Comparative genomic and transcriptomic analysis guides to further enhance the biosynthesis of erythromycin by an overproducer

Xiaobo Li¹, Xiang Ke¹, Ju Chu², Lijia Qiao¹, and Yufei Sui¹

¹East China University of Science and Technology ²State Key Laboratory of Bioreactor Engineering

October 18, 2021

Abstract

Omics approaches have been applied to understand the boosted productivity of natural products by industrial high-producing microorganisms. Here, with the updated genome sequence and transcriptomic profiles derived from high-throughput sequencing, we exploited comparative omics analysis to further enhance the biosynthesis of erythromycin in an industrial overproducer, Saccharopolyspora erythraea HL3168 E3. By comparing the genome of E3 with the wild type NRRL23338, we identified fragment deletions inside 56 coding sequences and 255 single nucleotide polymorphisms over the genome of E3. Substantial numbers of genomic variations were observed in genes responsible for pathways which were interconnected to the biosynthesis of erythromycin by supplying precursors/cofactors or by signal transduction. Through comparative transcriptomic analysis, L-glutamine/L-glutamate and 2-oxoglutarate were identified as reporter metabolites. Around the node of 2-oxoglutarate, genomic mutations were also observed. Furthermore, the transcriptomic data suggested that genes involved in the biosynthesis of erythromycin were significantly up-regulated constantly, whereas some genes in biosynthesis clusters of other secondary metabolites contained nonsense mutations and were expressed at extremely low levels. Based on the omics association analysis, readily available strategies were proposed to engineer E3 by simultaneously overexpressing sucB (coding for 2-oxoglutarate dehydrogenase E2 component) and sucA (coding for 2-oxoglutarate dehydrogenase E1 component), which increased the erythromycin titer by 71% compared to E3 in batch culture. This work provides more promising molecular targets to engineer for enhanced production of erythromycin by the overproducer.

Comparative genomic and transcriptomic analysis guides to further enhance the biosynthesis of erythromycin by an overproducer

Xiaobo Li^a, Xiang Ke^a, Lijia Qiao^a, Yufei Sui^a and Ju Chu^a*

^aState Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, China

*Correspondence:

Ju Chu: juchu@ecust.edu.cn; Tel. (+86) 021-64253021; 130 Meilong Road, Shanghai 200237, China.

Abstract:

Omics approaches have been applied to understand the boosted productivity of natural products by industrial high-producing microorganisms. Here, with the updated genome sequence and transcriptomic profiles derived from high-throughput sequencing, we exploited comparative omics analysis to further enhance the biosynthesis of erythromycin in an industrial overproducer, *Saccharopolyspora erythraea* HL3168 E3. By comparing the genome of E3 with the wild type NRRL23338, we identified fragment deletions inside 56 coding sequences and 255 single nucleotide polymorphisms over the genome of E3. Substantial numbers of genomic variations were observed in genes responsible for pathways which were interconnected to the biosynthesis of erythromycin by supplying precursors/cofactors or by signal transduction. Through comparative transcriptomic analysis, L-glutamine/L-glutamate and 2-oxoglutarate were identified as reporter metabolites. Around the node of 2-oxoglutarate, genomic mutations were also observed. Furthermore, the transcriptomic data suggested that genes involved in the biosynthesis of erythromycin were significantly upregulated constantly, whereas some genes in biosynthesis clusters of other secondary metabolites contained nonsense mutations and were expressed at extremely low levels. Based on the omics association analysis, readily available strategies were proposed to engineer E3 by simultaneously overexpressing sucB (coding for 2-oxoglutarate dehydrogenase E2 component) and sucA (coding for 2-oxoglutarate dehydrogenase E1 component), which increased the erythromycin titer by 71% compared to E3 in batch culture. This work provides more promising molecular targets to engineer for enhanced production of erythromycin by the overproducer.

Key words: comparative omics, *Saccharopolyspora erythraea*, erythromycin, metabolic engineering, 2-oxoglutarate

Introduction

As one of the largest bacterial orders, actinobacteria are well known as prolific producers of numerous antibiotics, biofuels, materials, and commodity chemicals (Sayed, Abdel-Wahab, Hassan, & Abdelmohsen, 2019). Increasing numbers of new natural bioactive products are discovered in both actinomycetes inhabiting hostile environments or well-known production/laboratory strains (Newman & Cragg, 2016; Weber, Blin, et al., 2015). However, in most cases wild type strains isolated from nature only produce limited amounts of the desired compounds. To meet the industrial demand, several rounds of random mutagenesis and screening with subsequent optimization of cultivation process are usually required. Although many cases have shown the success of mutate-and-screen methods, this classical breeding strategy requires intensive labor and challenges the physiological stability of the strain, as various subculturing steps involving different cultivation conditions are required (Zeng, Guo, Xu, Chen, & Zhou, 2020). To some extent, the widely applied strategy of mutate-and-screen in industry is inevitable due to the lack of either genome/transcriptome information or feasible genetic tools for specific strains. On the other hand, most industrial mutants are likely overmutagenized after several rounds of mutate-and-screen, resulting in a large number of unnecessary genomic variations relative to the beneficial mutations (Peano et al., 2014). The underlying correlation between genomic variations and phenotypic changes is still obscure in most cases. With the great strides made in recent years on DNA and RNA sequencing, in combination with the development of efficient genetic tools for actinomycetes (Tong, Charusanti, Zhang, Weber, & Lee, 2015), rational metabolic engineering approaches provide faster and more sparking ways to improve the production of desired products (Y. Liu et al., 2019). For instance, reverse engineering combined with omics analysis has been used to identify key factors responsible for high production of secondary metabolites and to further boost the biosynthesis of desired metabolites (Lum, Huang, Hutchinson, & Kao, 2004; Peano et al., 2014; Zhuo et al., 2010; wang et al., 2019).

Even though, most mechanisms by which industrial actinomycetes increase yield or production of desired metabolites remain ambiguous. The desired products from actinomycetes are usually derived from secondary metabolism, which is interconnected tightly with primary metabolism, morphological differentiation, transcriptional regulation and post-translational modification (Butler et al., 2002; Chng, Lum, Vroom, & Kao, 2008; Manteca & Yague, 2018). Furthermore, actinomycetes have larger genomes compared to *E. coli*, *Corynebacterium glutamicum* or some other widely-used industrial strains (Redenbach, Scheel, & Schmidt, 2000). Larger genomes usually imply more complex regulatory networks for the biosynthesis of desired secondary metabolites, more genes with unknown functions and more dependence of products profile on cultivation conditions. All these factors make the rational engineering approaches in actinomycetes difficult to enhance production of desired secondary metabolites.

Saccharopolyspora erythraea is of particularly interest as it is the producer of erythromycin in industry. S. erythraea has a large circular genome with 8 Mb (Oliynyk et al., 2007), which is also a host of several kinds of novel enzyme or natural products (Sayed et al., 2019). Erythromycin serves as a broad-spectrum macrolide antibiotic against pathogenic gram-positive bacteria. A series of derivatives of erythromycin, such as azithromycin, roxithromycin and clarithromycin, have been sought out for antiparasitic, antineoplas-

tic, immunosuppressant, neurotrophic, antiinflammatory and gastroenteric therapeutic activities (Mironov, Sergienko, Nastasyak, & Danilenko, 2004). For instance, azithromycin combined with hydroxychloroquine served as a drug candidate to treat the pandemic COVID-19 (Andreani et al., 2020). However, the industrial production of erythromycin remains lower titer compared to some other antibiotics, e.g. penicillin (Y. Chen et al., 2013) and there is still large economic drive to improve the microbial production of erythromycin. *S. erythraea* is also considered as a model microorganism for investigating the production of antibiotics. The biosynthesis of erythromycin starts with the assembly of six methylmalonyl-CoA molecule and one propionyl-CoA molecules, and methylmalonyl/propionyl-CoA are common precursors of some other bioactive compounds. In addition, *S. erythraea* and other actinomycetes, e.g., *Streptomyces coelicolor* and *Streptomyces avermitilis* share high similarities regarding life cycle and regulation networks (Liao et al., 2015; Zhuo et al., 2010).

Hitherto, most previous attempts for improving the production of erythromycin by rational engineering were realized by genetic manipulations predominantly focused on stimulating the accumulation of erythromycin precursors (Reeves et al., 2006, 2007) or by optimizing the cultivation process (Y. Wang et al., 2007). Studies on transcriptional regulation of erythromycin biosynthesis gene cluster (BGC) also managed to boost the production of erythromycin through global or pathway-specific regulators (Chng et al., 2008; J. Liu et al., 2017; Z. Xu, Liu, & Ye, 2018). With fast advances of nucleotide sequencing, omics analysis could further provide more rational engineering strategies to enhance the production of erythromycin. The whole genome of S. erythraea NRRL23338 (the wild type, WT) was first reported in 2007 and initiated comparative omics analysis of S. erythraea (Oliynyk et al., 2007). Genomic and transcriptomic comparisons have been conducted between NRRL23338 and several high-erythromycin producing strains, which were isolated using classical mutagenesis and screening (Carata et al., 2009; Karnicar et al., 2016; Y. Li et al., 2013; Peano et al., 2012). Omics comparison illustrated underlying key features potentially associated with the phenotype of high-erythromycin production at the specific omics level. However, only few engineering strategies were proposed and validated. Particularly noteworthy is that all published transcriptomic data of S. erythraea were collected by using microarrays (Carata et al., 2009; Karnicar et al., 2016; Y. Li et al., 2013; Peano et al., 2007; Peano et al., 2012), which has inevitable disadvantages compared to the high throughput RNA sequencing. As considerable amount of information about S. erythraea can be obtained after DNA or RNA sequencing, more rational engineering strategies would be proposed and deserve implementation.

