

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

Differential transcription of expanded gene families in central carbon metabolism of  
*Streptomyces coelicolor* A3(2)

**Authors:** Jana K Schniete<sup>1,4</sup>, Richard Reumerman<sup>2</sup>, Leena Kerr<sup>3</sup>, Nicholas P Tucker<sup>1</sup>,  
Iain S Hunter<sup>1</sup>, Paul R Herron<sup>1</sup>, and Paul A Hoskisson<sup>1\*</sup>

Jana K Schniete, Richard Reumerman, Leena Kerr, Nicholas P Tucker, Iain S Hunter,  
Paul R Herron, and Paul A Hoskisson

**Corresponding Author:** paul.hoskisson@strath.ac.uk

**Affiliations:** <sup>1</sup>Strathclyde Institute of Pharmacy and Biomedical Sciences, University  
of Strathclyde, 161 Cathedral Street, Glasgow, G4 0RE, UK.

<sup>2</sup>Isomerase Therapeutics, Cambridge, CB10 1XL, UK

<sup>3</sup>Institute of Earth and Life Sciences, School of Energy, Geoscience, Infrastructure and  
Society, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS, UK.

<sup>4</sup>Current Address: Biology Department, Edge Hill University, St Helens Road,  
Ormskirk, Lancashire, L39 4QP, UK.

25 **Abstract**

26 **Background:** Streptomycete bacteria are prolific producers of specialised  
27 metabolites, many of which have clinically relevant bioactivity. A striking feature of  
28 their genomes is the expansion of gene families that encode the same enzymatic  
29 function. Genes that undergo expansion events, either by horizontal gene transfer or  
30 duplication, can have a range of fates: genes can be lost, or they can undergo neo-  
31 functionalisation or sub-functionalisation. To test whether expanded gene families in  
32 *Streptomyces* exhibit differential expression, an RNA-Seq approach was used to  
33 examine cultures of wild-type *Streptomyces coelicolor* grown with either glucose or  
34 tween as the sole carbon source.

35 **Results:** RNA-Seq analysis showed that two-thirds of genes within expanded gene  
36 families show transcriptional differences when strains were grown on tween compared  
37 to glucose. In addition, expression of specialised metabolite gene clusters  
38 (actinorhodin, isorenieratane, coelichelin and a cryptic NRPS) was also influenced by  
39 carbon source.

40 **Conclusions:** Expression of genes encoding the same enzymatic function had  
41 transcriptional differences when grown on different carbon sources. This  
42 transcriptional divergence enables partitioning to function under different physiological  
43 conditions. These approaches can inform metabolic engineering of industrial  
44 *Streptomyces* strains and may help develop cultivation conditions to activate the so-  
45 called silent biosynthetic gene clusters.

46 **Key words:** *Streptomyces*, RNA-Seq, central carbon metabolism, metabolic plasticity,  
47 gene redundancy, silent biosynthetic clusters, metabolic engineering

48

## 49 **Introduction**

50 Streptomyces bacteria are major source of clinically useful bioactive natural products  
51 including antibiotics, immunosuppressive and anti-cancer agents. A remarkable  
52 feature of their genomes is that there are often several genes that appear to encode  
53 the same biochemical function (Bentley *et al.*, 2002; Schniete *et al.*, 2018). This is  
54 often referred to as 'genetic redundancy', where two or more genes are performing the  
55 same biochemical function (Nowak *et al.*, 1997). While genetic redundancy does occur  
56 in nature, many so called 'redundant' genes have evolved divergent functions and  
57 provide the functional diversity and evolutionary robustness that can be observed in  
58 genomes from a wide range of organisms (Wagner, 2008b). There are two main  
59 mechanisms that contribute to the expansion of gene families within genomes - the  
60 duplication of genes or horizontal gene transfer events (Treangen & Rocha, 2011;  
61 Wagner, 2008a, b). Such gene family expansions are well known in streptomyces  
62 regulatory genes (Bush, 2018; Chater & Chandra, 2006; Clark & Hoskisson, 2011;  
63 Girard *et al.*, 2013) and in specialised metabolite biosynthesis (Jenke-Kodama *et al.*,  
64 2005; Ridley *et al.*, 2008). Yet, there has been limited attention focussed on the genes  
65 of central metabolism despite a number of gene expansion events being identified  
66 from biochemical and phylogenomic studies. There is an increasing appreciation that  
67 gene expansion events in central metabolism may facilitate the evolution of  
68 specialised metabolites (Borodina *et al.*, 2008; Cruz-Morales *et al.*, 2016; Fernández-  
69 Martínez & Hoskisson, 2019; Noda-García & Barona-Gómez, 2013; Noda-García *et al.*,  
70 2013; Schniete *et al.*, 2018). Whilst expansion of gene families is widespread,  
71 expression differences, gene dosage, cofactor variation, allostery or substrate  
72 affinities are seldom taken into account during the construction of metabolic models  
73 for *Streptomyces*. As a result, homologous functions often combined into a single flux  
74 pathway that may not reflect the physiological nature of each gene product. This is  
75 especially stark when attempting to understand the supply of precursor molecules from  
76 central metabolism to the production of specialised metabolites.

