Comparative genomics and transcriptional profiles of *Saccharopolyspora erythraea* NRRL 2338 and a classically improved erythromycin over-producing strain

Clelia Peano¹, Adelfia Talà², Giorgio Corti¹, Daniela Pasanisi², Miriana Durante³, Giovanni Mita³, Silvio Bicciato⁴, Gianluca De Bellis¹* and Pietro Alifano²*

**Abstract**

**Background:** The molecular mechanisms altered by the traditional mutation and screening approach during the improvement of antibiotic-producing microorganisms are still poorly understood although this information is essential to design rational strategies for industrial strain improvement. In this study, we applied comparative genomics to identify all genetic changes occurring during the development of an erythromycin overproducer obtained using the traditional mutate-and-screen method.

**Results:** Compared with the parental *Saccharopolyspora erythraea* NRRL 2338, the genome of the overproducing strain presents 117 deletion, 78 insertion and 12 transposition sites, with 71 insertion/deletion sites mapping within coding sequences (CDSs) and generating frame-shift mutations. Single nucleotide variations are present in 144 CDSs. Overall, the genomic variations affect 227 proteins of the overproducing strain and a considerable number of mutations alter genes of key enzymes in the central carbon and nitrogen metabolism and in the biosynthesis of secondary metabolites, resulting in the redirection of common precursors toward erythromycin biosynthesis. Interestingly, several mutations inactivate genes coding for proteins that play fundamental roles in basic transcription and translation machineries including the transcription anti-termination factor NusB and the transcription elongation factor Efp. These mutations, along with those affecting genes coding for pleiotropic or pathway-specific regulators, affect global expression profile as demonstrated by a comparative analysis of the parental and overproducer expression profiles. Genomic data, finally, suggest that the mutate-and-screen process might have been accelerated by mutations in DNA repair genes.

**Conclusions:** This study helps to clarify the mechanisms underlying antibiotic overproduction providing valuable information about new possible molecular targets for rationale strain improvement.

**Keywords:** *Saccharopolyspora erythraea*, Secondary metabolism, Antibiotic fermentation, Strain improvement, Comparative genomics

* Correspondence: gianluca.debellis@itb.cnr.it; pietro.alifano@unisalento.it

1 Institute of Biomedical Technologies, National Research Council, Via Fratelli Cervi 93, 20090 Segrate, Milano, Italy
2 Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, Via Monteroni, 73100 Lecce, Italy
Full list of author information is available at the end of the article

© 2012 Peano et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Background

Actinomycetes are ecologically important microorganisms that hold a prominent position in the industry due to their ability to produce a wide range of secondary metabolites with biological activities including antibiotics, anti-tumour agents and immuno-suppressants [1]. However, these microorganisms must often be genetically improved for higher production before they can be used in an industrial setting. Historically, strain improvement has been empirically carried out by multiple rounds of random mutagenesis and screening [2]. Since the late 1970s, the availability of molecular genetics tools and information about the biosynthetic pathways and genetic control for most of secondary metabolites of commercial interest has opened the way for improving strains through engineering-based approaches [3,4]. More recently, these rational strain improvement strategies benefit from the support of genomic, transcriptomic, proteomic, and metabolomic technologies [5-12]. Combining classical and recombinant strain improvement with a solid fermentation development program represents the optimal synergy to design commercially successful processes.

The erythromycin fermentation has been improved by the traditional mutate-and-screen method over the past 50 years. Erythromycin biosynthesis in the mycelial actinomycete, Saccharopolyspora erythraea, has been widely studied as a model system for antibiotic production [13-16] and erythromycin and its semi-synthetic derivatives are widely used in the clinic. As such, the development of improved producers still represents a challenging and up-to-date issue. Erythromycin A is obtained through a three-stage pathway [17], i.e., i) assembly of the 14-membered macrolactone 6-deoxyerythronolide B (6DEB) from one propionyl-CoA and six (25S)-methylmalonyl-CoA units by multifunctional modular polyketide synthase followed by ii) its hydroxylation to erythronolide B (EB), formation of the deoxysugars mycarose and desosamine from glucose and their addition to EB to make erythromycin D, and then iii) C-12 hydroxylation and C-3′ O-methylation of the latter compound to produce erythromycin A [18,19].

Extensive genetic studies have provided some insight into the genes involved in erythromycin biosynthesis [20,21]. The erythromycin gene cluster contains 20 genes arranged in four major polycistronic units [22]. Evidence for regulatory genes has been missing for a long time hampering efforts to enhance erythromycin production other than by medium manipulation, random mutagenesis and selection. In recent times, the availability of the entire genome sequence of S. erythraea and the advent of metabolic engineering opened the possibility to deeply investigate the molecular mechanisms controlling erythromycin production [23-25]. Whole-genome approaches led, for instance, to the identification of BldD, a key developmental regulator in actinomycetes [26,27], as one of the major regulators of erythromycin synthesis [28]. At the same time, metabolic engineering evidenced that, manipulating the methylmalonyl-CoA metabolite node in S. erythraea and in Aeromicrobium erythreum, a non-filamentous erythromycin A producer [29,30], i.e., increasing the flux through feeder metabolic pathways, strongly influences the erythromycin yields.