In the present study, a high-erythromycin producing mutant *S. erythraea* HL3168 E3 was subjected to high throughput sequencing followed by comparative omics analysis. First, the whole genome was re-sequenced with Illumina Hiseq2000 platform (C. Chen et al., 2017; Y. Li et al., 2013). Then, time-series transcriptomic profiles of E3 and the wild type NRRL23338 were obtained by RNAseq. We identified more details about genomic variations relative to the first genome sequence of E3, which was determined by pyrosequencing (Y. Li et al., 2013). By integrating genomic and transcriptomic analysis, we proposed promising molecular targets to further boost the production of erythromycin by E3. Based on the comparative omics analysis, the node of 2-oxoglutarate was manipulated in E3, which contributed to a further enhancement of erythromycin production by 71%. This work showed that comparative omics analysis can provide readily available strategies to further enhance the production of secondary metabolites even in the overproducer.

Results and Discussion

Phenotypes of E3 and NRRL23338

Phenotypes of NRRL23338 and E3 were compared on agar medium and in liquid culture with glucose as a carbon source. With respect to the glossy colonies of NRRL23338 on agar medium, E3 produced gray spores more intensively (Fig. 1a). In order to show the exclusive effects of genome mutation on phenotypes as clear as possible, batch cultivation of *S. erythraea* in a minimal liquid medium was performed (Hopwood, 1985). The growth rate of E3 was reduced by 14% compared to NRRL23338, and the biomass formation of E3 was decreased by approximately 18% (Fig. 1b). Both strains showed a two-step diauxic growth in the minimal medium, with growth suspension happening approximately at 30 h (Marcellin et al., 2013) followed by a growth slowdown until 45 h. The glucose consumption rate of E3 also decreased markedly compared to

NRRL23338 (Fig. 1c). Approximately 450 mg/L of erythromycin was produced by E3 in the liquid culture, whereas NRRL23338 produced only 20 mg/L (Fig. 1d). Regarding the production of the diffusible pigment (Cortés et al., 2002), NRRL23338 colonies appeared more reddish by excreting more pigment out of the cells than retaining it within (Fig. 1a). The reddish pigment 7-O-rahmosyl flaviolin is a kind of secondary metabolite, which is derived from malonyl-CoA (Cortés et al., 2002).

Genomic comparison of E3 and NRRL23338

Re-sequencing of the whole genome of E3 was performed using an Illumina Hiseq2000 platform resulting in approximately 1.2×10^7 raw reads of 150 bp. The overall features of the E3 genome are highly similar to NRRL23338 (Table 1) (Oliynyk et al., 2007). The length of E3 genome is 8,181,083 bp with a GC content of 71.81%. 7,669 coding sequences were predicted bioinformatically, including 7607 protein-coding sequences, 12 rRNA genes and 50 tRNA genes. E3 presents an identical ncRNA profile to NRRL23338, which indicated a relatively conserved translation process after mutagenesis.

The E3 genome is 31.72 kb shorter than that of NRRL23338 (Oliynyk et al., 2007). However, 16 coding sequences are multiplied in E3 relative to NRRL23338 (Table S1). 7 out of the 16 multi-copy genes encode transposases, e.g., 10 copies of SACE_2214, 8 copies of SACE_2314 and 9 copies of SACE_2316, which scatter over the genome of E3. Some transposases (e.g., IS 30 and IS 256) can alter the expression levels of their neighboring genes by either forming new strong promoter or transcriptional termination signals (Nagy & Chandler, 2004). Therefore, the rearrangement of transposase genes might be one of the factors leading to transcriptional changes of some genes. It may also indicate a more active DNA transposition in E3. Furthermore, highly active transposable elements with increased copy numbers might be useful tools for heterologous expression of gene clusters of other secondary metabolites in S. erythraea (Kiss, Szabo, & Olasz, 2003). Interestingly, the copy numbers of genes in Table S1 other than coding for transposases are all doubled in E3. One subgroup of the doubled genes includes acetyl-CoA synthetase gene (SACE_0337) and proline-specific permease (SACE-4252), which are associated with supplying precursors for the synthesis of erythromycin. A second subgroup of doubled genes code for 3-phenylpropionate dioxygenase ferredoxin subunit (SACE_4486), superoxide dismutase (SACE_0619) and aldehyde dehydrogenase (SACE_2377), whose functions are in association with regulating the intracellular redox status. A putative transcriptional regulator (SACE_2715) are also doubled.

On the other hand, in total 31 kb nucleotides which distribute inside 56 coding sequences are deleted in E3 (Table S2). The deletion herein was defined as the missing of fragments larger than 100 bp. The functions of these 56 damaged genes were considered blocked due to the missing nucleotides. It's thus reasonable to assume that the deleted sequences do not play important roles in the biosynthesis of erythromycin. Except for deletions in sequences with unknown functions, quite large amount of deletions are observed in genes responsible for replication and repair. A long fragment of SACE_0239~0248 is entirely deleted and was predicted as a prophage (Y. Li et al., 2013). Deletions in five genes (SACE_2777/2787/2907/4254/5157) coding for substrate transporters implied that the substrate profile of E3 was likely different from that of NRRL23338. Another complete deletion is observed at SACE_4326~4331 which is linked to terpenoids metabolism. However, the relation between terpenoids and erythromycin remains poorly understood.

In E3, we observed a total of 255 single nucleotide polymorphism (SNP) (Table S3). 160 SNPs locate in the core region of *S. erythraea*genome (Oliynyk et al., 2007). 128 nonsynonymous/nonsense/stop-mutation SNPs are located inside sequences of 118 coding regions. The protein conformation could change owing to the nonsynonymous SNPs, resulting in a modified activity of that protein (Ferro et al., 2017; Nakken, Alseth, & Rognes, 2007). 51 mutations occur in intergenic regions of 44 coding sequences, which probably affect the processing and segmental stability of a transcript containing multiple coding regions (Smolke & Keasling, 2002). To include underlying mechanisms by which the biosynthesis of erythromycin in E3 was enhanced, the functional analysis of genes with SNPs was performed (Fig. 2). Top five categories in terms of the amount of SNPs were [K] transcription, [E] amino acid metabolism and transport, [C] energy production and conversion, [G] carbohydrate metabolism and transport, and [Q] secondary metabolites biosynthesis, transport and catabolism. By mapping genes with nonsynonymous/intergenic-regions SNPs to metabolic pathways, several key nodes/pathways for the biosynthesis of erythromycin were identified (Figure S1). In general, the key nodes/pathways were tightly related to the biosynthesis of erythromycin by supplying precursors and cofactors or by signal transduction (Z. Xu, You, Tang, Zhou, & Ye, 2019). The key nodes/pathways were [a] isocitrate to 2-oxoglutarate or glyoxylate shunt, [b] oxidative phosphorylation, [c] 2-oxoglutarate to glutamate, [d] lipid metabolism, [e] S-Adenosylmethionine (SAM) metabolism, [f] valine, leucine, isoleucine metabolism, [g] propanoate metabolism, [h] pyruvate synthesis, [i] thiomine metabolism, [j] tryptophan/tyrosine metabolism and [k] pentose phosphate pathway (PPP)/metabolism of phosphoribosyl diphosphate (PRPP). Among genes with nonsynonymous or intergenic SNPs, genes encoding isocitrate lyase (SACE_1449), isocitrate dehydrogenase (SACE_6636) or 2-oxoglutarate dehydrogenase subunits (SACE_1638&6385) surround the node of 2-oxoglutarate. Under high-erythromycin producing conditions, higher carbon flux through isocitrate was drained into succinyl-CoA via 2-oxoglutarate compared to the flux into the glyoxylate shunt (Hong, Huang, Chu, Zhuang, & Zhang, 2016). We can assume that in E3 more carbon flux through isocitrate node flows towards succinyl-CoA and subsequently to methylmalonyl-CoA. The re-direction of carbon flow at the isocitrate node is likely accomplished by varying enzymatic activities owing to nonsynonymous SNPs. Another important SNP is the one in the intergenic region of SACE_3400, which encodes acetyl-CoA carboxylase (Acc) subunit. 7 genes (SACE_0934/1282/1341/1764/2786/4937/7125) involved in PPP or downstream purine metabolism are mutated, which indicated the importance of PPP or purine metabolism to the biosynthesis of erythromycin. There are also two nonsynonymous SNPs in erythromycin BGC. One nonsynonymous mutation occur inside SACE_0718 (eryCVI) and one intergenic mutation is observed in front of SACE_0720 (eryBIV). These two genes were also mutated in another high erythromycin producer, Px (Peano et al., 2012). Given the necessity of eryCVI and eryBIV for the biosynthesis of erythromycin, the two mutations are very likely to enhance the total enzymatic activities of their respective coding products (Summers et al., 1997).

Apart from the nonsynonymous mutations, 6 CDSs present nonsense SNPs, which likely blocked the transcription of the genes (Table 2). Two polyketide synthases (SACE_0019 & SACE_2875) are affected by nonsense mutations. SACE_0019 locates in pfa cluster, which appears to govern the biosynthesis of polyunsaturated fatty acids such as eicosapentaenoic acid (Oliynyk et al., 2007). SACE_2875 is part of the pks3gene cluster involved in the production of an aromatic polyketide antibiotic with low molecular weight. As a result, synthesis of these two kinds of secondary metabolites should be inactivated or diminished by nonsense mutations. The biosynthesis of erythromycin also depends on polyketide synthases (PKS), therefore the inactivation or diminishment of pks- related synthesis pathways may weaken precursor competitions between the synthesis of erythromycin and other polyketides.

Transcriptomic comparison of E3 and NRRL23338

Most physiological events originate at the transcriptional level. In order to characterize the impacts of genomic variations on gene expression profiles, both samples of NRRL23338 and E3 taken from the minimal liquid medium (Hopwood, 1985) were subjected to the high throughput RNAseq. Cell samples were harvested in exponential phase (10 h) and stationary phase (50 h) to extract RNA. 205 million reads or 107 million reads were generated for NRRL23338 or E3, respectively. After mapping all the reads to the respective genome, the coverage showed that the general transcription of core region was higher than that of the noncore region both in NRRL23338 and E3 (Figure S2).