77 It was hypothesised that gene families that have undergone gene expansion events  
78 would exhibit different expression profiles if they have diverged functionally. To  
79 understand the role of these gene expansions in *Streptomyces* and their impact on  
80 central metabolism and specialised metabolism, an RNA-Seq approach was taken  
81 using *S. coelicolor* grown either on glucose or tween as sole carbon source. Cultures

82 were compared at a single point during growth (mid-log phase) to understand how  
83 transcription of central metabolic genes varies when the cultures are growing primarily  
84 via glycolysis (glucose as the sole carbon source) or gluconeogenically (with the  
85 mono-oleate tween). It was found that expanded gene families exhibit different  
86 responses to growth on different carbon sources. These data will enable prioritisation  
87 of targets for metabolic engineering and will be informative for the construction of  
88 metabolic models.

89

## 90 **Results and discussion**

### 91 **Expanded gene families in carbon metabolism exhibit different transcriptional** 92 **profiles**

93 To investigate the differences between glycolytic and gluconeogenic growth  
94 conditions, *S. coelicolor* M145 (an A3(2) strain) was grown respectively on glucose  
95 and Tween40 as sole carbon source. Growth of liquid cultures was monitored by cell  
96 dry weight and samples were removed for total RNA extraction at mid-exponential  
97 phase, which for glucose was 19 h and for tween was 36 h (Supp. Fig. 1). None of the  
98 pigmented antibiotics (actinorhodin or undecylprodigiosin) could be detected at time  
99 points RNA was harvested. The specific growth rate of *S. coelicolor* M145 was 0.18 h<sup>-1</sup>  
100 when grown on glucose and 0.1 h<sup>-1</sup> for growth on tween (See Supp. Fig. 1).

101 RNA samples from these cultures showed global effects on transcription, when  
102 analysed by RNA-Seq, when the strain was grown on different carbon sources. Growth  
103 on tween resulted in 644 genes being differentially expressed when compared to  
104 growth on glucose. Of the 644 differentially expressed genes, 37% were predicted to  
105 encode hypothetical proteins, 9% encoded regulatory genes, 12% encoded central  
106 carbon metabolic enzymes, 6% were transporters, 4% were associated with  
107 specialised metabolism, 2% encoded genes associated with stress metabolism and 1  
108 % of genes were associated with DNA replication, nitrogen metabolism and genes  
109 associated with metal metabolism (See Supp. Table 1-6). Growth on Tween40  
110 resulted in an up-regulation of genes associated with gluconeogenesis and fatty acid  
111 degradation when compared to cultures grown on glucose (Fig. 1C and Supp. Tab.1-  
112 6).

113 Gene Ontology (GO) enrichment analysis (using PANTER <http://www.pantherdb.org>;  
114 geneontology.org, Mi et al., 2013) was used to investigate the enrichment of specific  
115 gene ontologies between the two growth conditions. Analyses were performed for  
116 genes upregulated on Tween and one for genes downregulated on Tween. The  
117 glycolytic genes are downregulated when *S. coelicolor* was grown on tween as would  
118 be expected, with GO enrichment analysis highlighting *tpiA* (SCO1945), *pyk1*  
119 (SCO2014) and *pyk2* (SCO5423), one of the three GAPDH (SCO7511) and a single  
120 copy of the pyruvate dehydrogenase E1 complex (SCO2183), it was noted that all  
121 genes belonged in the oxidoreductase gene ontology grouping.

122 GO enrichment analysis for genes upregulated during growth on Tween were found  
123 to be those associated with fatty acid catabolism, gluconeogenesis (*pyc*; SCO0546)  
124 and another of the three GAPDH homologues (SCO7040).

125 To examine transcription of expanded gene families, two categories of expression  
126 profiles were considered to reflect changes across the gene family that are predicted  
127 to encode the same enzymatic function: Type 1, where all genes within an expanded  
128 gene family behaved similarly under the growth conditions, and Type II where  
129 members of an expanded gene family exhibit differential gene expression i.e. one or  
130 more member of the family increased, while expression of other members of the gene  
131 family, increased to a different degree, decreased or remained the same (Fig. 1A and  
132 Supp. Table 1). We identified 34 enzymatic reactions in central metabolism whose  
133 expression differed when cultures were grown on tween or glucose. Of these, 21 had  
134 more than one enzyme predicted to encode the same enzymatic function, i.e. were  
135 expanded gene families (Fig. 1; Supp. Table 1). When the expression profiles of these  
136 genes were examined further, 9/21 enzymatic reactions had families that showed Type  
137 I expression profiles and 12/21 expanded gene families exhibited Type II patterns of  
138 expression. This indicates that despite their identity in genome annotation, members  
139 of some expanded gene families are not redundant in function but have distinctive  
140 physiological roles (Borodina *et al.*, 2008; Gubbens *et al.*, 2012; Schniete *et al.*, 2018).  
141 It is unlikely that functional differences can be attributed at the transcriptional level.  
142 However, the data enable prioritisation of targets for metabolic engineering, such as  
143 modulating the expression poorly expressed members of gene families.

144 The glycolysis pathway, unsurprisingly, showed reduced expression when the cultures  
145 were grown on Tween40 as sole carbon source. Only two glycolytic enzyme families  
146 exhibited Type II expression, with an increase in expression for the minor glucose  
147 kinase (SCO0063; 5-fold increase in expression), whereas expression of the primary  
148 *glk* (SCO2126) remained unchanged under both conditions. GAPDH also showed  
149 Type II expression profiles, with expression of SCO7040 increased on tween, whereas  
150 the other two copies of the genes encoding GAPDH had reduced expression  
151 (SCO1947 and SCO7511). A previous proteomic-based study also identified an  
152 increase in abundance of the SCO7040 protein when grown on a non-glucose carbon  
153 source (fructose; Gubbens *et al.*, 2012). The two PK genes, as we have previously

154 demonstrated, exhibited Type I expression (Schniete *et al.*, 2018), with both copies  
155 being down regulated when cultures are grown on tween.