Lately, new global approaches based on “RNA polymerase and ribosome engineering” have been successful used to improve erythromycin production under laboratory conditions. It has been shown that several mutations affecting rpsL (coding for the ribosomal protein S12) result in a marked enhancement of erythromycin production, accompanied by increased transcription of bldD [31]. It has been reported also that several mutations in rpoB (coding for the beta subunit of RNA polymerase) deeply change the transcriptional profile of S. erythraea. In particular, the expression of genes coding for key enzymes of carbon (and energy) and nitrogen central metabolism is dramatically altered affecting in turn the flux of metabolites through erythromycin feeder pathways [32]. Mutations in ribosomal protein- and RNA polymerase subunit-encoding genes can be easily selected in the presence of drugs opening the way for a new approach to strain improvement. Very recently, complete biosynthesis of erythromycin A and designed analogs has been obtained using E. coli as a heterologous host suggesting alternative strategies to improve erythromycin production [33].

The focus of the present study is to investigate the molecular mechanisms leading to erythromycin overproduction in a classically improved strain by using a genomic approach. To this purpose, we identified all genetic changes that occurred in S. erythraea NRRL 2338/Px (hereafter indicated as S. erythraea Px), an erythromycin overproducer obtained through the traditional mutate-and-screen method. Compared with parental S. erythraea NRRL 2338, a total of 117 deletion sites, 78 insertion sites and 12 transposition sites were found across the genome of the overproducer. Moreover, single nucleotide variations affecting a total of 144 CDSs were identified between the two genomes. All genetic changes have been carefully mapped in S. erythraea Px genome and genomic information has been used to elucidate the molecular mechanism underlying the overproduction of erythromycin by this strain. Genomic comparison has been supported by comparative transcriptome, an approach that has been also successful used before this study [34], and phenotypic analysis.
**Results and discussion**

**Genomic comparison of *S. Erythraea* px and *S. Erythraea* NRRL 2338**

Phenotypic differences between *S. erythraea* Px and the reference strain NRRL 2338 are shown in Figure 1. With respect to NRRL 2338, Px exhibited slower growth and reduced sporulation in both Yeast Starch (YS) and Oatmeal Yeast (OMY) agar plates, less pigmentation in OMY agar, while the phenotypes of the two strains were more similar in Soluble Complete Medium (SCM) agar. SCM broth was used in fermentation and microarray experiments (see below).

To understand the genetic bases underlying phenotypic differences between the two strains and improved erythromycin A production in Px, whole genome of this strain was sequenced and compared with that of NRRL 2338. The genome of Px consists of a single circular chromosome of 8,212,111 bp with an average G + C content of 71.14%. The size of the Px chromosome is 694 bp smaller than that of NRRL 2338. Genomic comparison revealed highly conserved gene content and gene order between these two strains (Figure 2A). The two genomes are 99.1% identical and there is no remarkable change in the chromosome structure as clearly shown in the dot-plot comparison.

Compared with NRRL 2338, a total of 117 deletion sites, 78 insertion sites and 12 transposition sites were found across the Px genome. Among them, 71 sites are single nucleotide insertions/deletions (indels), which affect coding sequences leading to frame-shift mutations, and only 11 are indels larger than 100 bp. The largest insertion is 1127 bp near position 5,393,000 which is located between an oxidoreductase (SACE_4827) and a transcriptional regulator (SACE_4828), while the largest deletion is 1147 bp in 6,849,979-6,851,124 (corresponding to SACE_6108, an ATP-dependent helicase) which is located between a methyltransferase (SACE_6107) and a glycosyl transferase (SACE_6109)

Apart from frame-shift mutations, single nucleotide variations (SNVs) affecting a total of 144 CDSs were identified between the two chromosomes. The SNVs generate 110 missense and 10 nonsense mutations. The variations, also including 4 CDS duplications, 7 transposon/transposable element insertions and 1 transposon excision, affect a total of 227 proteins in the Px genome. Chromosome map of genetic variations distinguishing Px from NRRL 2338 is shown in Figure 2B. Mutations are homogeneously scattered along the whole chromosome without “core genome” vs. “non-core genome” preference. The mutated CDSs, the corresponding gene/locus names, the nature of variations

---

![Image of agar plates showing phenotypes of *S. erythraea* NRRL 2338 and *S. erythraea* Px.](image-url)
and presumed functions are reported in Tables 1, 2, 3, 4, 5, 6 with reference to the following aspects: i.) carbon, nitrogen and sulfur metabolism; ii.) substrate uptake, membrane permeability and secretion; iii.) biosynthesis of secondary metabolites; iv.) transcription and translation processes; v.) cell division, DNA replication and repair, transposition and phage integration; vi.) CDSs of unknown functions.

Transcriptome comparison of *S. Erythraea* px and *S. Erythraea* NRRL 2338

To gain further insight about the molecular mechanisms underlying improved erythromycin A production in Px strain, DNA microarray of *S. erythraea* were manufactured and used for comparative analysis between Px and NRRL 2338. Microarray data for NRRL 2338 strains were already available [24]. In DNA microarray experiments Px and NRRL were cultivated under standard batch-culture conditions in SCM broth [29] and erythromycin production was evaluated at different time points (Figure 3).