A total of 822 genes were identified to be differentially expressed in the exponential phase. In order to obtain an overview of the general expression difference, we conducted functional enrichment analysis based on KEGG and GO annotations (Figure 3). KEGG enrichment analysis of differentially expressed genes (DEGs) showed that most affected pathways were [a] biosynthesis of 12-, 14- and 16-membered macrolides, [b] synthesis of type I polyketides, [c] polyketide sugar unit biosynthesis, [d] phosphotransferase system (PTS), [e] sesquiterpenoid and triterpenoid biosynthesis (Fig. 3a). The up-regulation of genes involved in [a] biosynthesis of 12-, 14- and 16-membered macrolides, [b] synthesis of type I polyketide sugar unit biosynthesis contribute directly to the biosynthesis of erythromycin. The significant change in expression of PTS genes suggested the unneglectable role of phosphate metabolism in regulating erythromycin

synthesis (Y. Xu, You, Yao, Chu, & Ye, 2019). On the contrary, purine/pyrimidine metabolism, biosynthesis of ribosome/peptidoglycan, translation and DNA replication/repair were down-regulated significantly in E3. This was in accordance with the repressed growth rates of E3. Besides the above pathways/functions, GO enrichment analysis implied that in E3 activities of oxidoreductase and NADH dehydrogenase were altered significantly (Fig. 3b), which hinted alterations regarding the intracellular level of cofactors.

Next, a genome-scale metabolic model (GSMM) was used to identify reporter metabolites, around which the most significant changes in expression occurred (Licona-Cassani, Marcellin, Quek, Jacob, & Nielsen, 2012). L-glutamate/L-glutamine and 2-oxoglutarate were observed among the high scoring reporter metabolites both in the exponential and the stationary phase, along with some precursors of erythromycin and molecules essential for electron transport (Figure 4). The production onset of erythromycin in *S. erythraea* has proven associated with nitrogen starvation (Zou, Hang, Chu, Zhuang, & Zhang, 2009a), while L-glutamate/L-glutamine and 2-oxoglutarate are tightly interconnected with nitrogen metabolism. Therefore, results of the reporter metabolites stressed the importance of nitrogen metabolism for the production of erythromycin.

Hierarchical clustering of DEGs was conducted to show alterations regarding specific subset of genes (Figure 5 & Table S4). It is noteworthy that all genes in erythromycin BGC and genes related to the biosynthesis of dTDP-sugar were drastically up-regulated in E3 in the exponential phase. Transcription of erythromycin BGC depends on the positive regulation by BldD and PhoP (Chng et al., 2008; Y. Xu et al., 2019). However, transcription of neither *bldD* nor *phoP* was stimulated in E3. This indicated that the regulatory networks affecting the transcription of erythromycin BGC changed in E3.

Transcription of 40 genes encoding (putative) transcriptional regulators were significantly changed in E3, 32 of which were up-regulated at both sampling points (Table S5). The transcription of SACE_4906 in E3 was stimulated most by 41.6 folds. The gene product of SACE_4906 comprises a cyclic N-Methyl pyrrolidone (cNMP) binding domain and is predicted as a gene involved in carbon catabolite repression. This implied that the substrate uptake system in E3 may vary from that of NRRL23338. In addition, two WhiB-like regulators encoded by SACE_5583 and SACE_6426 were stimulated significantly. In streptomycetes, WhiB-like regulators usually affect stress response and sporulation processes, which are interconnected with biosynthesis of secondary metabolites (Molle, Palframan, Findlay, & Buttner, 2000; Zheng, Long, & Xie, 2012). Although functions of the remaining transcriptional regulators are unknown, this analysis still narrowed the range of transcriptional regulator candidates, which probably have impact on the biosynthesis of erythromycin.

Transcriptional divergence was observed for genes involved in glycolysis and the TCA cycle (Figure S3). Expression of genes encoding the pyruvate dehydrogenase complex, which catalyzes the key step from glycolysis to the TCA cycle, and *pyc*responsible for the reaction from pyruvate to oxaloacetate were repressed. Genes encoding aconitate hydratase were down-regulated slightly. Expression of *icd*, *sucAB*, *korAB* and *sucCD* were up-regulated, whereas genes coding for succinate dehydrogenase and fumarate hydratase were repressed more significantly in E3. Given the transcriptional up-regulation of *mutAB*, carbon flux from isocitrate in E3 might flow into methylmalonyl-CoA metabolism via 2-oxoglutarate and then succinyl-CoA rather than to the dehydrogenation of succinate. Furthermore, *gabT* coding for the enzyme to catalyze conversion from 2-oxoglutarate to L-glutamate was down-regulated. Most of genes involved in the glutamate and glutamine metabolism were repressed, which could exert a nitrogen-starvation signal and triggered the biosynthesis commence of erythromycin (Z. Xu et al., 2019). As mentioned, nonsynonymous mutations were also identified in *icd* and *sucAB* (Table S3). Therefore, the genomic and transcriptomic variations regarding reactions from isocitrate to succinyl-CoA addressed the importance of these pathways in accumulating precursors of erythromycin.

Expression of genes involved in propanoate metabolism were generally much lower than those in TCA cycle both for NRRL23338 and E3 (Table S4), which indicated an inactive propanoate metabolism because the cultivation in the minimal medium had no propanol addition. However, *mmsA* and *mmsB* encoding enzymes to catalyze the degradation of branched amino acids into propionyl-CoA were significantly stimulated. This was beneficial for the biosynthesis of erythromycin in E3 (Z. Xu et al., 2018). All genes coding for acetyl-CoA carboxylase were repressed, and this implied that the lower accumulation of malonyl-CoA led to the reduced production of pigment by E3.

Since the degradation of fatty acids provides NADPH and coenzyme A, the up-regulation of genes related to the degradation of fatty acids benefits the production of erythromycin (Karnicar et al., 2016). SACE_-3084/6363 encoding acetyl-CoA acyltransferase and SACE_5380/5383 encoding acyl-CoA dehydrogenase were up-regulated significantly in E3 (Table S4). Another important source of NADPH is the oxidative pentose phosphate pathway (PPP), which partly connected with Entner-Doudoroff (ED) pathway. It is noteworthy that all genes involved in ED pathway were stimulated to different extents in E3. This expression change in the ED pathway was consistent with a previous conclusion drawn by¹³C labeling experiments that the ED pathway was the main glucose utilization pathway of E3 cultivated in minimal or chemically defined meidum (Hong et al., 2016). Relative to Embden-Meyerhof-Parnas pathway (EMP), the utilization of glucose predominantly through ED pathway not only increased the supply of NADPH for erythromycin production, but also decreased the generation of NADH, that exerts a negative effect on expression of erythromycin BGC via c-di-GMP (X. Li, Chu, & Jensen, 2020; Z. Xu et al., 2019). prsAwhich controls the conversion from PPP to purine/pyrimidine metabolism was down-regulated notably. Genes (purF / purH / pyrC) engaged in purine/pyrimidine metabolism, which is linked to the signal transduction through secondary messengers such as (p)ppGpp and c-di-GMP (Sivapragasam & Grove, 2019), were also transcriptionally down-regulated in E3 starting from the exponential phase (Table S4).

Transcriptional changes regarding specific genes with genomic variations

Interestingly, most of the doubled genes in Table S1 exhibited relatively low transcriptional levels both in NRRL23338 and E3 (Table 3). As an exception, SACE_0337 was expressed 4 to 6 times higher in the exponential phase in E3 compared to NRRL23338. Since in principle the accumulation of acetyl-CoA benefits the formation of precursors of erythromycin through the TCA cycle, the upregulation of SACE_0337 in E3 emphasized the importance of synthesis of acetyl-CoA for erythromycin production.

Apart from CDSs with large-fragment deletions (Table S2), 214 CDSs showed extremely low transcriptional levels close to zero, both in exponential and stationary phase (Table S6). 25 gene clusters in S. erythraea genome were responsible for biosynthesis of secondary metabolites (Oliynyk et al., 2007). However, we observed low-transcribed genes in 9 out of the 25 clusters, including rpp, nrps1, nrps4, nrps7, pks1, pks2 pks3, pks4 and hop clusters in E3 (Table S6). Particularly in the rpp cluster, that is responsible for the reddish pigment synthesis, several genes coding for acetyl-CoA carboxylase were transcribed lowly. This transcriptional repression of acc cluster along with the intergenic SNP around SACE_3400 turn less acetyl-CoA into malonyl-CoA and finally reduce the biosynthesis of the reddish pigment in E3 (Cortés et al., 2002). This was in line with the reduced production of the reddish pigment by E3. Therefore, it appeared that biosynthesis of secondary products other than erythromycin in E3 was more likely to decrease rather than to disappear. Secondary products, particularly Pks-related products, are assembled from some common starter units (Yuzawa, Keasling, & Katz, 2017). The low transcriptions of several pks clusters manifested themselves as engineering targets to enhance the production of erythromycin, as their synthesis may competent with erythromycin for precursors. Furthermore, low transcriptional levels were also observed in several genes of acyl-, methyl- or glycosyl-transferase. Biosynthesis of most secondary products requires post-PKS modification, e.g. methylation or glycosylation of the molecule backbone (Dhakal, Sohng, & Pandey, 2019). Thus, the transferase genes with low transcription were likely associated with modifications of the secondary products of the 9 clusters in Table S6. Taken together, we observed the genome variation and the low transcription of gene clusters responsible for biosynthesis of secondary metabolites other than erythromycin, which addressed the importance of construction of super-producer through genome minimization i.e., removal of gene clusters associated with other secondary metabolites (Weber, Charusanti, et al., 2015). Another 30 CDSs with low transcriptional levels were associated with transport processes, such as the transport of mannose, mannitol, arabinose, 1,4-digalacturonate or C4-dicarboxylate. Accordingly, two more genes (SACE_0179/0186) encoding arabinosyltransferases were also transcribed lowly. Transcriptional repression in these transporter genes of E3 implied distinguished preferences upon substrate, which corresponded to the repression on the biosynthesis of erythromycin by mannose, mannitol or arabinose with respect to glucose as the carbon source (El-Enshasy, Mohamed, Farid, & El-Diwany, 2008). SACE_2955-2956 encodes the bd type terminal oxidase of the electron transport chain (ETC), which exhibited a lower energetic efficiency compared to bc_1 - aa_3 complex (Scott, Salmon, & Poole, 1992). The low transcription of SACE_2955-2956 implied that alternative operons (SACE_0142-0143) encoding the bd complex were operative or E3 preferred an alternative electron transport channel, such as through bc_1 - aa_3 complex. The low transcription of SACE_3011, which encodes a copper resistance gene, fitted the conclusion drawn in our recent research that appropriate exogenous addition of copper ions would increase the production of erythromycin in E3 (Qiao, Li, Ke, & Chu, 2020).