156 Overall, genes involved in gluconeogenesis were upregulated when cultures were  
157 grown on tween compared to glucose - as would be expected. The only expanded  
158 gene family encoding gluconeogenic function is that encoding the two PPKs, which  
159 both displayed substantially increased expression (Type I expression), with *ppdk1*  
160 (SCO0208) up regulated 11-fold up and *ppdk2* (SCO2494) upregulated 30-fold  
161 (*ppdk2*). This suggests that, under these conditions, *S. coelicolor* may be using  
162 unconventional gluconeogenic routes for anaplerotic reactions rather than via the  
163 glyoxylate shunt, as expression of ICL and both MS enzymes remain unchanged  
164 during growth on either carbon source.

165 Expression of the pentose phosphate pathway showed little differential expression  
166 under the two conditions studied. Type II expression profiles were observed for *zwf*,  
167 with expression of SCO6661 reduced when cells were grown on tween, whereas  
168 expression of SCO1937 was unchanged between the two conditions. This was  
169 consistent with the work of Gubbens *et al.* (2012), where SCO6661 was  
170 downregulated when cultures were grown on fructose rather than glucose. There was  
171 no evidence of changes in transcriptional activity for the putative Entner-Doudoroff  
172 (ED) pathway genes for KDPG aldolase (SCO2298, SCO3473 and SCO3495).  
173 However reduced expression of two of the three phosphogluconate dehydratase  
174 homologues (SCO3877 and SCO6658) was observed when cultures were grown on  
175 tween (Fig. 1C). Whilst it was reported previously that there is no active ED pathway  
176 in *S. coelicolor* (Gunnarsson *et al.*, 2004) these data suggest that putative  
177 phosphogluconate dehydratase homologs do respond to changes in carbon source.

178 In general, expression of genes encoding enzymes of the tricarboxylic acid (TCA)  
179 cycle remained stable under both growth conditions as would be expected, given this  
180 core part of metabolism is required for biosynthesis under all physiological conditions.  
181 The exception was expanded gene families, which exhibited differential expression  
182 (Type II): expression of the fumarase (SCO5042) was reduced two-fold on tween  
183 whilst SCO5044 expression remained unchanged; one copy of the malic enzyme  
184 (SCO2951) showed a two-fold decrease in expression, compared to SCO5261, a  
185 second malic enzyme, which remained unchanged under both conditions; the  
186 extensive expansion in *S. coelicolor* of genes encoding the PHDC<sub>E1</sub> subunit exhibited

187 a range of expression changes - expression profiles for SCO1269, SCO7124,  
188 SCO2371 and SCO3816/3817 were the same for both carbon sources, whereas  
189 SCO2183 had a 3.5-fold decrease and SCO1270 a six-fold increase on tween.

190 As expected fatty acid utilisation genes were expressed more when *S. coelicolor* was  
191 grown on tween, especially cholesterol esterase (SCO5420), which had an increase  
192 of almost 37-fold. The enzyme catalyses the hydrolysis of the head group of tween  
193 from the fatty acid palmitate, which are then utilised as carbon source (Plou *et al.*,  
194 1998; Pratt *et al.*, 2000; Sakai *et al.*, 2002). The other genes from this pathway showed  
195 between 3 to 12-fold increase in expression (Supp. Table 2).

196 To verify the RNA-Seq data, we performed qPCR on the RNA samples used for the  
197 RNA-Seq experiment using primer pairs for five different genes - *pyk1* (SCO2014) and  
198 *pyk2* (SCO5423) from glycolysis, *ppdk1* and *ppdk2* from gluconeogenesis with primary  
199 sigma factor *hrdB* (SCO5820) as control. These two pairs of genes were chosen as  
200 representative of expanded families, identified in a previous study (Schniete *et al.*,  
201 2018). RNA-Seq data showed that expression was strongly up or down-regulated  
202 under the chosen conditions, whilst *hrdB* (SCO5820, as expected) showed no  
203 difference under the two conditions tested. The fold-change difference in the qPCR  
204 data was similar to that observed in RNA-Seq experiments, corroborating the wider  
205 results (Supp. Fig. 2 & 3).

### 206 **Carbon source influences specialised metabolite gene expression**

207 Genes involved in isorenieratene biosynthesis (SCO0185-0191), showed an increase  
208 in transcription in cultures grown on tween, with a three to eight-fold increase in  
209 expression across the entire operon. Isorenieratene is associated with blue light  
210 exposure, with the pathway present in green photosynthetic bacteria and a few  
211 actinobacteria (Krügel *et al.*, 1999; Takano *et al.*, 2005). Isorenieratene is a carotenoid  
212 with antioxidative properties; it is synthesized via the mevalonate-independent  
213 pathway (MEP/DOXP pathway) from basic precursors of GAP and pyruvate, derived  
214 from glycolysis to form Isopentenyl pyrophosphate (IPP) and Dimethylallyl  
215 pyrophosphate (DMAP). These metabolites then enter the carotenoid biosynthetic  
216 pathway via phytoene, lycopene and beta-carotene. The MEP/DOXP pathway also  
217 had increased expression of some genes (SCO6768 - 2.2-fold, SCO5250 - 4.4-fold)  
218 when grown on glucose (Supp. Table 2). Given the only difference between the



