive Microbial Metabolites. *The Journal of Antibiotics, 58*(1), 1–26.

Blombach, B., Arndt, A., Auchter, M., & Eikmanns, B. J. (2009). L-valine production during growth of pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum* in the presence of ethanol or by inactivation of the transcriptional regulator SugR. *Applied and Environmental Microbiology, 75*, 1197–1200.

Borodina, I., Siebring, J., Zhang, J., Smith, C. P., van Keulen, G., Dijkhuizen, L., et al. (2008). Antibiotic overproduction in *Streptomyces coelicolor* A3(2) mediated by phosphofructokinase deletion. *The Journal of Biological Chemistry, 283*, 25186–25199.

Bramwell, H., Nimmo, H. G., Hunter, I. S., & Coggins, J. R. (1993). Phosphoenolpyruvate carboxylase from *Streptomyces coelicolor* A3(2): purification of the enzyme, cloning of the ppc gene and over-expression of the protein in a streptomycete. *Biochemical Journal, 293*(Pt 1), 131–136.

Buchholz, J., Schwentner, A., Brunnenkan, B., Gabris, C., Grimm, S., Gerstmeir, R., et al. (2013). Platform engineering of *Corynebacterium glutamicum* with reduced pyruvate dehydrogenase complex activity for improved production of l-lysine, l-valine, and 2-ketoisovalerate. *Applied and Environmental Microbiology, 79*, 5566–5575.

Butler, M. J., Bruheim, P., Jovetic, S., Marinelli, F., Postma, P. W., & Bibb, M. J. (2002). Engineering of primary carbon metabolism for improved antibiotic production in *Streptomyces lividans*. *Applied and Environmental Microbiology, 68*, 4731–4739.

Challis, G. L., & Hopwood, D. A. (2003). Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proceedings of the National Academy of Sciences USA, 100*(Suppl. 2), 14555–14561.

Chandra, G., & Chater, K. F. (2013). Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiology Reviews, 38*. n/a–n/a.

Clark, L. C., & Hoskisson, P. A. (2011). Duplication and evolution of devA-like genes in *Streptomyces* has resulted in distinct developmental roles. *PLoS One, 6*, e25049.
Copley, S. D. (2003). Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Current Opinion in Chemical Biology, 7*, 265–272.
Copley, S. D. (2012). Moonlighting is mainstream: paradigm adjustment required. *Bioessays, 34*, 578–588.
Copley, S. D. (2014). An evolutionary perspective on protein moonlighting. *Biochemical Society Transactions, 42*, 1684–1691.
Copley, S. D. (2015). *An evolutionary biochemist's perspective on promiscuity 1*–7. Elsevier Ltd.
Davies, J. (2013). Specialized microbial metabolites: functions and origins. *The Journal of Antibiotics, 66*, 361–364.
Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews, 74*, 417–433.
Eikmanns, B. J., & Blombach, B. (2014). The pyruvate dehydrogenase complex of *Corynebacterium glutamicum*: an attractive target for metabolic engineering. *Journal of Biotechnology, 192*(Pt B), 339–345.
Firn, R. D., & Jones, C. G. (2009). A Darwinian view of metabolism: molecular properties determine fitness. *Journal of Experimental Botany, 60*, 719–726.
Flärdh, K., & Buttner, M. J. (2009). *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology, 7*, 36–49.
Girard, G., Traag, B. A., Sangal, V., Mascini, N., Hoskisson, P. A., Goodfellow, M., et al. (2013). A novel taxonomic marker that discriminates between morphologically complex actinomycetes. *Open Biology, 3*, 130073.
Gomez-Escribano, J. P., & Bibb, M. J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology, 4*, 207–215.
Gomez-Escribano, J. P., & Bibb, M. J. (2012). *Streptomyces coelicolor* as an expression host for heterologous gene clusters. *Methods in Enzymology, 517*, 279–300.
Hartmann, T. (2008). The lost origin of chemical ecology in the late 19th century. *Proceedings of the National Academy of Sciences USA, 105*(12), 4541–4546.
Hobbs, G., Obanye, A. I., Petty, J., Mason, J. C., Barratt, E., Gardner, D. C., et al. (1992). An integrated approach to studying regulation of production of the antibiotic methylenomycin by *Streptomyces coelicolor* A3(2). *Journal of Bacteriology, 174*, 1487–1494.
Hoskisson, P. A., Hobbs, G., & Sharples, G. P. (2000). Response of *Micromonospora echinospora* (NCIMB 12744) spores to heat treatment with evidence of a heat activation phenomenon. *Letters in Applied Microbiology, 30*, 114–117.
Hyeon, J. E., Kang, D. H., Kim, Y. I., You, S. K., & Han, S. O. (2012). GntR-type transcriptional regulator PckR negatively regulates the expression of phosphoenolpyruvate carboxykinase in *Corynebacterium glutamicum*. *Journal of Bacteriology, 194*, 2181–2188.
Jenke-Kodama, H. (2005). Evolutionary implications of bacterial polyketide synthases. *Molecular Biology and Evolution, 22*, 2027–2039.
van Keulen, G., & Dyson, P. J. (2014). Production of specialized metabolites by *Streptomyces coelicolor* A3(2). *Advances in Applied Microbiology, 89*, 217–266.
Kim, J., & Copley, S. D. (2007). Why metabolic enzymes are essential or nonessential for growth of *Escherichia coli* K12 on glucose. *Biochemistry, 46*, 12501–12511.
Koffas, M., & Stephanopoulos, G. (2005). Strain improvement by metabolic engineering: lysine production as a case study for systems biology. *Current Opinion in Biotechnology, 16*, 361–366.
Li, R., & Townsend, C. A. (2006). Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in *Streptomyces clavuligerus*. *Metabolic Engineering, 8*, 240–252.