Due to the different length of lag phase, RNA samples were collected during growth of NRRL 2338 in the time interval 12-72 h, and in the time interval 24-84 for Px. Despite similar final values of biomass, shapes of growth

---

**Figure 2** Dot plot and chromosome map of genetic variations distinguishing *S. erythraea* Px from *S. erythraea* NRRL 2338

A) Dot plot comparison between the *S. erythraea* NRRL 2338 and Px strains generated by MUMmer software. B) Chromosome map of genetic variations distinguishing Px from NRRL 2338. The outer scale is numbered in megabases starting from the origin of replication (OriC), and indicates the core (blue) and noncore (yellow) regions. Outer circle: *S. erythraea* NRRL 2338 genes color-coded by their COG function (orange, cellular process and signaling; purple, information storage and processing; light blue, metabolism; pale green, poorly characterized); inner circle: positions of variations between Px and NRRL 2338 color-coded by type (green, frameshift; purple, nonsense; yellow, missense). Position of secondary metabolism cluster genes are shown.
| Gene locus in NRRL 2338 | Length (amino acids) | Gene/locus Variation<sup>a</sup> | Protein function and notes |
|-------------------------|----------------------|----------------------------------|-----------------------------|
| **Central carbon and energy metabolism** | | | |
| SACE_0618 | 286 | ccdA Missense (G38S) | Cytochrome c biogenesis protein |
| SACE_0633 | 221 | citA IPM | Citrate synthase (N-terminal) |
| SACE_1638 | 609 | sucB Missense (V574A) | 2-oxoglutarate dehydrogenase complex E2 component (dihydrolipoamide succinyltransferase) |
| SACE_3073 | 375 | hypD Nonsense (Y30*) | Hydrogenase expression/formation protein HypD |
| SACE_5291 | 335 | dhaK Frameshift (-C 717) | Dihydroxyacetone kinase N-terminal containing protein (DhaK) |
| SACE_6118 | 1125 | pyc Frameshift (-C 1222) | Pyruvate carboxylase |
| SACE_6385 | 1206 | kgd<sub>sucA</sub> Missense (1773 S) | 2-oxoglutarate dehydrogenase complex E1 component |
| **Nitrogen and amino acid metabolism** | | | |
| SACE_0635 | 122 | ureB Frameshift (-C 352) | Urease beta subunit |
| SACE_3800 | 689 | Missense (P601L) | Assimilatory nitrate reductase catalytic subunit (selenocysteine-containing) |
| SACE_4319 | 426 | aspB Missense (Y185C) | Aspartate aminotransferase |
| SACE_5299 | 433 | Missense (G103R) | D-amino acid deaminase |
| SACE_4527 | 417 | Frameshift | D-amino acid dehydrogenase, small |
| SACE_5741 | 500 | gltD Frameshift (-G 1301) | Glutamate synthase NADH/NADPH, small subunit |
| SACE_6330 | 421 | Frameshift (-G 324) | Fusion protein (ligase/carboxylase and argininosuccinate lyase) |
| SACE_6668 | 288 | sucD In frame insertion (-2081, -209V, -210M, -2111, -212G, -213E) | Succinyl-CoA synthetase, alpha subunit |
| **Sulfur and amino acid metabolism** | | | |
| SACE_3346 | 517 | Missense (C286R) | Sulfite oxidase/oxidoreductase, molybdopterin binding |
| SACE_4434 | 269 | tauC Missense (L223F) | Taurine transport system permease protein TauC |
| SACE_4651 | 279 | tauD Missense (G177S) | Taurine dioxygenase, 2-oxoglutarate-dependent |
| SACE_6133 | 158 | cdo2 Missense (P151S) | Cysteine dioxygenase |
| **Sugar metabolism** | | | |
| SACE_0652 | 383 | iroB Missense (A325V) | Glycosyl transferase, related to UDP-glucuronosyl transferase |
| SACE_1940 | 788 | lacZ2 Missense (A769V) | Beta-galactosidase/beta-glucuronidase |
| SACE_3069 | 207 | gmrA Frameshift (-G 429) | Phosphoheptose isomerase |
| SACE_3071 | 223 | gmrA Frameshift (-C 75, -C 90, -C 91, -T 95, -G 98, -G 99, -C 100, -G 126) | Phosphoheptose isomerase |
| SACE_3737 | 1001 | ama2 Missenses (R121A,N122Q, T123A, F124L, I125H, V127R) | Alpha-mannosidase |
| SACE_4596 | 814 | Missense (L810Q) | Beta-galactosidase/beta-glucuronidase |
| SACE_5147 | 1099 | embC Nonsense (S399*) | Arabinosyl transferase |
| SACE_5208 | 1098 | embC Missense (S799F) | Arabinosyl transferase |
| SACE_5734 | 657 | Frameshift (+A 1027/1028) | Galactofuranosyl transferase |
| SACE_6416 | 201 | mlC Missense (A173T) | dTDP-4-dehydrorhamnose 3,5-epimerase |
| SACE_6564 | 328 | deoC Missense (G16S) | Deoxyribosyl-phosphate aldolase |
| SACE_6765 | 501 | Missense (R207G, P208R) | Ribokinase |
| SACE_6971 | 433 | Missense (S23F) | Ribokinase |
| **Fatty acid metabolism** | | | |
| SACE_0351 | 281 | estC Missense (A211T) | Esterase |
and erythromycin production curves were markedly different between the two strains. Erythromycin production by both strains was detectable after 12 h. Consistently with previous findings [24], three distinct phases could be distinguished during erythromycin fermentation in the reference strain: an initial period of rapid increase of antibiotic concentration lasting until 60 h (phase a), followed by a period of production slowdown until 72 h (phase b), and a second period of moderate increase of erythromycin titers from 72 to 84 h (phase c) before entering the stationary phase. In the erythromycin overproducer phase a was protracted longer up to 72 h, phase b lasted until 84 h, while phase c was characterized by gradual increase of erythromycin titers until 108 h with a further period of production slowdown between 108 and 120 h followed by a further increase up to 132. Therefore, while during growth of NRRL 2338 erythromycin titers remained stable over the stationary phase, during growth of Px the antibiotic concentration continued to increase in the exhausted medium after entering the stationary phase reaching final titers about five-fold higher than in NRRL 2338.