219 cultures was the carbon source, we hypothesise that cultures experienced oxidative  
220 stress. This is in part supported by the GO enrichment analysis where 51 genes with  
221 GO terms matching 'oxidoreductases' exhibiting increased expression on Tween as  
222 the sole carbon source, with a seven-fold increase in expression of a superoxide  
223 mutase (SCO0999) and an 8-fold increase of expression of a putative bacterioferritin  
224 co-migratory protein (SCO7353; Supp. Table 2, Table XY). Caution should be  
225 exercised however as many genes in primary metabolism also share GOs with  
226 'oxidoreductases'.

227 A non-ribosomal peptide synthetase (NRPS) pathway (SCO6429-6438), the  
228 siderophore coelichelin biosynthetic gene cluster (SCO0491-0498) and the  
229 actinorhodin cluster (SCO5071-5092) all had decreased expression when grown on  
230 tween compared to glucose by two to five-fold, two to three-fold and three to 11-fold  
231 respectively (Fig. 1 and Supp. Table 2). This may reflect a tighter control of entry into  
232 specialised metabolism when cultures are grown on tween rather than glucose,  
233 although further experiments are needed to confirm this.

234 Studies such as this can inform metabolic modelling approaches for strain  
235 improvement enabling expression data and metabolic flux analysis to be taken into  
236 account with respect to isoenzymes within expanded gene families rather than  
237 combining the activities of all copies of a gene family into a single flux (Fernández-  
238 Martínez & Hoskisson, 2019). The identification of metabolic engineering targets can  
239 often be complicated through the presence of multiple genes that putatively encode  
240 the same enzymatic function in streptomycetes. Studies such as this enables the  
241 identification of transcriptionally active genes under relevant conditions to be identified  
242 and prioritised for manipulation. This is exemplified by GAPDH in *S. coelicolor* where  
243 three genes are annotated as having this function, with one (SCO7040) increasing  
244 expression on tween when the other two copies show decreased expression under  
245 the same conditions. This knowledge will help inform choices and enable prioritisation  
246 of metabolic engineering targets for industrial streptomycetes and the supporting  
247 dataset will also inform researchers on the role of hypothetical proteins and regulators,  
248 during growth on two different carbon sources that drive metabolism along two  
249 different pathways.

250

## 251 **Materials and Methods**

### 252 **Bacterial strains and growth conditions**

253 *Streptomyces coelicolor* A3(2) M145 (Kieser *et al.*, 2000) was used throughout the  
254 study. Spores were germinated in 50 ml of 2x YT medium in flasks containing a metal  
255 spring and were incubated for up to 8h at 30°C and 250 rpm until emerging germ tubes  
256 were visible under a microscope (Kieser *et al.*, 2000). The cultures were harvested by  
257 centrifugation, washed twice with 0.25 M TES buffer (pH 7.2) and pellets were  
258 resuspended in media. Growth curves were performed at 400 ml scale with minimal  
259 medium (Hobbs *et al.*, 1989) in 2 L flasks containing a metal spring at 30°C, shaken  
260 at 250 rpm. Carbon source added was adjusted to ensure that each culture contained  
261 166.5 mM equivalent of carbon.

262

### 263 **RNA extraction and sample preparation**

264 *S. coelicolor* biomass (15 ml sample) from liquid cultures was harvested by  
265 centrifugation (5 min, 4°C, 4000 x g). Cell pellets were resuspended in an equal volume  
266 of RNAprotect (Qiagen) for 5 min at room temperature. Following centrifugation (5 min,  
267 4°C, 6000 x g) biomass was then resuspended in 1ml 1x TE buffer containing 15 mg/ml  
268 lysozyme. Tubes were vortexed for 10 s and incubated at room temperature for 60  
269 min whilst shaking. 1 ml RLT buffer (Qiagen RNA Isolation Kit) + 10 µl β-  
270 mercaptoethanol was added and vortexed and a phenol chloroform extraction followed  
271 by an ethanol precipitation was carried out. The sample was then purified using a  
272 commercial RNA isolation Kit (Qiagen). The isolated RNA was treated with RNase  
273 free DNase (Ambion, Life Technologies) as specified by the manufacturer.

274 Quantification of RNA was carried out using Qubit® (Life Technologies). The quality  
275 and integrity of the RNA was assessed using a Bioanalyzer (Agilent). Furthermore, the  
276 RNA samples were also used as templates for a generic PCR in order to check for  
277 DNA contamination using primers for *hrdB* (SCO5820).

278 To enrich the samples for mRNA, rRNA depletion was performed (rRNA depletion Kit  
279 Ribo Zero Magnetic Kit for Gram-positive bacteria; Epicentre [Illumina]) according to  
280 the manufacturer's instructions. The rRNA-depleted samples were then precipitated  
281 with ethanol and resuspended according to the manufacturer's instructions. The

282 quality and integrity of the samples were then analysed on the BioAnalyzer (Agilent)  
283 and the concentration was determined using Qubit (Life Technologies).