Transcriptional levels of genes with nonsense mutations were all repressed significantly in E3, although their transcription in NRRL23338 were also pretty low (Table 4). Particularly, transcription of SACE_-0019/2875/3073/7243 were as low as they seemed to be deleted completely in E3 genome. The results confirmed again that nonsense mutations can be considered as gene deletion. It seemed that nonsense mutations in SACE_1076 and SACE_1257 played ambiguous effects on stopping expression by some unknown mechanisms. SACE_0019 encodes an *o* -succinylbenzoate-CoA ligase, which catalyzes a reaction consuming ATP and coenzyme A. SACE_3073 encodes a hydrogenase expression/formation protein HypD and SACE_7243 encodes a FAD-dependent oxygenase. Therefore, nonsense mutations in SACE_0019/3073/7243 indicated that the intracellular redox status might be changed in E3.

Putative mechanisms and promising molecular targets/strategies by which E3 enhances its biosynthesis of erythromycin

With the comparative omics analysis of *S. erythraea*, it is possible to propose putative mechanisms by which E3 enhances its production of erythromycin. In other words, these mechanisms provide promising molecular targets to engineer for the further enhancement of erythromycin production. Here, we concluded the cellular change of E3 compared to NRRL23338 at three aspects related to [a] intracellular biochemical environment, [b] signal transduction along with transcriptional regulation, and [c] the supply of precursors for erythromycin (Fig. 6).

Firstly, it is very likely that E3 created an intracellular pseudo-limited condition by genomic variations, which triggered the onset of secondary metabolism much earlier compared to NRRL23338. For instance, the transcriptional repression on glutamine/glutamate metabolism in E3 could exert a nitrogen-starvation signal (Figure 4 & Table S4), which benefits the earlier onset of the biosynthesis of erythromycin (Liao et al., 2015; Z. Xu et al., 2019). The extremely low transcription of transporter genes and the transcriptional change of PTS genes in E3 implied distinguished preferences of substrate utilization in E3 (Figure 3 & Table S6), which probably resulted in that E3 tended to have a different intracellular biochemical environment relative to NRRL23338 (Liao et al., 2015). In addition, one pair of bd type oxidase genes were lowly transcribed (Table S6), which can have a significant effect on the intracellular energy level (Figure S4) or redox status (Fischer, Falke, Naujoks, & Sawers, 2018). To engineer E3 for a better production of erythromycin, several strategies regarding the intracellular biochemical environment are worth trying, e.g., to repress the nitrogen metabolism by down-regulating the flux from 2-oxoglutarate to L-glutamate/glutamin, to optimize the components of industrial medium, or to directly regulate the intracellular energy or redox status. Based on the present mechanism, in our recent research we improved the production of erythromycin in E3 28% by enhancing the intracellular energy level (X. Li, Chen, Andersen, Chu, & Jensen, 2020) and 60% by reducing the intracellular redox status with more oxygen supply (X. Li, Chu, et al., 2020).

Secondly, E3 presented differently but unclearly altered regulatory networks which also contributed to the stimulation of erythromycin BGC transcription (Figure 5). E3 did not stimulate the transcription of the direct activator BldD or PhoP of erythromycin BGC (Chng et al., 2008; Y. Xu et al., 2019), which indicated a more complex regulatory network influencing the transcription of erythromycin BGC in E3. Regulator genes with significantly different transcriptional levels in Table S5 are candidates which can affect the transcription of erythromycin BGC. Besides, a two component system (TCS) regulating the copper homeostasis in S. erythraea was identified in our recent study based on the present comparative genomics (Qiao et al., 2020). Copper functioned as a cofactor of bc_1 - aa_3 terminal oxidase (Fujimoto, Chijiwa, Nishiyama, Takano, & Ueda, 2016). The addition of copper in the medium may increase the efficiency of ATP synthesis and then benifits

the biosynthesis of erythromycin. The genomic mutation in this TCS gene i.e. SACE_0101 enhanced the copper tolerance of E3 compared to NRRL23338. Gene deletion of SACE_0101 in NRRL23338 increased its erythromycin titer by 110% and the addition of copper enhanced the erythromycin titer of E3 by 17% (Qiao et al., 2020).

Thirdly, the supply of precursors for erythromycin appeared higher in E3. In general, the decreased biomass formation ensures that more carbon flows to secondary metabolism (Figure 1). The stimulated transcription of the ED pathway, enhancing the dehydrogenation of 2-oxoglutarate to succinyl-CoA and the degradation of branched amino acids, increased the supply of NADPH, propionyl-CoA and methylmalonyl-CoA, respectively, for the synthesis of erythromycin (Figure S3). In addition, repressed synthesis of other secondary metabolites in E3, such as pigment and other polyketides, drains more carbon flux to the biosynthesis of erythromycin (Table S6). The repressed synthesis of other secondary metabolites was also observed in another erythromycin overproducer Px (Peano et al., 2012).

Further improvement of erythromycin production in E3 by engineering the molecular targets

Specific genes functioning to supply precursors for erythromycin were chosen as targets in the present study for metabolic engineering to further enhance the production of erythromycin by the overproducer E3. 2-oxoglutarate was considered one of the reporter metabolites identified by transcriptional analysis between E3 and NRRL23338 (Fig. 4). Additionally, nonsynonymous or intergenic variations were also found in genes (*icd*, *sucA* and *sucB*) related to 2-oxoglutarate metabolism (Table S3). Several genes surrounding 2-oxoglutarate, i.e., *icd*, *sucA* and *sucB*, were stimulated in E3 with respect to NRRL2338 (Fig. 7a). As such, the metabolic flux from isocitrate to succinyl-CoA was supposed to be increased in E3. Although this still requires metabolomic data to support, metabolic engineering regarding 2-oxoglutarate was worthwhile.

Metabolic manipulations upon pathways around 2-oxoglutarate were performed by straightforwardly overexpressing *icd*, *sucA*and/or *sucB* in E3 (Fig. 7b). The overexpression of *icd* in E3 decreased the production of erythromycin by 45% and the seperate overexpression of *sucA* or *sucB* exhibited no significant effects on the production. Inspiringly,the simultaneous overexpression of *sucB* and *sucA* boosted the erythromycin titer by 40% in shake flask culture. Based on the result, bacth experiments were porformed in 5-L bioreactors with minimum essential medium (Fig. 8). Growth curves revealed that overexpression of sucBA led to a significant delaying on growth. However, the growth of E3::sucBA expedited with a faster rate after 36 h, and reached into the stationary phase at a similar time to E3. Similar trends were observed on the carbon dioxide evolution rate (CER) too. Fig.8c reveals that a rapid increase in CER occurred with both strains, with a lag of approximately 12h. After 60 h when both strains exhibited CER peaks, the CER of E3::sucBA droped rapidly, and maintained at lower levels than E3 during the stational phase. The change of CER indicated that the combinational expression of sucB and sucA accelerate the metabolic transition between exponential phase and stational phase of *S. erythraea*. The erythromycin titer of E3::sucBA was increased by 71% compated to E3, reaching to 967 mg/L. In consideration of glucose consumption rates, the overexpression of sucB and sucA led to a higher yield of erythromycin of E3::sucBA compared to E3.

The overexpression of *icd* is expected to enhance the intracellular 2-oxoglutarate pool generated from isocitrate . 2-oxoglutarate plays a role of indicating the cellular nitrogen status, and the anticipated larger 2-oxoglutarate pool in E3::icd relative to E3 signaled nitrogen deficiency that would lead to a stimulated nitrogen metabolism (Radchenko, Thornton, & Merrick, 2013), which exerts negative effects on the biosynthesis of erythromycin (Liao et al., 2015; Zhang et al., 2014). However, the overexpression of *sucBA* could avoid the accumulation of 2-oxoglutarate. 2-oxoglutarate dehydrogenase contributes to the degeneration of 2-oxoglutarate to succinyl-CoA. Via the catalysis by the methylmalonyl-CoA mutase, succinyl-CoA is then converted to methylmalonyl-CoA and subsequently to the biosynthesis of erythromycin. The anticipated increase of CoA-related precursors for erythromycin and decrease of 2-oxoglutarate for nitrogen metabolism might account for the titer enhancement and growth retardation by *sucBA* overexpression, respectively. It suggests that the overexpression of *sucBA* effectively reproduced the evolution between the wild type NRRL23338 and industrial overproducer E3.