Table 1 Variations within genes related to central carbon, nitrogen and sulfur metabolism (Continued)

| Strain | Gene 1 | Gene 2 | Description |
|--------|--------|--------|-------------|
| SACE_3132 | 275 | abmC | Missense (V246G) Frameshift (-C 742) Enoyl-CoA hydratase/isomerase (rev hydr) Putative: 2-ketocyclohexanecarboxyl-CoA hydrolase (amino benzoate degradation) |
| SACE_3216 | 526 | Frameshift (-G 1366) | Long-chain acyl-CoA synthetase |
| SACE_3361 | 329 | Misse (T312I) | Glycerophosphoryl diester phosphodiesterase |
| SACE_3450 | 304 | IPM | Enoyl-CoA hydratase/isomerase |
| SACE_3478 | 568 | IPM | Long-chain acyl-CoA synthetase |
| SACE_3745 | 255 | Misse (L187G) | Enoyl-CoA hydratase/isomerase |
| SACE_3936 | 246 | Misse (L48H) | 3-oxoacyl-[acyl-carrier protein] reductase |
| SACE_4589 | 388 | Frameshift (+CACC 10/11) | Acyl-CoA dehydrogenase |

Nucleotide metabolism

| Strain | Gene 1 | Gene 2 | Description |
|--------|--------|--------|-------------|
| SACE_1282 | 360 | ndB | Missense (S2N) Ribonucleotide-diphosphate reductase subunit β |
| SACE_2080 | 432 | pyrC | Misse (A90V) Dihydroorotase |
| SACE_6664 | 521 | purH | Misse (P545) Bifunctional phosphoribosylaminomimidazolcarboxyamide formyltransferase/IMP cyclohydrolase No parologue |
| SACE_7125 | 538 | purF | Misse (P497S) Glutamine phosphoribosylpyrophosphate amidotransferase No parologue |
| SACE_2398 | 165 | prsA | Misse (T123P) Phosphoribosylpyrophosphate synthetase |
| SACE_5196 | 506 | codA | Misse (Y322S) Cytosine deaminase |

Vitamin and cofactor metabolism

| Strain | Gene 1 | Gene 2 | Description |
|--------|--------|--------|-------------|
| SACE_0506 | 376 | thiO | Misse (A44V) Amino acid oxidase flavoprotein ThiO, putative Glycine/D-amino acid oxidases |
| SACE_0511 | 547 | thiC | Misse (T30) Thiamine biosynthesis protein ThiC |
| SACE_5955 | 1201 | cobN | Misse (A1091T) Cobaltochelatase subunit CobN |