284

### 285 **Library preparation**

286 cDNA synthesis and library preparation was carried out using the Ion Total RNA-Seq  
287 Kit v2 Revision E from Ion Torrent, Life Technologies. The manufacturer's protocol for  
288 less than 100 ng rRNA depleted samples was followed. The yield and size distribution  
289 of the amplified cDNA was assessed using BioAnalyzer (Agilent). The three samples  
290 harvested from glucose cultures and three samples from tween grown cultures were  
291 barcoded and pooled to a concentration of 20 pM as specified by the manufacturer.  
292 The Ion OneTouch 2 system using the Ion PGM Template OT200 Kit (Life  
293 Technologies) was used for template preparation which includes the steps of  
294 emulsification, amplification and enrichment of the library. The libraries were checked  
295 using the quality control assay for the Qubit (Life Technologies).

296

### 297 **RNA Sequencing**

298 Sequencing of the samples was carried out using a Ion Torrent Personal Genome  
299 Machine System (PGM; Life Technologies) on a 316v2 chip following the procedures  
300 in the manual, Ion PGM Sequencing 200 Kit v2 User Guide Revision 3.0. All sequence  
301 data were deposited on the Sequence Read Archive (SRA) under the BioProject:  
302 PRJNA566372 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA566372>).

303

### 304 **Data analysis**

305 Sequencing data were downloaded from the Ion Torrent server version Torrent Suite  
306 4.0.2 in Fastq format. Reads were trimmed and low quality data removed. Reads were  
307 mapped to the reference genome of *S. coelicolor* (Bentley *et al.*, 2002); Genbank:  
308 NC\_003888.3). The data were analysed using CLC Genomics Workbench (Version  
309 7.5, Qiagen). The software showed on average 99.8% and 99.7% alignment of the  
310 reads to the reference sequence. The CLC differential gene expression tool was used  
311 to determine differential gene expression. To examine differential gene expression  
312 analysis within the dataset, the cut-off was set at a p-value of 0.05. Genewise

313 dispersions are estimated by conditional maximum likelihood using the total count for  
314 the gene of interest followed by empirical Bayes to obtain a consensus value (Smyth  
315 and Verbyla, 1996; Robinson and Smyth, 2007). The differential expression is then  
316 assessed using Fisher's exact test adjusted to over-dispersed data (Robinson and  
317 Smyth, 2008). The raw data output from CLC Genomics Suite can be found in Supp  
318 Table 8, and a list of all differentially expressed genes is shown in Supp. Table 5. For  
319 the heatmap representation in Fig 1C, all differentially expressed genes were  
320 normalised to the maximum (+1) and minimum (-1) of all significantly different  
321 expressed genes and this can be found in Supp. Table 6. The code to create Fig 1C  
322 can be found in the Supp. Code 1 ( <https://doi.org/10.6084/m9.figshare.10008914.v3>).  
323 The colour code was expressed relative to the highest and lowest expression change  
324 in the data shown ranging from green to red respectively.  
325 Gene ontology enrichment analysis was performed using Panther classification from  
326 Gene Ontology at NCBI (<https://www.ncbi.nlm.nih.gov/pubmed/23868073>) and  
327 summarised in Supp. Table 7.

328

### 329 **qPCR**

330 qPCR was performed on the same samples as the RNA-Seq in order to provide  
331 independent confirmation of the data obtained. Each primer pair was tested using  
332 genomic DNA as template. cDNA was synthesized from the RNA samples using  
333 qPCRBIO cDNA synthesis Kit (PCR Biosystems) following the manufacturer's  
334 instructions. Quantification of the cDNA was carried out using the QuantiFluor ssDNA  
335 system (Promega) with the Qubit Fluorometer (Life Technologies).

336 All cDNA samples were diluted to a concentration of 10 ng/μl and each reaction  
337 contained 10 ng of cDNA. Samples were mixed with qPCRBio MasterMix and the  
338 respective primers (Supp. Table 9; Kit 2x qPCRBIO SyGreen Mix Lo-ROX from  
339 PCRBIOSYSTEMS). Using a Corbett Research 6000 (Qiagen) machine, PCR  
340 reactions were subjected to a three-stage thermocycling reaction (one 3 min step at  
341 95 °C, followed by 40 cycles of 5 s at 95 °C and one 25 s step at 60 °C. Each reaction  
342 was carried out in duplicate and a no template control was included for each set of  
343 primers. To allow quantification, standard curves for each gene were prepared (in  
344 triplicate) using purified PCR product from a genomic DNA PCR. This template was

345 diluted to create seven different standards ranging from  $10^1$ - $10^7$  molecules/per  
346 reaction. (Supplementary Figure 3) and were used to calculate the concentrations of  
347 the unknown samples obtained in the RNA Sequencing.

348

349

350 **Funding Information**

351 This work was funded through PhD studentships from the Scottish Universities Life  
352 Sciences Alliance (SULSA) to JKS and RR.

353

354 **Author Contributions**

355 Conceptualization: JKS, PAH, NPT, LK

356 Data curation: JKS & RR

357 Formal analysis: JKS & RR

358 Funding acquisition: PAH, PRH, ISH

359 Methodology: JKS, RR, PAH, PRH, NPT, LK

360 Project administration: PAH & PRH

361 Supervision: PAH, ISH & PRH

362 Writing – original draft: JKS and PAH

363 Writing – review and editing: JKS, RR, LK, NPT, ISH, PRH, PAH

364

365 **Acknowledgements**

366 The authors would like to thank Jan Schniete for writing the MATLAB code for the  
367 heatmap representation in Figure 1C. PAH would like to acknowledge the support of  
368 NERC (grant: NE/M001415/1) and BBSRC/NPRONET (grant: NPRONET POC045).  
369 PAH and PRH acknowledge SULSA funding for PhD studentships to JKS and RR.