Conclusions

In the present study, we re-sequenced the whole genome of an erythromycin overproducer *S. erythraea* HL3168 E3 and carried out comparative genomic and transcriptomic analysis of E3 with respect to the wild type NRRL23338. Genomic variations occurred in E3 in terms of gene transposition, fragment deletions in 56 sequences and 255 SNPs, which were partly distinguished from a previous E3 genome obtained via pyrosequencing (Y. Li et al., 2013). Comparative omics analysis revealed that the phenotypic difference between E3 and NRRL23338 resulted mainly from genomic and transcriptomic changes in central carbon and nitrogen metabolism, biosynthesis of secondary metabolites, and cofactor metabolism.

Within transcriptomic analysis the top ranking reporter metabolites were L-glutamine/L-glutamate and 2oxoglutarate, which pointed to the important role of nitrogen metabolism and the signal transduction in the biosynthesis of erythromycin. Genes coding for direct regulators of erythromycin biosynthesis gene cluster i.e. BldD and PhoP in E3 were not stimulated, we predicted that the regulatory networks of erythromycin BGC in E3 might be different from our anticipation. Expression of genes involved in the TCA cycle, ED pathway and the degradation of fatty acid changed in E3, which facilitated to supply more precursors for the biosynthesis of erythromycin. Moreover, several genes involved in biosynthesis of secondary metabolites other than erythromycin were expressed at extremely low levels in E3, which addressed the importance of construction of super-producer through genome minimization (Peano et al., 2007).

Based on the omics analysis, we proposed putative mechanisms by which E3 boosted the biosynthesis of erythromycin. The mechanisms provided molecular targets to strengthen the productivity of erythromycin by E3. As an example, we enhanced the supply of erythromycin precursors by manipulating targets derived from one of the proposed mechanisms. The overexpression of sucBA resulted in an enhancement of the erythromycin titer by 71% in E3. The present study showed that comparative omics analysis could reveal the differences between the physiologies of wild type strains and their overproducing derivatives, and also point out strategies for enhancing the production of secondary metabolites in overproducers even further.

Materials and methods

Strains, plasmids and primers

The wild type NRRL23338 and an industrial mutant HL3168 E3 were compared in this study (Table S7). E3 was obtained through classical mutate-and-screen method by our group. Primers utilized in this study were listed (Table S8).

Genetic manipulations

The genome of E3 was used as the template to amplify genetic fragments. The coding sequence of *icd*, *sucA* and *sucB* were amplified with primers icdF/icdR, sucAF/sucAR and sucBF/sucBR, respectively. The fragments were inserted into pIB139 between Nde I and Xba I sites using gibson assembly, yielding plasmid pIB-icd, pIB-sucA and pIB-sucB, respectively. Primers sucAF3 and sucAR3 were used for the amplification of leader and coding sequence of *sucA*. The resulting fragment was next cloned into pIB-sucB at *EcoR* I via gibson assembly, yielding the plasmid pIB-sucBA.

Culture conditions of S. erythraea

XM solid medium was used for the growth of *S. erythraea* spores which were incubated at 34 °C for 6 days. About 1 cm² XM solid medium covered with dense spores was picked into 300 mL flasks with 30 mL preculture medium. The component of XM solid medium and preculture medium were described previously (C. Chen et al., 2017). After a preculture of 48 h, 3 mL seed culture was collected, resuspended in 1 mL PBS buffer (pH=7.4) and inoculated into flasks with 27 mL modified minimal liquid medium for phenotype characterization. The modified minimal liquid medium (MMLM) contained the following per 800 mL: 2 g (NH₄)₂SO₄, 5 g casamino acids (DifcoTM, BD), 0.6 g MgSO₄[?]7H₂O, 0.001 g ZnSO₄[?]7H₂O, 0.001 g MnCl₂[?]4H₂O, 0.001 g CaCl₂. 800 mL MMLM was then dispensed in 80 ml aliquots. 15 mL NaH₂PO₄/K₂HPO₄buffer (0.1 M, pH=6.8) was added in 80 mL liquid medium for pH

buffering. After autoclaving, 25% (w/v) glucose as sole carbon source was injected in the medium to a final concentration 20 g/L. Liquid culture in flasks was performed in shaking incubator with an agitation speed of 220 rpm at 34 degC. The fermentations of E3 and E3::sucBA were carried out in a 5-L fermenter (Shanghai Guoqiang Bioengineering Equipment Co., Ltd., China). Samples were taken every 12 or 24 h for monitoring cell growth and erythromycin production. CER were determined by process mass spectrometer.

Phenotype characterization

Cell samples from cultures in duplicate or triplicate shake flasks were collected at various time points throughout the fermentation. Dry cell weight (DCW) in 3 mL culture was measured after filtration for monitoring cell growth. The concentration of residual glucose in culture was determined by salicylic acid method (Zou, Hang, Chu, Zhuang, & Zhang, 2009b). The erythromycin titer was measured by the modified colorimetric method (Zou et al., 2009a).

Extraction of the whole genome and total RNA

Single colony was picked up from XM solid medium and inoculated into TSB liquid medium (Sigma, Germany) in flasks. 5 mL cell samples were collected after the shaking incubation at 34 degC for 48⁷² h and were used for genome extraction. 400 to 600 µL TSB culture was centrifuged at 12000 rpm for 3 min for pellet collection. Pellets were then washed by 600 µL 0.05 M N-[tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid sodium salt (TES) buffer (pH=8.0) twice and resuspended in 700 µL TES buffer. 50 mg/mL of lysozyme and 4 mg/mL RNase were sequentially added to a final concentration of 5 mg/mL or 40 µg/mL, respectively. Pellets with enzymes were incubated at 37 °C for at least 1 h. The tubes should be gently inverted 4 to 5 times during incubation until the liquid became sticky. 10% sodium dodecyl sulfate was added to a final concentration of 1²2%. 5 µL of 20 mg/mL protease K was also needed for protein denaturation. The mixture was then incubated at 37 °C for 30 min, during which the tubes were vigorously inverted 2 to 3 times until the liquid was transparent. To precipitate the protein, 100 µL of 5 M NaCl and 50 µL of 10% CTAB preheated to 65 °C were added in tubes. The mixture was then incubated at 65 °C for 1 h. The tubes were inverted each 15 min. To extract DNA from the liquid, equivalent volume of phenol:chloroform: isoamyl alcohol (25:24:1) saturated with 10 mM Tris, pH 8.0, 1 mM EDTA was used and then vibrated in vortex for 15 s. The mixture was centrifuged at 12000 rpm for 10 min. Clear supernatant was transferred into new tubes and extraction by phenol:chloroform: isoamyl alcohol (25:24:1) was repeated 3 times. To precipitate DNA, 3 M NaAc was added to a final concentration of 0.3 M. 4 °C isopropanol was added before DNA samples were kept in -20 °C overnight. The mixture was then centrifuged at 12000 rpm for 1 min at 4 °C and the supernatant was discarded. 2 mL cold ethanol was used to wash samples and the supernatant was discarded after 12000 rpm centrifugation for 10 min. 1 mL cold 70% ethanol was then added and supernatant was also discarded. At last, 30 to 50 µL TE buffer preheated to 55 °C was added in tubes to dissolve genome. Concentration of genome was measured by NanoDrop spectro-photometer. Genome sequencing was performed on Illumina Hiseq2000 platform (BGI, China).

For RNA isolation, two replicates from independent cultures in modified minimal liquid cultures were used. Cells were harvested at two time points, i.e. 10 h in the early exponential phase and 50 h at the onset point of erythromycin biosynthesis, for total RNA extraction. 10 mL culture was centrifuged at $4000 \times g$ at 4 °C for 10 min. The broth supernatant was discarded and the cells were resuspended in 2 mL RNAlaterTM solution (Invitrogen, USA). Total RNA was extracted with an RNeasy Plus Mini kit (Qiagen, Germany) using glass beads to mechanically disrupting cells with a FastPrep (30 s ×4; M.P. Biomedical, USA). DNase treatment by RNase-Free DNase Set (Qiagen, Germany) aided to digest DNA in samples. The RNA integrality was analyzed by 1% agarose gel electrophoresis and Bioanalyzer (Agilent, USA). The RNA quality was also determined by Bioanalyzer (Agilent, USA). 1 µg total RNA was reverse transcribed using a PrimeScriptTM RT Reagent Kit with gDNA Eraser (Takara, Japan) for RT-qPCR. RNA sequencing was accomplished using BGIseq 500 next-gen sequencer or Illumina Hiseq4000 sequencer, which was commercially provided by BGI, China or Shanghai Majorbio Bio-pharm Technology Co.,Ltd, respectively. The original image data obtained from sequencers was transferred into sequence data via base calling, which is defined as raw data or raw reads and saved as FASTQ file. Datasets consisted of at least 30 M pair-ended reads per sample with a 100-

or 150-bp read length. After sequencing, the raw reads were filtered includes removing adaptor sequences, contamination and low-quality reads from raw reads. Especially for genome raw reads, reads with a large amount of duplications or N base were also removed.

Raw reads assembly and analysis of E3 genome sequence

For genome sequence, software SOAPdenovo2.04 and Platanus1.2.4 was used for reads assembly. After mapping reads to contigs, the assembly was optimized according to relation between paired-end and overlapping. Software FGAP was then used to fix gaps in the genome.

After genome sequencing and assembly, CDSs were predicted using software Glimmer3.02. Annotation of CDSs were conducted based on blast of protein sequences in database, i.e. KEGG, COG, SwissProt, TrEMBL, phi and iprscan. Repetitive sequences were recognized by blasting contigs to transposon database by software RepeatMasker and TRF (tandem repeat finder). MUMmer was used for SNP calling by comparing E3 genome to the reference genome of NRRL23338. After SNP calling, SNP filtration was performed by removing SNPs with low quality. LASTZ1.01.50 was adopted to identify InDel mutations in E3 genome.