Monooxygenases/oxidoreductases/methyltransferases/hydrolases of unknown function

| Strain | Gene 1 | Gene 2 | Description |
|--------|--------|--------|-------------|
| SACE_0651 | 422 | cypA | Frameshift (+C 16/17; +C 22/23) Cytochrome P450 |
| SACE_0781 | 191 | Misse (R80S) NADPH-dependent FMN reductase |
| SACE_4325 | 341 | Frameshift (-C 890) Radical SAM family protein Fe-S oxidoreductase |
| SACE_4560 | 326 | Frameshift (+A 58/59) SAM-dependent methyltransferase |
| SACE_4563 | 515 | Misse (S9F) Coproporphyrinogen III oxidase |
| SACE_4854 | 366 | Frameshift (+G 477/478) Amine oxidase, flavin-containing protein |
| SACE_5012 | 454 | Misse (G114E) Haem peroxidase |
| SACE_5030 | 372 | Frameshift (-G 553) Oye family NADH-dependent flavin oxidoreductase |
| SACE_5053 | 229 | Misse (L204F) Amidohydrolase 2 |
| SACE_6588 | 491 | Frameshift (-C 633) Monooxygenase, FAD-binding |
| SACE_7243 | 462 | Nonsense (W188*) FAD-dependent oxygenase/FAD/FMN-containing dehydrogenase |
| Gene locus in NRRL 2338 | Length (amino acids) | Gene/locus Variation* | Protein function and notes |
|-------------------------|----------------------|-----------------------|---------------------------|
| **Transport system**    |                      |                       |                           |
| SACE_0429               | 646                  | Missense (P416L)      | ABC-type transport system, permease component |
| SACE_0924               | 482                  | Frameshift (-C212; -C 221; -C 228; -G 264; -G 273) | Extracellular solute binding protein |
| SACE_0925               | 337                  | IPM                   | ABC-type transport system, permease component |
| SACE_0926               | 284                  | Missense (V18M, A89T) | ABC-type sugar transport system, permease component |
| SACE_0991               | 300                  | Frameshift (-G 370)  | Dipeptide transport system permease |
| SACE_1587               | 297                  | Frameshift (-C 585)  | Sugar transport system permease |
| SACE_1909               | 152                  | Missense (L90R)       | Twin-arginine translocation pathway signal |
| SACE_2131               | 278                  | Missense (L275F)      | Permease of the drug/metabolite transporter (DMT) superfamily |
| SACE_2701               | 276                  | Frameshift (-C 750)  | ABC-type transport system, ATP binding component |
| SACE_3038               | 545                  | Frameshift (+C 1407/1408) | ABC-type transport system, ATP binding component |
| SACE_3524               | 489                  | Frameshift (+G 976/977) | Permease of the major facilitator superfamily |
| SACE_4034               | 426                  | Missense (L221F)      | Permease of the major facilitator superfamily |
| SACE_4066               | 343                  | Frameshift (+C 268/269) | C4 dicarboxylate transporter/malic acid transport protein |
| SACE_4307               | 309                  | Missense (A90T)       | Integral membrane transport protein |
| SACE_4347               | 555                  | Missense (A277G)      | Sodium:solute symporter |
| SACE_4454               | 466                  | Missense (G348S)      | Permease of the major facilitator superfamily |
| SACE_4982               | 407                  | Missense (L145F)      | Integral membrane transport protein |
| SACE_5435               | 688                  | Missense and in frame deletion (N244T, G245A, A246P, D247D, R248A, L249A, H250R, G251R, G252L, L253T, Q255C, L256A, Q258H, A259R, T260G, L263-, P264-, T254S) | ABC-type transport system, permease component |
| SACE_5787               | 318                  | Missense (G300R)      | Dipeptide ABC transporter, ATP-binding protein |
| SACE_6087               | 200                  | kdpC                  | Potassium-transporting ATPase, C chain |
| SACE_6246               | 229                  | Missense (R206C)      | Phosphoenolpyruvate-protein phosphotransferase |
| SACE_6319               | 455                  | Frameshift (-C 1095) | Permease of the major facilitator superfamily |
| SACE_6323               | 315                  | Frameshift (-AT 913 and 914) | Dipeptide/oligopeptide/Nichel ABC transporter, permease component |
| SACE_6326               | 444                  | Frameshift (-G 1026) | Permease of the major facilitator superfamily |
| SACE_6578               | 364                  | Missense (G270R)      | ABC-type sugar transport system, permease component |
| SACE_6927               | 392                  | Missense (P121L)      | Permease of the drug/metabolite transporter (DMT) superfamily |
| SACE_6972               | 314                  | Missense (L224F)      | Permease of the major facilitator superfamily |
Moreover, while NRRL 2338 produced essentially erythromycin A during fermentation, Px yielded erythromycin A along with significant amounts of its immediate precursors erythromycin B and erythromycin C (Figure 3).

Gene expression data were analyzed to identify transcripts modulated during the growth curve. Considering each time point replicate as an independent entry and setting the confidence threshold at \(q\)-value \(\leq 0.001\) (see Methods), the EDGE algorithm identified a total of 404 differentially expressed genes (DEGs) in NRRL 2338 (6.22% of total probeset) with 220, 32 and 152 genes up-regulated, respectively, during phase a, b and c. In the Px strain the number of DEGs identified increased to 577 (8.88% of total probeset) and only two clusters of genes up-regulated, respectively during phases a (459 genes) and phase b (104 genes) could be clearly distinguished. Microarray analysis confirmed changes in global control of cell cycle in P. xanthus with respect to NRRL 2338 (Figure 4).

Inspection of functional classification of cell cycle-regulated genes showed that the increment in the number of DEGs in Px was mostly due to genes belonging to the following clusters of orthologous groups (COGs): I.5 Intracellular trafficking, secretion, and vesicular transport; I.6 Posttranslational modification, protein turnover, chaperones (in the functional category Cellular processes and signaling); II.12 Translation, ribosomal structure and biogenesis; III.8 Nucleotide transport and metabolism; III.9 Posttranslational modification, protein turnover, chaperones (in the functional category Metabolism) (Table 7). At the same time the following clusters of orthologous group resulted under-represented: II.11 Transcription and III.3 Cell wall/membrane/envelope biogenesis. In Table 7 the COGs have been considered over-represented (bold numbers in Table 7) if the percentage of the genes belonging to each category respect to the total of DEGs resulted 2 times more than the percentage of the genes belonging to each category respect to the total of GeneChip probe-sets. On the contrary they were considered under-represented (oblique numbers in Table 7) if the percentage of the genes belonging to each category respect to the total of DEGs resulted less than half of the percentage of the genes belonging to each category respect to the total probesets. A complete list of the cell cycle-regulated genes in NRRL 2338 and Px is available as supplemental data files (Additional files 1, 2), while the lists of genes up-regulated during phase a either in both strains or in Px or in NRRL 2338 are reported, respectively, in Additional files 3, 4, 5.