Lucas, X., Senger, C., Erxleben, A., Grüning, B. A., Döring, K., Mosch, J., et al. (2013). StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Research, 41*, D1130–D1136.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research, 39*, W339–W346.

Meza, E., Becker, J., Bolivar, F., Gosset, G., & Wittmann, C. (2012). Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*. *Microbial Cell Factories, 11*, 127.

Muñoz, E., & Ponce, E. (2003). Pyruvate kinase: current status of regulatory and functional properties. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 135*, 197–218. Elsevier.

Noda-García, L., & Barona-Gómez, F. (2013). Enzyme evolution beyond gene duplication: a model for incorporating horizontal gene transfer. *Mobile Genetic Elements, 3*, e26439.

Noda-García, L., Camacho-Zarco, A. R., Medina-Ruíz, S., Gaytán, P., Carrillo-Tripp, M., Fülöp, V., et al. (2013). Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Molecular Biology and Evolution, 30*, 2024–2034.

Obanye, A. I. C., Hobbs, G., Gardner, D. C. J., & Oliver, S. G. (1996). Correlation between carbon flux through the pentose phosphate pathway and production of the antibiotic methylenomycin in *Streptomyces coelicolor* A3(2). *Microbiology, 142*, 133–137.

Olano, C., Méndez, C., & Salas, J. A. (2010). Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Natural Product Reports, 27*, 571–616.

Peters-Wendisch, P. G., Kreutzer, C., Kalinowski, J., Pátek, M., Sahm, H., & Eikmanns, B. J. (1998). Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the pyc gene. *Microbiology, 144*(Pt 4), 915–927.

Peters-Wendisch, P. G., Schiel, B., Wendisch, V. F., Katsoulidis, E., Möckel, B., Sahm, H., et al. (2001). Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. *Journal of Molecular Microbiology and Biotechnology, 3*, 295–300.

Projan, S. J. (2003). Why is big Pharma getting out of antibacterial drug discovery? *Current Opinion in Microbiology, 6*, 427–430.

Reddy, T. B. K., Thomas, A. D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., et al. (2015). The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Research, 43*, D1099–D1106.

Ridley, C. P., Lee, H. Y., & Khosla, C. (2008). Evolution of polyketide synthases in bacteria. *Proceedings of the National Academy of Sciences USA, 105*, 4595–4600.

Riedel, C., Rittmann, D., Dangel, P., Möckel, B., Petersen, S., Sahm, H., et al. (2001). Characterization of the phosphoenolpyruvate carboxykinase gene from *Corynebacterium glutamicum* and significance of the enzyme for growth and amino acid production. *Journal of Molecular Microbiology and Biotechnology, 3*, 573–583.

Rodriguez, E., Navone, L., Casati, P., & Gramajo, H. (2012). Impact of malic enzymes on antibiotic and triacylglycerol production in *Streptomyces coelicolor*. *Applied and Environmental Microbiology, 78*, 4571–4579.

Ryu, Y. G., Butler, M. J., Chater, K. F., & Lee, K. J. (2006). Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. *Applied and Environmental Microbiology, 72*, 7132–7139.

Sauer, U., & Eikmanns, B. J. (2005). The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiology Reviews, 29*, 765–794.

Stirrett, K., Denoya, C., & Westpheling, J. (2009). Branched-chain amino acid catabolism provides precursors for the Type II polyketide antibiotic, actinorhodin, via pathways that are nutrient dependent. *Journal of Industrial Microbiology and Biotechnology, 36*, 129–137.

Treangen, T. J., & Rocha, E. P. C. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genetics,* 7, e1001284.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., et al. (2007). Genomics of actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews,* 71, 495–548.

Wentzel, A., Sletta, H., Consortium, S., Ellingsen, T. E., & Bruheim, P. (2012). Intracellular metabolite pool changes in response to nutrient depletion induced metabolic switching in *Streptomyces coelicolor*. *Metabolites, 2*, 178–194.

Wu, Y. J. (2000). Highlights of semi-synthetic developments from erythromycin A. *Current Pharmaceutical Design, 6*, 181–223.

Wu, J., Zhang, Q., Deng, W., Qian, J., Zhang, S., & Liu, W. (2011). Toward improvement of erythromycin A production in an industrial *Saccharopolyspora erythraea* strain via facilitation of genetic manipulation with an artificial attB site for specific recombination. *Applied and Environmental Microbiology,* 77, 7508–7516.