RNA raw reads assembly and analysis for NRRL23338 and E3

RNA sequencing raw reads were uploaded into the software Geneious Prime. After trimming and filtering using the built-in function in Geneious Prime, all the raw reads were mapped into their respective reference genome of NRRL23338 or E3. Expression levels of each CDS were then calculated by Geneious Prime. DEseq2 method was used for differentially expression analysis.

Clusters of orthologous identified by NCBI batch group were entrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez) and EggNOG database (Huerta-Cepas et al., 2016). iPATH helped visualize the distribution of DEGs over different metabolic pathways (Darzi, Letunic, Bork, & Yamada, 2018). Gene set analysis and reporter metabolites analysis were conducted by R package Piano (Varemo, Nielsen, & Nookaew, 2013) based on a genome-scale metabolic model (11). Hierarchical clustering of DEGs was carried out by a web-based tool, Morpheus (https://software.broadinstitute.org/GENE-E).

Data availability

This genome sequence of E3 has been deposited at DDBJ/ENA/GenBank under the accession JABNNH000000000. Raw reads of RNA sequencing for E3 and NRRL23338 used in this study can be obtained from the NCBI Gene Expression Omnibus (accession number: GSE134767 for E3 and GSE150974 for NRRL23338).

Abbreviations of gene names

pyc, SACE_6118, pyruvate carboxylase; acn, SACE_3754/3811/4339, aconitate hydratase; icd, SACE_6636, isocitrate dehydrogenase; korA, SACE_3927, 2-oxoglutarate ferredoxin oxidoreductase subunit alpha; korB, SACE_3926, 2-oxoglutarate ferredoxin oxidoreductase subunit beta; sucC, SACE_6669, succinyl-CoA synthetase subunit beta; sucD, SACE_6668, succinyl-CoA synthetase subunit alpha; sdh, SACE_1170/1171/6582/6583/6584/6585, succinate dehydrogenase; fum, SACE_0930/1784, fumarate hydratase; gabT, SACE_6022, 4-aminobutyrate aminotransferase; mmsA, SACE_1456, methylmalonate-semialdehyde dehydrogenase; mmsB, SACE_1459, 3-hydroxyisobutyrate dehydrogenase; prsA, SACE_2398/0316, ribose-phosphate pyrophosphokinase; purF, SACE_7125, amidophosphoribosyltransferase; purH, SACE_6664, bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase; pyrC, SACE_2080, dihydroorotase; rpp, red-brown pigment-producing; nrps, nonribosomal peptide synthetase; hop, hopanoids.

Author information

Corresponding authors

Ju Chu: juchu@ecust.edu.cn; Tel. (+86) 021-64253021

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Corresponding authors

All authors read and approved the manuscript, contributed significantly to the work, and conceived the project. XL and JC designed the experiments. XL, YS and JC analyzed the results. XL wrote the manuscript with the help of JC. XL, XK and LQ performed the experiments and supported by JC.

Acknowledgements

This work was financially supported by the National Key Research Development Program of China (No. 2019YFA0904800), the National Natural Science Foundation of China (No. 21276081) and the National Scientific and Technological Major Special Project (Significant Creation of New drugs, No. 2011ZX09203-001-03). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

We thank Kuipu Zhang from East China University of Science and Technology for help with graphics and Robin Dorau from Technical University of Denmark for constructive criticism of the manuscript.

Supplemental Materials

Figure S1 Mapping nonsynonymous/synonymous within CDSs and variations in intergenic regions to metabolism pathways by iPATH.

Figure S2 Transcription coverage after mapping all raw RNA reads to the genome.

Figure S3 Visualization of differences in central carbon metabolism at transcription level.

Figure S4 Comparison of energy level in E3 relative to NRRL23338.

Table S1 Repeated sequences in E3

Table S2 All SNP in E3 with respect to NRRL23338

Table S3 Hierarchical clustering of other key subgroups of genes rather than genes in BGCs

Table S4 Large fragment deletion in E3

Table S5 CDSs with low transcriptional coverage

Table S6 Regulator genes with significant expressional change at 10 h

Table S7 Strains and plasmids used in this study

Table S8 Primers used in this study

Reference

Andreani, J., Le Bideau, M., Duflot, I., Jardot, P., Rolland, C., Boxberger, M., . . . Raoult, D. (2020). In vitro testing of combined hydroxychloroquine and azithromycin on SARS-CoV-2 shows synergistic effect. *Microbial Pathogenesis*, 145 . doi: 10.1016/j.micpath.2020.104228

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. *Appl Environ Microbiol, 68* (10), 4731-4739. doi: 10.1128/aem.68.10.4731-4739.2002

Carata, E., Peano, C., Tredici, S. M., Ferrari, F., Tala, A., Corti, G., . . . Alifano, P. (2009). Phenotypes and gene expression profiles of Saccharopolyspora erythraea rifampicin-resistant (rif) mutants affected in erythromycin production. *Microb Cell Fact, 8*, 18. doi: 10.1186/1475-2859-8-18

Chen, C., Hong, M., Chu, J., Huang, M., Ouyang, L., Tian, X., & Zhuang, Y. (2017). Blocking the flow of propionate into TCA cycle through a mutB knockout leads to a significant increase of erythromycin production by an industrial strain of Saccharopolyspora erythraea. *Bioprocess Biosyst Eng*, 40 (2), 201-209. doi: 10.1007/s00449-016-1687-5

Chen, Y., Huang, M., Wang, Z., Chu, J., Zhuang, Y., & Zhang, S. (2013). Controlling the feed rate of glucose and propanol for the enhancement of erythromycin production and exploration of propanol metabolism fate by quantitative metabolic flux analysis. *Bioprocess Biosyst Eng*, 36 (10), 1445-1453. doi: 10.1007/s00449-013-0883-9

Chng, C., Lum, A. M., Vroom, J. A., & Kao, C. M. (2008). A key developmental regulator controls the synthesis of the antibiotic erythromycin in Saccharopolyspora erythraea. *Proc Natl Acad Sci U S A*, 105 (32), 11346-11351. doi: 10.1073/pnas.0803622105

Cortés, J., Velasco, J., Foster, G., Blackaby, A. P., Rudd, B. A., & Wilkinson, B. (2002). Identification and cloning of a type III polyketide synthase required for diffusible pigment biosynthesis in Saccharopolyspora erythraea. *Mol Microbiol*, 44 (5), 1213-1224. doi: doi.org/10.1046/j.1365-2958.2002.02975.x

Darzi, Y., Letunic, I., Bork, P., & Yamada, T. (2018). iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res*, 46 (W1), W510-W513. doi: 10.1093/nar/gky299

Dhakal, D., Sohng, J. K., & Pandey, R. P. (2019). Engineering actinomycetes for biosynthesis of macrolactone polyketides. *Microb Cell Fact, 18* (1), 137. doi: 10.1186/s12934-019-1184-z

El-Enshasy, H. A., Mohamed, N. A., Farid, M. A., & El-Diwany, A. I. (2008). Improvement of erythromycin production by Saccharopolyspora erythraea in molasses based medium through cultivation medium optimization. *Bioresour Technol*, 99 (10), 4263-4268. doi: 10.1016/j.biortech.2007.08.050

Ferro, A. M., Ramos, P., Guerreiro, O., Jeronimo, E., Pires, I., Capel, C., . . . Goncalves, S. (2017). Impact of novel SNPs identified in Cynara cardunculus genes on functionality of proteins regulating phenylpropanoid pathway and their association with biological activities. *BMC Genomics*, 18 (1), 183. doi: 10.1186/s12864-017-3534-8

Fischer, M., Falke, D., Naujoks, C., & Sawers, R. G. (2018). Cytochrome bd oxidase has an important role in sustaining growth and development of Streptomyces coelicolor A3 (2) under oxygen-limiting conditions. *J Bacteriol, 200* (16), e00239-00218.

Fujimoto, M., Chijiwa, M., Nishiyama, T., Takano, H., & Ueda, K. (2016). Developmental defect of cytochrome oxidase mutants of Streptomyces coelicolor A3(2). *Microbiology*, 162 (8), 1446-1455. doi: 10.1099/mic.0.000332

Hong, M., Huang, M., Chu, J., Zhuang, Y., & Zhang, S. (2016). Impacts of proline on the central metabolism of an industrial erythromycin-producing strain Saccharopolyspora erythraea via (13)C labeling experiments. *J Biotechnol, 231*, 1-8. doi: 10.1016/j.jbiotec.2016.05.026

Hopwood, D. A. (1985). Genetic manipulation of Streptomyces: a laboratory manual.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., . . . Kuhn, M. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*, 44 (D1), D286-D293. doi: doi.org/10.1093/nar/gkv1248

Karnicar, K., Drobnak, I., Petek, M., Magdevska, V., Horvat, J., Vidmar, R., . . . Petkovic, H. (2016). Integrated omics approaches provide strategies for rapid erythromycin yield increase in Saccharopolyspora erythraea. *Microb Cell Fact, 15*, 93. doi: 10.1186/s12934-016-0496-5

Kiss, J., Szabo, M., & Olasz, F. (2003). Site-specific recombination by the DDE family member mobile element IS30 transposase. *Proc Natl Acad Sci U S A*, 100 (25), 15000-15005. doi: 10.1073/pnas.2436518100

Li, X., Chen, J., Andersen, J. M., Chu, J., & Jensen, P. R. (2020). Cofactor engineering redirects secondary metabolism and enhances erythromycin production in Saccharopolyspora erythraea. ACS Synth Biol, 9 (3), 655-670. doi: 10.1021/acssynbio.9b00528

Li, X., Chu, J., & Jensen, P. R. (2020). The expression of NOX from synthetic promoters reveals an important role of the redox status in regulating secondary metabolism of Saccharopolyspora erythraea. *Front Bioeng Biotechnol, 8*, 818. doi: 10.3389/fbioe.2020.00818