The Locally Adaptive Statistical Procedure (LAP) [35] of the PREDA package [ref] was used to identify differentially expressed chromosomal regions (Figure 5). LAP analysis showed significant up-regulation of many genes clustered in the “core” region of the chromosome (mostly containing essential genes) in both NRRL 2338 and Px strains during phase a, including the erythromycin biosynthetic cluster (ery) and the ribosomal proteins-encoding operons (Figure 5A and 5B). In contrast most of genes clustering in the “non-core” region of the chromosome (mostly containing “contingency” genes) were down-regulated during this growth phase. The only notable exception was represented by pks6 cluster.
coding for an unknown type II polyketide [23], which was up-regulated in Px strain during phase a. Moreover, while in NRRL 2338 the biosynthetic clusters nrps3, nrps5, tpc2, tpc3 and tpc4 clearly appeared to be down-regulated during phase a, in Px only two clusters, tpc4 and tpc5, exhibited this behavior suggesting profound differences in control of secondary metabolism between the two strains (Figure 5A and 5B). The LAP analysis failed to identify regions exhibiting significant differential gene expression between phase b and phase c in NRRL 2338 see Peano et al. 2007 [24].

Central carbon and nitrogen metabolism in S. Erythraea px

Table 1 shows that a considerable number of mutated genes are involved in central carbon (and energy) and nitrogen metabolism, or related to substrate uptake and utilization (Table 2). This is not surprising because erythromycin biosynthesis is strictly connected with central metabolism consistently with both genomic and expression data. In general, secondary metabolism is believed dispensable for survival, and most of gene clusters coding for secondary metabolites occupy non-core genomic regions and are maximally expressed during late growth phases. In contrast, the ery cluster maps in the core
| Gene locus in NRRL 2338 | Length (amino acids) | Gene (locus) | Variation\(a\) | Protein function and notes |
|------------------------|---------------------|-------------|----------------|--------------------------|
| SACE_2076              | 148                 | nusB        | Frameshift (–G 316) | NusB antiterminator factor No parologue. |
| **Transcription factors** |                     |             |                |                          |
| SACE_0891              | 206                 | TetR        | Frameshift (+G 581/582) | TetR family transcriptional regulator |
| SACE_1040              | 163                 | MarR        | Missense (R30C) | MarR family transcriptional regulator |
| SACE_1895              | 281                 | XRE         | Frameshift (+C 603/604) | XRE family transcriptional regulator |
| SACE_2927              | 918                 | SARP        | Missense (P52L) | SARP family transcriptional regulator |
| SACE_3079              | 294                 | LysR        | Frameshift (–C 506) | LysR family transcriptional regulator |
| SACE_3348              | 96                  | XRE         | Missense (E60A) | XRE family transcriptional regulator |
| SACE_4500              | 233                 | GntR        | Frameshift (–C 129; –G 222) | GntR family transcriptional regulator |
| SACE_4536              | 327                 | Lacl        | Missense (T246N, L249M) | Lacl family transcriptional regulator |
| SACE_4775              | 246                 | MerR        | Missense (A96V) | MerR family transcriptional regulator |
| SACE_5410              | 287                 | XRE         | Missense (P140F) | XRE family transcriptional regulator |
| SACE_5425              | 505                 | PucR-       | Missense and in frame deletion (V297S, V298W, A299R, V300C, Q301R, A302P, H303R, D304-, G305R) | PucR- family transcriptional regulator |
| SACE_5448              | 126                 | SACE_5449   | Duplication of SACE_5448 and SACE_5449 | SACE_5448: Unknown function SACE_5449: LuxR family transcriptional regulator |
| SACE_5739              | 282                 | XRE         | Missense (K206T) | XRE family transcriptional regulator |
| SACE_6021              | 484                 | PucR-       | Frameshift (+A 1212/1213) | PucR- family transcriptional regulator |
| **Signal transduction** |                     |             |                |                          |
| SACE_1833              | 241                 | Two-component system response regulator | Two-component system response regulator |
| SACE_1879              | 587                 | Serine/threonine protein kinase | Serine/threonine protein kinase |
| SACE_1988              | 150                 | Universal stress protein UspA | Universal stress protein UspA |
| SACE_2583              | 541                 | N-acyl D-amino acid deacylase | N-acyl D-amino acid deacylase |
| SACE_4937              | 414                 | Mandelate racemase/starvation sensing protein | Mandelate racemase/starvation sensing protein |
| SACE_5284              | 286                 | stpP        | IPM            | Endonuclease/Exonuclease/Phosphatase family protein |
| SACE_5286              | 224                 | Frameshift (+C 419, +G 430) | Two-component system response regulator |
| SACE_5301              | 533                 | Frameshift (+C 1585/1586) | Two-component system response regulator |
| SACE_6086              | 879                 | Osmosensitive K+ channel histidine kinase | Osmosensitive K+ channel histidine kinase |
| SACE_6447              | 228                 | mtrA        | In frame deletion (V96-, H197-) | Two-component system response regulator |
| SACE_6490              | 424                 | Missense (S37R) | Two-component system sensor kinase |
| SACE_6720              | 526                 | phoD        | Missense (C328Y, P382F) | Phosphodiesterase/alkaline phosphatase D (phosphate starvation) |
| SACE_7263              | 382                 | Missense (A222T) | Two-component system sensor kinase |
| **Translation machinery** |                     |             |                |                          |
| SACE_0443              | 464                 | cysS        | Missense (G198S) | Cysteinyl-tRNA synthetase |
| SACE_0709              | 593                 | metS        | Missense (P569S) | Methionyl-tRNA synthetase |
| SACE_2075              | 188                 | elfP        | Frameshift (+C 504) | Elongation factor P |
| SACE_3403              | 474                 | gatA        | Missense (A243E) | Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit |
| SACE_5926              | 1027                | infB        | Missense (A278V) | Translation initiation factor 2 |
| SACE_5919              | 536                 | Frameshift (–G 821) | Pseudouridine synthase |
| **Protein turnover and chaperones** |                 |             |                |                          |
| SACE_1339              | 860                 | IPM         | Aminopeptidase N | Aminopeptidase N |
region of the *S. erythraea* chromosome, and is transcribed during the middle pseudo-exponential growth phase when the activities of the carbon and nitrogen central metabolic pathways are maximal (Figures 4 and 5) [23,24,32,36].