370 **Conflicts of interest**

371 The authors declare that there are no conflicts of interest

372

373 **Ethical statement**

374 No ethical approval was required.

375

376

377 **References**

- 378 Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R.,  
379 James, K. D., Harris, D. E., Quail, M. A., Kieser, H. & other authors. (2002).  
380 Complete genome sequence of the model actinomycete *Streptomyces coelicolor*  
381 A3(2). *Nature* 417, 141–147.
- 382 Borodina, I., Siebring, J., Zhang, J., Smith, C. P., van Keulen, G., Dijkhuizen, L. &  
383 Nielsen, J. (2008). Antibiotic overproduction in *Streptomyces coelicolor* A3(2)  
384 mediated by phosphofructokinase deletion. *J Biol Chem* 283, 25186–25199.
- 385 Bush, M. J. (2018). The actinobacterial WhiB-like (Wbl) family of transcription factors.  
386 *Molecular Microbiology* 110, 663–676.
- 387 Chater, K. F. & Chandra, G. (2006). The evolution of development in *Streptomyces*  
388 analysed by genome comparisons. *FEMS Microbiology Reviews* 30, 651–672.
- 389 Clark, L. C. & Hoskisson, P. A. (2011). Duplication and evolution of *devA*-like genes  
390 in *Streptomyces* has resulted in distinct developmental roles. *PLoS ONE* 6,  
391 e25049–.
- 392 Cruz-Morales, P., Kopp, J. F., Martínez-Guerrero, C., Yáñez-Guerra, L. A., Selem  
393 Mojica, N., Ramos-Aboites, H., Feldmann, J. & Barona-Gómez, F. (2016).  
394 Phylogenomic analysis of natural products biosynthetic gene clusters allows  
395 discovery of arseno-organic metabolites in model streptomycetes. *Genome*  
396 *Biology and Evolution* 8, 1906–1916.
- 397 Fernández-Martínez, L. T. & Hoskisson, P. A. (2019). Expanding, integrating, sensing  
398 and responding: the role of primary metabolism in specialised metabolite  
399 production. *Curr Opin Microbiol* 51, 16–21.
- 400 Girard, G., Traag, B. A., Sangal, V., Mascini, N., Hoskisson, P. A., Goodfellow, M. &  
401 van Wezel, G. P. (2013). A novel taxonomic marker that discriminates between  
402 morphologically complex actinomycetes. *Open Biology* 3.
- 403 Gubbens, J., Janus, M., Florea, B. I., Overkleeft, H. S. & van Wezel, G. P. (2012).  
404 Identification of glucose kinase dependent and independent pathways for carbon  
405 control of primary metabolism, development and antibiotic production in  
406 *Streptomyces coelicolor* by quantitative proteomics. *Molecular Microbiology* 86,  
407 1490-1507.

408 Hobbs, G., Frazer, C., Gardner, D. J., Cullum, J. & Oliver, S. (1989). Dispersed growth  
409 of *Streptomyces* in liquid culture. *Applied Microbiology and Biotechnology* **31**, 272-  
410 277.

411 Jenke-Kodama, H., Sandmann, A., Müller, R. & Dittmann, E. (2005). Evolutionary  
412 Implications of Bacterial Polyketide Synthases. *Molecular Biology and Evolution*  
413 **22**, 2027–2039.

414 Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. (2000). *Practical*  
415 *Streptomyces Genetics*. John Innes Foundation.

416 Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. (2013). Large-scale gene  
417 function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551-  
418 1566.

419 Noda-García, L. & Barona-Gómez, F. (2013). Enzyme evolution beyond gene  
420 duplication: A model for incorporating horizontal gene transfer. *Mob Genet*  
421 *Elements* **3**, e26439.

422 Noda-García, L., Camacho-Zarco, A. R., Medina-Ruíz, S., Gaytán, P., Carrillo-Tripp,  
423 M., Fülöp, V. & Barona-Gómez, F. (2013). Evolution of substrate specificity in a  
424 recipient's enzyme following horizontal gene transfer. *Molecular Biology and*  
425 *Evolution* **30**, 2024–2034.

426 Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. (1997). Evolution of genetic  
427 redundancy. *Nature* **388**, 167–171.

428 Plou, F. J., Ferrer, M., Nuero, O. M., Calvo, M. V., Alcalde, M., Reyes, F. & Ballesteros,  
429 A. (1998). Analysis of Tween 80 as an esterase/ lipase substrate for lipolytic  
430 activity assay. *Biotechnology Techniques* **12**, 183–186.

431 Pratt, J., Cooley, J. D., Purdy, C. W. & Straus, D. C. (2000). Lipase Activity from Strains  
432 of *Pasteurella multocida*. *Curr Microbiol* **40**, 306–309.

433 Ridley, C. P., Lee, H. Y. & Khosla, C. (2008). Evolution of polyketide synthases in  
434 bacteria. *Proc Natl Acad Sci USA* **105**, 4595–4600.

435 Sakai, Y., Hayatsu, M. & Hayano, K. (2002). Use of tween 20 as a substrate for assay  
436 of lipase activity in soils. *Soil Science and Plant Nutrition* **48**, 729–734.