Li, Y., Chang, X., Yu, W., Li, H., Ye, Z., Yu, H., . . . Ye, B. (2013). Systems perspectives on erythromycin biosynthesis by comparative genomic and transcriptomic analyses of S. erythraea E3 and NRRL23338 strains. *BMC Genomics*, 14 (1), 523. doi: doi.org/10.1186/1471-2164-14-523

Liao, C., Yao, L., Xu, Y., Liu, W., Zhou, Y., & Ye, B. (2015). Nitrogen regulator GlnR controls uptake and utilization of non-phosphotransferase-system carbon sources in actinomycetes. *Proc. Natl. Acad. Sci. U. S.* A., 112 (51), 15630-15635. doi: doi:10.1073/pnas.1508465112

Licona-Cassani, C., Marcellin, E., Quek, L. E., Jacob, S., & Nielsen, L. K. (2012). Reconstruction of the Saccharopolyspora erythraea genome-scale model and its use for enhancing erythromycin production. *Antonie van Leeuwenhoek*, 102 (3), 493-502. doi: 10.1007/s10482-012-9783-2

Liu, J., Chen, Y., Wang, W., Ren, M., Wu, P., Wang, Y., . . . Zhang, B. (2017). Engineering of an Lrp family regulator SACE_Lrp improves erythromycin production in Saccharopolyspora erythraea. *Metab Eng*, 39, 29-37. doi: 10.1016/j.ymben.2016.10.012

Liu, Y., Ren, C. Y., Wei, W. P., You, D., Yin, B. C., & Ye, B. C. (2019). A CRISPR-Cas9 Strategy for Activating the Saccharopolyspora erythraea Erythromycin Biosynthetic Gene Cluster with Knock-in Bidirectional Promoters. *ACS Synth Biol*, 8 (5), 1134-1143. doi: 10.1021/acssynbio.9b00024

Lum, A. M., Huang, J., Hutchinson, C. R., & Kao, C. M. (2004). Reverse engineering of industrial pharmaceutical-producing actinomycete strains using DNA microarrays. *Metab Eng*, 6 (3), 186-196. doi: 10.1016/j.ymben.2003.12.001

Manteca, A., & Yague, P. (2018). Streptomyces Differentiation in Liquid Cultures as a Trigger of Secondary Metabolism. *Antibiotics (Basel)*, 7 (2). doi: 10.3390/antibiotics7020041

Marcellin, E., Mercer, T. R., Licona-Cassani, C., Palfreyman, R. W., Dinger, M. E., Steen, J. A., . . . Nielsen, L. K. (2013). Saccharopolyspora erythraea's genome is organised in high-order transcriptional regions mediated by targeted degradation at the metabolic switch. *BMC Genomics*, 14 (1), 15. doi: doi.org/10.1186/1471-2164-14-15

Mironov, V., Sergienko, O., Nastasyak, I., & Danilenko, V. (2004). Biogenesis and regulation of biosynthesis of erythromycins in Saccharopolyspora erythraea. *Appl. Biochem. Microbiol.*, 40 (6), 531-541. doi: doi.org/10.1023/B:ABIM.0000046985.66328.7a

Molle, V., Palframan, W. J., Findlay, K. C., & Buttner, M. J. (2000). WhiD and WhiB, homologous proteins required for different stages of sporulation in Streptomyces coelicolor A3 (2). *J Bacteriol, 182* (5), 1286-1295. doi: Biogenesis and regulation of biosynthesis of erythromycins in Saccharopolyspora erythraeadoi: 10.1128/JB.182.5.1286-1295.2000

Nagy, Z., & Chandler, M. (2004). Regulation of transposition in bacteria. *Res Microbiol*, 155 (5), 387-398. doi: 10.1016/j.resmic.2004.01.008

Nakken, S., Alseth, I., & Rognes, T. (2007). Computational prediction of the effects of non-synonymous single nucleotide polymorphisms in human DNA repair genes. *Neuroscience*, 145 (4), 1273-1279. doi: 10.1016/j.neuroscience.2006.09.004

Newman, D. J., & Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod*, 79 (3), 629-661. doi: 10.1021/acs.jnatprod.5b01055

Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., . . . Leadlay, P. F. (2007). Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. *Nat Biotechnol*, 25 (4), 447-453. doi: 10.1038/nbt1297

Peano, C., Bicciato, S., Corti, G., Ferrari, F., Rizzi, E., Bonnal, R. J., . . . De Bellis, G. (2007). Complete gene expression profiling of Saccharopolyspora erythraea using GeneChip DNA microarrays. *Microb Cell Fact*, 6, 37. doi: 10.1186/1475-2859-6-37

Peano, C., Damiano, F., Forcato, M., Pietrelli, A., Palumbo, C., Corti, G., . . . Alifano, P. (2014). Comparative genomics revealed key molecular targets to rapidly convert a reference rifamycin-producing bacterial strain into an overproducer by genetic engineering. *Metab Eng*, 26, 1-16. doi: 10.1016/j.ymben.2014.08.001

Peano, C., Tala, A., Corti, G., Pasanisi, D., Durante, M., Mita, G., . . . Alifano, P. (2012). Comparative genomics and transcriptional profiles of Saccharopolyspora erythraea NRRL 2338 and a classically improved erythromycin over-producing strain. *Microb Cell Fact, 11*, 32. doi: 10.1186/1475-2859-11-32

Qiao, L., Li, X., Ke, X., & Chu, J. (2020). A two-component system gene SACE_0101 regulates copper homeostasis in Saccharopolyspora erythraea. *Bioresources and Bioprocessing*, 7 (1), 12. doi: 10.1186/s40643-020-0299-8

Radchenko, M. V., Thornton, J., & Merrick, M. (2013). P(II) signal transduction proteins are ATPases whose activity is regulated by 2-oxoglutarate. *Proc Natl Acad Sci U S A*, 110 (32), 12948-12953. doi: 10.1073/pnas.1304386110

Redenbach, M., Scheel, J., & Schmidt, U. (2000). Chromosome topology and genome size of selected actinomycetes species. *Antonie van Leeuwenhoek*, 78 (3-4), 227-235. doi: doi.org/10.1023/A:1010289326752

Reeves, A. R., Brikun, I. A., Cernota, W. H., Leach, B. I., Gonzalez, M. C., & Weber, J. M. (2006). Effects of methylmalonyl-CoA mutase gene knockouts on erythromycin production in carbohydrate-based and oil-based fermentations of Saccharopolyspora erythraea. *J Ind Microbiol Biotechnol, 33* (7), 600-609. doi: 10.1007/s10295-006-0094-3

Reeves, A. R., Brikun, I. A., Cernota, W. H., Leach, B. I., Gonzalez, M. C., & Weber, J. M. (2007). Engineering of the methylmalonyl-CoA metabolite node of Saccharopolyspora erythraea for increased erythromycin production. *Metab Eng*, 9 (3), 293-303. doi: 10.1016/j.ymben.2007.02.001

Sayed, A. M., Abdel-Wahab, N. M., Hassan, H. M., & Abdelmohsen, U. R. (2019). Saccharopolyspora: an underexplored source for bioactive natural products. *J Appl Microbiol* . doi: 10.1111/jam.14360

Scott, R. I., Salmon, I., & Poole, R. K. (1992). The cytochromes of the filamentous bacteria Streptomyces clavuligerus and Saccharopolyspora erythraea (FormerlyStreptomyces erythraeus). *Curr Microbiol*, 24 (2), 105-109. doi: doi.org/10.1007/BF01570906

Sivapragasam, S., & Grove, A. (2019). The Link between Purine Metabolism and Production of Antibiotics in Streptomyces. *Antibiotics (Basel)*, 8 (2). doi: 10.3390/antibiotics8020076

Smolke, C. D., & Keasling, J. D. (2002). Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. *Biotechnol Bioeng*, 80 (7), 762-776. doi: 10.1002/bit.10434

Summers, R. G., Donadio, S., Staver, M. J., Wendt-Pienkowski, E., Hutchinson, C. R., & Katz, L. (1997). Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of Saccharopolyspora erythraea that are involved in L-mycarose and D-desosamine production. *Microbiology*, 143 (10), 3251-3262.

Tong, Y., Charusanti, P., Zhang, L., Weber, T., & Lee, S. Y. (2015). CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. ACS Synth Biol, 4 (9), 1020-1029. doi: 10.1021/acssynbio.5b00038

Varemo, L., Nielsen, J., & Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res*, 41 (8), 4378-4391. doi: 10.1093/nar/gkt111

Wang, W., Li, S., Li, Z., Zhang, J., Fan, K., Tan, G., . . . Zhang, L. (2020). Harnessing the intracellular triacylglycerols for titer improvement of polyketides in Streptomyces. *Nat. Biotechnol.*, 38 (1), 76-83. doi: doi.org/10.1038/s41587-019-0335-4

Wang, Y., Wang, Y., Chu, J., Zhuang, Y., Zhang, L., & Zhang, S. (2007). Improved production of erythromycin A by expression of a heterologous gene encoding S-adenosylmethionine synthetase. *Appl Microbiol Biotechnol*, 75 (4), 837-842. doi: 10.1007/s00253-007-0894-z

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., . . . Medema, M. H. (2015). antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*, 43 (W1), W237-243. doi: 10.1093/nar/gkv437

Weber, T., Charusanti, P., Musiol-Kroll, E. M., Jiang, X., Tong, Y., Kim, H. U., & Lee, S. Y. (2015). Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes. *Trends Biotechnol*, 33 (1), 15-26. doi: 10.1016/j.tibtech.2014.10.009

Xu, Y., You, D., Yao, L. L., Chu, X., & Ye, B. C. (2019). Phosphate regulator PhoP directly and indirectly controls transcription of the erythromycin biosynthesis genes in Saccharopolyspora erythraea. *Microb Cell Fact*, 18 (1), 206. doi: 10.1186/s12934-019-1258-y