As shown in Figure 6, these pathways are strictly connected to erythromycin biosynthesis that requires one propionyl-CoA and six (2S)-methylmalonyl-CoA units for assembly of the 14-membered macrolactone 6DEB. Different metabolic routes may accomplish

**Table 4 Variations within genes related to transcription and translation processes** (Continued)

| Gene | Length (amino acids) | Variation | Protein function and notes |
|------|----------------------|-----------|--------------------------|
| SACE_2951 | 174 | clpC Frameshift (+G 468/469) | ATPases with chaperone activity, ATP-binding subunit |
| SACE_3756 | 768 | clpA Missense (I284T) | ATPases with chaperone activity, ATP-binding subunit |
| SACE_6784 | 610 | Frameshift (-C 1573) | Molecular chaperone |
| SACE_6113 | 481 | IPM? | Aminopeptidase |

*IPM, imperfect match

**Table 5 Variations within genes related to cell division, DNA replication and repair, transposition and phage integration**

| Gene locus in NRRL 2338 | Length (amino acids) | Gene (locus) | Variation* | Protein function and notes |
|-------------------------|----------------------|--------------|------------|--------------------------|
| DNA replication and repair |                      | mfd | Frameshift (+A 218/219) | Transcription/repair coupling factor |
|                          |                      | mutM, fpg | Missense (M199T) | Formamidopyrimidine DNA glycosylase |
|                          |                      | ogtl | Missense (G41K) | Methylated DNA-protein cysteine methyltransferase |
|                          |                      | poiA | Frameshift (-C617) | DNA polymerase I |
|                          |                      | aikA | Missense (P68F) | 3-methyladenine DNA glycosylase |
|                          |                      | uvrD | IFM | ATP-dependent DNA helicase UvrD-like |
|                          |                      |              | Missense and in-frame deletion | ATP-dependent DNA helicase UvrD |
| Cell division |                      | SACE_0667 | 941 | Missense (F874L) | DNA segregation ATPase FtsK/SpoIIIE |
|                          |                      | SACE_6104 | 263 | IPM | Cell division initiation protein |
|                          |                      | SACE_6929 | 579 | Missense (P159R) | ATPase involved in chromosome partitioning |
| Transposition and phage integration |                      | SACE_0657 | 106 | Excision | IS1647-like transposase |
|                          |                      | SACE_2154 | 350 | IPM | Transposase, IS891/IS1136/IS1341 |
|                          |                      | SACE_2214 and similar CDSs | Insertion between SACE_4827 and SACE_4228 | SACE-2214: Transposase SACE_4827: Oxidoreductase SACE_4228: Lac family transcriptional regulator |
|                          |                      | SACE_2313 and similar CDSs | Insertion between SACE_2371 and SACE_2372 | SACE_2313: IS4-like transposase SACE_2371: IS4-like transposase SACE_2372: IS1647-like transposase |
|                          |                      | SACE_2322 | 374 | Missense (V391I) | Phage-related integrase/Site-specific recombinase XerC |
|                          |                      | SACE_3579 | 469 | trp | Missense (R135, A73S, T118A, V309A) | Transposase, IS1X5 |
|                          |                      | SACE_4072 | 469 | trp | Missense (S178L, Q191R, S199A,H218R, N370T, G430S, T431V) | Transposase, IS1xxS |
|                          |                      | SACE_5073 | 459 | IPM | Transposase, IS3508i |
|                          |                      |              | Missense and in-frame insertion | |
|                          |                      | SACE_5268 | 395 | (S73T, A74P, H129N,134T, -135V,L191V, I216V, A296V, P388R, S396L) | Transposase inactivated by frameshift mutation |
|                          |                      | SACE_5430 | 232 | Missense (H15R, R193W) | Transposase, Tn5714 |