437 Schniete, J. K., Cruz-Morales, P., Selem Mojica, N., Fernández-Martínez, L. T.,



438 Hunter, I. S., Barona-Gómez, F. & Hoskisson, P. A. (2018). Expanding Primary  
439 Metabolism Helps Generate the Metabolic Robustness To Facilitate Antibiotic  
440 Biosynthesis in *Streptomyces*. *MBio* 9, e02283–17.

441 Treangen, T. J. & Rocha, E. P. C. (2011). Horizontal transfer, not duplication, drives  
442 the expansion of protein families in prokaryotes. *PLoS Genet* 7, e1001284.

443 Wagner, A. (2008a). Robustness and evolvability: a paradox resolved. *Proceedings of*  
444 *the Royal Society B: Biological Sciences* 275, 91–100.

445 Wagner, A. (2008b). Gene duplications, robustness and evolutionary innovations.  
446 *Bioessays* 30, 367–373.

447

448 **Figure Legends**

449 **Figure 1) Carbon source-dependent expression of genes in *S. coelicolor*. 1A).**

450 Schematic overview of expression patterns observed for expanded gene families Type  
451 I-expression where genes behaved in the same manner under the different growth  
452 conditions and Type II, where members of an expanded gene family exhibited  
453 differential gene expression across multiple gene family members (Green = up-  
454 regulated; Red = down-regulated & white = no expression change) **B**). Schematic  
455 overview with main metabolites of the central carbon metabolism grouped according  
456 to pathways: glycolysis (green), pentose phosphate pathway (PPP, yellow), Entner-  
457 Doudoroff pathway (not complete in *S. coelicolor*; blue), tricarboxylic acid cycle (TCA,  
458 purple), gluconeogenesis (orange arrows). **C**). Visualisation of gene expression using  
459 Heatmap by pathway (gene expression was normalised to maximum and minimum  
460 values of the entire dataset, and heat map colouring was based on maximum and  
461 minimum values of genes represented, where green is up-regulated and red is down-  
462 regulated)

463 **Legend for abbreviations:** Catalytic functions: Glk = glucose kinase, Zwf = Glucose-  
464 6-phosphate 1-dehydrogenase, Pgl = 6-phosphogluconolactonase, GhD = 6-  
465 phosphogluconate dehydrogenase, Rpi = Ribose-5-phosphate isomerase, Rpe  
466 = Ribulose 5-Phosphate 3-Epimerase, Tal = transaldolase, Tkt = transketolase, Pgd =  
467 6-phosphogluconate dehydratase , KDPGal = KDPG aldolase, Eno = enolase, Fba =  
468 fructose-1,6- bisphosphate aldolase, Gap = glyceraldehyde-3-phosphate  
469 dehydrogenase, Pfk = phosphofructokinase, Pgi = Phosphoglucose isomerase, Pgc =  
470 Phosphoglycerate kinase, Pgm = Phosphoglycerate mutase, Pyk = pyruvate kinase,  
471 Tpi = triosephosphate isomerase, FBPase = FBP bisphosphatase, PEPCK = PEP  
472 carboxykinase, PPS = PEP synthase, PPDK = pyruvate phosphate dikinase, Pyc =  
473 Pyruvate carboxylase, CS = citrate synthase, Aco = aconitase, Idh = isocitrate  
474 dehydrogenase, AKGdh = Alpha-ketoglutarate dehydrogenase, Suc = succinyl-CoA  
475 synthetase, Sdh = succinate dehydrogenase, Fum = Fumarase, Mdh = malate  
476 dehydrogenase, me = malic enzyme, Icl = isocitrate lyase, Ms = malate synthase,  
477 PDHC = pyruvate dehydrogenase complex. **Metabolites:** Glc = Glucose, G6P =  
478 Glucose-6-Phosphate, 6-PGLU = 6-phosphogluconate, Ru5P = Ribulose-5-  
479 phosphate, X5P = Xylose-5-Phosphate, KDPG = 2-keto-3-deoxy-6-  
480 phosphogluconate, F6P = Fructose-6-Phosphate, FBP = Fructose 1.6-bisphosphate,

481 DHAP = dihydroxyacetone phosphate , Ri5P = Ribose-5-Phosphate, S7P=  
482 Seduheptulose-7-Phosphate, E4P = Erythrose-4-Phosphate, GAP = glyceraldehyde-  
483 3-phosphate , 1.3BGP = 1.3-bisphosphoglycerate, 3PG = 3-phosphoglycerate, 2PG =  
484 2-phosphoglycerate, PEP= Phosphoenolpyruvate, PYR = Pyruvate, ACoA = Acetyl-  
485 CoA, Cit = Citrate, cAco = cisAconitate, ICit = Isocitrate, A-KG-=  $\alpha$ -Ketoglutarate,  
486 SucCoA = Succinyl-CoA, Suc = Succinate, Fum = Fumarate, Mal = Malate, OAA =  
487 Oxaloacetate

## 488 **Supplementary Figure Legends**

489 **All supplementary data is available here**  
490 <https://doi.org/10.6084/m9.figshare.10008914.v3>

491  
492  
493 **Supplementary Figure 1** Growth profiles of liquid *Streptomyces coelicolor* M145  
494 cultures grown in glucose or tween with sampling point indicator (green arrow) at mid  
495 log phase. Means of three biological replicates are shown as *ln* of cell dry weight  
496 (CDW) over time. Each experimental condition was carried out with three biological  
497 replicates. Each data point is the mean of three independent experiments and an error  
498 bar represents the standard deviation for that data.