Xu, Z., Liu, Y., & Ye, B. C. (2018). PccD Regulates Branched-Chain Amino Acid Degradation and Exerts a Negative Effect on Erythromycin Production in Saccharopolyspora erythraea. *Appl Environ Microbiol, 84* (8). doi: 10.1128/AEM.00049-18

Xu, Z., You, D., Tang, L. Y., Zhou, Y., & Ye, B. C. (2019). Metabolic engineering strategies based on secondary messengers (p)ppGpp and C-di-GMP to increase erythromycin yield in Saccharopolyspora erythraea. ACS Synth Biol, 8 (2), 332-345. doi: 10.1021/acssynbio.8b00372

Yuzawa, S., Keasling, J. D., & Katz, L. (2017). Bio-based production of fuels and industrial chemicals by repurposing antibiotic-producing type I modular polyketide synthases: opportunities and challenges. J Antibiot (Tokyo), 70 (4), 378-385. doi: 10.1038/ja.2016.136

Zeng, W., Guo, L., Xu, S., Chen, J., & Zhou, J. (2020). High-Throughput Screening Technology in Industrial Biotechnology. *Trends Biotechnol*. doi: 10.1016/j.tibtech.2020.01.001

Zhang, Q., Chen, Y., Hong, M., Gao, Y., Chu, J., Zhuang, Y.-p., & Zhang, S.-l. (2014). The dynamic regulation of nitrogen and phosphorus in the early phase of fermentation improves the erythromycin production by recombinant Saccharopolyspora erythraea strain. *Bioresources and Bioprocessing*, 1 (1), 15. doi: doi.org/10.1186/s40643-014-0015-7

Zheng, F., Long, Q., & Xie, J. (2012). The function and regulatory network of WhiB and WhiB-like protein from comparative genomics and systems biology perspectives. *Cell Biochem Biophys*, 63 (2), 103-108. doi: 10.1007/s12013-012-9348-z

Zhuang, Z., Huang, M., & Chu, J. (2018). In silico reconstruction and experimental validation of Saccharopolyspora erythraea genome-scale metabolic model iZZ1342 that accounts for 1685 ORFs. *Bioresour Bioprocess*, 5 (1), 26. doi: doi.org/10.1186/s40643-018-0212-x

Zhuo, Y., Zhang, W., Chen, D., Gao, H., Tao, J., Liu, M., . . . Zhang, Q. (2010). Reverse biological engineering of hrdB to enhance the production of avermeetins in an industrial strain of Streptomyces avermitilis. *Proceedings of the National Academy of Sciences*, 107 (25), 11250-11254. doi: doi:10.1073/pnas.1006085107

Zou, X., Hang, H.-f., Chu, J., Zhuang, Y.-p., & Zhang, S.-l. (2009a). Enhancement of erythromycin A production with feeding available nitrogen sources in erythromycin biosynthesis phase. *Bioresour Technol*, 100 (13), 3358-3365. doi: doi.org/10.1016/j.biortech.2009.01.064

Zou, X., Hang, H.-f., Chu, J., Zhuang, Y.-p., & Zhang, S.-l. (2009b). Oxygen uptake rate optimization with nitrogen regulation for erythromycin production and scale-up from 50 L to 372 m3 scale. *Bioresour Technol*,

Component of the genome	NRRL 23338	E3
Length	8,212,805 bp	8,181,083 bp
G+C content	71.10%	71.18%
Coding density	84.90%	87.19%
Coding sequences	7264	7669
rRNA	16	12
tRNA	50	50
$CDSs^*$	7198	7607
Average CDS length	968 bp	$931 \mathrm{\ bp}$

Table 1. Genomic features of NRRL23338 and E3

* CDSs: coding sequences

Table 2. Nonsense SNPs in E3 with respect to NRRL23338

Old locus tag	$\begin{array}{l} {\rm NRRL23338_base^*} \\ {\rm <=> E3_base^{**}} \end{array}$	NRRL23338 codon*** <=> E3_codon****	Gene products
SACE_0019	C<=>T	CAG<=>TAG	AMP-dependent synthetase/ligase
SACE_1076	$C \ll T$	TGG<=>TAG	chitinase
SACE_1257	$G \ll A$	CAG<=>TAG	hypothetical protein
SACE_2875	G<=>A	TGG<=>TGA	modular polyketide synthase
SACE_3073	C<=>G	TAC<=>TAG	hydrogenase expression protein HypD
SACE_7243	C<=>T	TGG<=>TGA	FAD-dependent oxygenase

Table 3. Transcriptional comparison of doubled genes in E3 with respect to NRRL23338

Locus tag	TPM value*	TPM value*	TPM value*	TPM value*	Product
	WT_{-10}	WT_{-50}	$E3_{10}$	$E3_{50}$	
$SACE_{-0337}$	50,01	68,29	296,22	$237,\!37$	acetyl-CoA synthase
$SACE_{-}0619$	249,92	1745,04	18,37	$2251,\!27$	superoxide dismutase
$SACE_{-1134}$	$13,\!44$	49,705	1,38	7,99	hypothetical protein
$SACE_{2377}$	7,11	30,55	3,39	6,94	aldehyde dehydrogenase
$SACE_{2715}$	13,84	31,16	0,39	$0,\!67$	LytR family transcriptional regulator
SACE_3150	$6,\!97$	$23,\!52$	$5,\!43$	7,64	aromatic ring-hydroxylating dioxygenase
$SACE_4252$	$13,\!69$	29,05	52,04	5,01	proline permease
SACE_4486	$9,\!90$	32,53	3,04	9,72	bifunctional 3-phenylpropionate/cinnamic
SACE_4850	9,31	$36,\!65$	7,31	6,82	hypothetical protein

* Average TPM (transcripts perkilobase million) value of duplicates.

Table 4. Transcriptional change of nonsense genes in E3 with respect to NRRL23338

Locus tag	Function	TPM value* E3_10 h	TPM value* E3_50 h	TPM value* WT_10 h	TPN WT
$SACE_{-}0019$	o-succinylbenzoate-CoA synthetase/ligase	3,81	17,86	28,33	35,59
$SACE_{-}1076$	chitinase	$19,\!35$	$33,\!35$	20,59	78,22
$SACE_{-1257}$	4-diphosphocytidyl-2c-methyl-d-erythritol synthase	21,46	$27,\!35$	26,32	117,6
$SACE_{2875}$	modular polyketide synthase	0,78	6,80	14,17	78,49
SACE_3073	hydrogenase expression formation protein HypD	1,54	4,73	26,82	279,4
$SACE_{-7243}$	FAD linked oxidase domain protein	$5,\!00$	8,84	$35,\!12$	91,13

* Average TPM values of duplicates.

Figure 1. Physiological comparison between E3 and NRRL23338. Spores on XM agar plates over time (a), biomass formation (b), residual glucose concentration (c), and erythromycin titer (d). Error bars show the standard deviation from three independent experiments.

Figure 2. COG annotation of CDSs with nonsynonymous or intergenic SNP variations.

Figure 3. Functional enrichment analysis of DEGs in exponential phase based on KEGG (a) and GO (b) database. Y axis represents KEGG or GO terms, X axis shows the rich factor, which refers to the ratio of the number of DEGs belonging to the KEGG or GO term (sample number) compared to the number of all genes categorized in the KEGG or GO term (background number). The greater the rich factor is, the greater the enrichment degree is. The size of the node indicates the number of DEGs of the KEGG or GO term. The color of the node corresponds to different FDR (False Discovery Rate, corrected P-value) value.

Figure 4. Reporter metabolites in the exponential phase (a) and in the stationary phase (b) identified by Piano. Nodes represent metabolites around which the most significant changes in expression occurred. The number in the center of the node corresponds to the metabolites name in the table below (only top 10 metabolites were listed). The color of nodes indicates the P-value of significance. The size represents the number of genes around the metabolite according to the GSMM (Licona-Cassani et al., 2012; Zhuang, Huang, & Chu, 2018). The postfix, i.e. [c], of reporter metabolites means "intracellular". QH2, ubiquinol; Q, ubiquinone; XMP, xanthosine 5'-phosphate; IMP, inosine monophosphate.

Figure 5. Hierarchical clustering of genes in erythromycin biosynthesis gene cluster and genes related to polyketide sugar unit biosynthesis. Red, up-regulation in E3. Blue, down-regulation in E3.

Figure 6. Schematic map of putative mechamisms by which E3 enhances its erythromycin production. Pathways shown on red background were stimulated in E3, while a green background indicates the repression and grey indicates altered activity. KDPG, 2-keto-3-deoxy-6-phospho-gluconate.

Φιγυρε 7. Μεταβολις ενγινεερινη αρουνό τηε νοδε οφ 2-οξογλυταρατε (α- \times ετογλυταρατε) \approx E3. Identification of the key node, α-ketoglutarate (a). # indicates that SNP occurred in intergenic region of the gene. * indicates that nonsynonymous mutations occurred inside the CDS. The inset on the up-right showed the transcriptional change of genes around α-ketoglutarate node. Erythromycin titers of engineered strains and E3 in shake flasks (b). Strains were cultivated in 300 mL shake flasks with the minimal liquid medium. Data was collected from six independent replicates. Boundaries of the boxes indicate the 1st and the 3rd quartile of the sample populations and horizontal lines represent the median values. The squares inside the boxes represent the average values. Whiskers indicate the highest and lowest values of the results, whereas the solid rhombus indicate abnormal outliers.

Figure 8. Comparison of physiological process parameters between industrial strain E3 and mutant strain E3::sucBA. Dry cell weight (DCW) (a), consumed glucose (b), CO₂ emission rate (CER) (c), erythromycin titer (d).

Hosted file

Figures.docx available at https://authorea.com/users/441702/articles/542140-comparativegenomic-and-transcriptomic-analysis-guides-to-further-enhance-the-biosynthesis-oferythromycin-by-an-overproducer