499 **Supplementary Figure 2.** Verification of fold change data from RNA-Seq and qPCR  
500 obtained from three biological replicates with standard deviation of five genes, one  
501 constitutively expressed gene (*hrdB*), two from glycolysis (*pyk1*, *pyk2*) and two from  
502 gluconeogenesis (*ppdk1*, *ppdk2*). Experimental details are in the Materials and  
503 Methods. Pyk = pyruvate kinase; PPDK = pyruvate phosphate dikinase; *hrdB* =  
504 housekeeping sigma factor

505 **Supplementary Figure 3** Standard curves of each gene for the quantification of  
506 transcripts in qPCR experiments A purified PCR product from a PCR for each gene  
507 from genomic DNA served as template (Supp. Table 9). This template was diluted to  
508 derive seven different standards ranging from  $10^1$ - $10^7$  molecules. The standard curves  
509 were prepared in triplicate and used to calculate the concentration in the unknown  
510 samples, which were then compared to the results obtained in the RNA Sequencing.  
511 Pyk = pyruvate kinase; PPDK = pyruvate phosphate dikinase; HrdB = housekeeping  
512 sigma factor.

513

514 **Supplementary Table Legends**

515 **Supplementary Table 1** Differential gene expression (DE) and expression category  
516 for growth on tween versus glucose in central carbon metabolism showing all genes  
517 annotated for the function Legend: green = up, red = down, yellow highlighted genes  
518 = expanded gene in Streptomyces, '-' symbol indicates no significant change in  
519 expression detected, expression type meanings: I) same direction of change or no  
520 change in all genes II) different direction of change in all genes.

521 **Supplementary Table 2** Specialised metabolite gene clusters with individual genes  
522 showing differential expression for growth on tween and glucose, showing SCO  
523 number, gene function, fold change and p-value.

524 **Supplementary Table 3** Genes involved in fatty acid metabolism and EM-CoA  
525 pathway showing differential expression for growth on tween and glucose, showing  
526 SCO number, gene function, fold change and p-value.

527 **Supplementary Table 4** Regulatory genes showing differential expression for growth  
528 on tween and glucose showing SCO number, gene function, fold change and p-value.

529 **Supplementary Table 5** List of all genes showing differential expression for growth  
530 on tween and glucose showing SCO number, gene function, fold change and p-value.

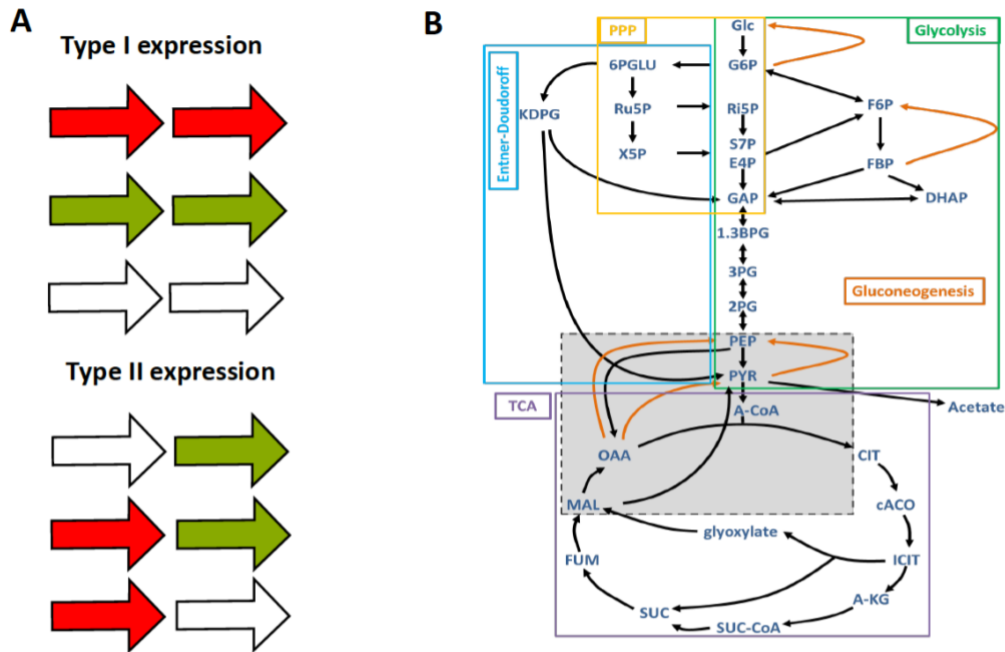
531 **Supplementary Table 6** List of all genes showing differential expression for growth  
532 on tween and glucose showing SCO number, normalised fold change (+1 highest  
533 increase in expression to -1 highest decrease in expression), fold change and p-value.

534 **Supplementary Table 7 – GO Enrichment analysis (<http://www.pantherdb.org>)**  
535 **,<https://www.ncbi.nlm.nih.gov/pubmed/23868073>)**

536  
537 **Supplementary Table 8** - Raw data output from analysis of RNA-Seq data from CLC  
538 Genomics Workbench 7.5.

539 **Supplementary Table 9** Primers utilised for qPCR specifying for which gene,  
540 direction, sequence, melting temperature, amplicon size.

541  
542  
543  
544  
545  
546  
547



**C** PRIMARY METABOLISM

SPECIALISED METABOLISM