*IPM, imperfect match
| Gene locus in NRRL 2338 | Length (amino acids) | Gene/locus Variation | Protein function and notes |
|--------------------------|----------------------|----------------------|--------------------------|
| **Unknown function**     |                      |                      |                          |
| SACE_0062 35             | Insertion between SACE_2351 and SACE_2352 | **SACE_0062** Unknown function. **SACE_2351**: Tn5714-like transposase. **SACE_2352**: IS 111 a/IS 1328A/IS1533-like transposase. |
| SACE_0062 35             | Insertion in SACE_5430, between SACE_5429 and SACE_5431 | **SACE_0062** Unknown function. **SACE_5429**: Dephospho-CoA kinase. **SACE_5430**: Transposase. **SACE_5431**: 30S ribosomal protein S |
| SACE_0157 389            | Missense (P295L)     | Unknown function/Glutathionylspermidine synthase |
| SACE_0587 995            | Frameshift (+G 332/333) | Unknown function |
| SACE_0744 164            | Missense (R42H)      | Unknown function |
| SACE_0940 289            | IPM                  | Unknown function |
| SACE_0944 831            | Nonsense (Y412*)     | Unknown function |
| SACE_1128 294            | Missense (V113 A)    | Unknown function |
| SACE_1129 774            | Missense (A172T)     | Unknown function |
| SACE_1257 195            | Nonsense (Q60*)      | Unknown function/Uncharacterized MobA-related protein |
| SACE_1344 1638           | IPM                  | Unknown function/PE-PGRS family protein |
| SACE_1805 100            | Missense (S13F)      | Unknown function |
| SACE_1835 175            | IPM                  | Unknown function |
| SACE_1853 110            | Frameshift (+G 51/52) | Unknown function |
| SACE_2384 496            | Frameshift (+T 823/824) | Unknown function |
| SACE_2456 251            | Missense (A28V)      | Unknown function |
| SACE_2737 116            | Missense (P4R)       | Unknown function |
| SACE_3005 316            | Frameshift (-G 727)  | Unknown function |
| SACE_3102 1249           | Frameshift (+G 338/339) | Unknown function |
| SACE_3186 392            | Missense (R390S)     | Unknown function/tpcl cluster |
| SACE_3262 282            | Frameshift (+C 434/435) Insertion between | Unknown function. **SACE_3264**: Unknown secreted protein. |
| SACE_3264 117            | SACE_3572 and SACE_3573 | **SACE_3572**: Unknown secreted protein. **SACE_3573**: Unknown function |
| SACE_3748 277            | Frameshift (+C 696/697) | Unknown function |
| SACE_3850 261            | Frameshift (-G 388)  | Unknown function |
| SACE_3925 188            | Frameshift (+A 437/438) | Unknown function |
| SACE_4249 234            | Frameshift (-G 640)  | Unknown function |
| SACE_4310 480            | Frameshift (-C 1146) | Unknown function |
| SACE_4431 291            | In frame insertion (-102D,-103A,-104N,-105E,-106Q) | Unknown function |
| SACE_4989 38             | Insertion between SACE_5103 and SACE_5105 | **SACE_4989**: Unknown function. **SACE_5103**: Type III restriction enzyme. **SACE_5105**: Transposase |
| SACE_5311 835            | Missense (G411R) Frameshift | Unknown function |
| SACE_5423 184            | (-G 372,-G 473,-C 481,-G 493,-T 409,-C 540) | Unknown function |
| SACE_5446 1184           | Frameshift (-C 2898) | Unknown function |
| SACE_5460 330            | Missense (P162F)     | Unknown function |
| SACE_5482 632            | Missense (L509P)     | Unknown function |
| SACE_5483 11792          | Missense (L1329F, T5002A, P8537L, R9518H) | Unknown function |

---

**Table 6 Variations within genes of unknown function**

Peano et al. *Microbial Cell Factories* 2012, 11:32

http://www.microbialcellfactories.com/content/11/1/32

Page 12 of 25
Table 6 Variations within genes of unknown function (Continued)

| Gene     | Variation                                                                 | Function                                      |
|----------|---------------------------------------------------------------------------|-----------------------------------------------|
| SACE_5513| In frame deletion (S249-, E250-, G251-, T252-, T253-, G254-, G255-, G256-, G257-, G258-, A259-, G260-) | Unknown function                              |
| SACE_5523| Nonsense (L370*)                                                           | Unknown function                              |
| SACE_5655| Duplication of SACE_5655                                                  | Unknown function                              |
| SACE_5905| Missense and in frame insertion (A5V, T7G, I8C, R10H, V11H, Q12L, T13A, M14G, S15A, I16D, E17H, S18E, A19H, R21E, T22R, L24A, -25H, -26R, -27A) | Unknown function                              |
| SACE_6184| Frameshift (-G 169)                                                       | Unknown function                              |
| SACE_6567| Missense, in frame insertion, nonsense (A135G, -136R, -137R, 138G, -139A, -140A, -141R, -142W, -143G, -144G, -145R, -146R, -147S, -148R, A149R, L155R, G156A, T157R, A158R, A159R, V160R, V161R, T161R, F163R, A165R, A166R, I167R, A168G, T170R, V171R, V172R, K173R, D174R, W175R, F176R, V177R, A178R, A179R, A180R) | Unknown function                              |
| SACE_6773| Missense (D330E) Frameshift (-G 1654)                                      | Unknown function/PE-PGRS family protein       |
| SACE_7193| Frameshift (+C 729/730)                                                    | Unknown function                              |
| SACE_7240| Missense (P90H)                                                           | Unknown function                              |
| SACE_7316| Duplication of SACE_7316                                                  | Unknown function                              |

*IPM, imperfect match

Figure 3 Bioreactor cultures of S. erythraea NRRL2338 and S. erythraea Px: Biomass, erythromycin production, pO2 and pH were evaluated as described in the Materials and Methods section. Erythromycin production was monitored by HPLC. Total erythromycin is the sum of erythromycin A, B and C. Values are means ± S.D. of three independent experiments for each strain starting from the same frozen culture.
precursor supply for erythromycin biosynthesis. In addition to well-established propanoate metabolism leading to both propionyl-CoA and six (2S)-methylmalonyl-CoA units, alternative routes exist including valine, leucine and isoleucine degradation pathways, and glycine, serine and threonine metabolism. An additional proposed route to (2S)-methylmalonyl-CoA proceeds by the rearrangement of succinyl-CoA catalyzed by methylmalonyl-CoA mutase yielding the (2R)-isomer of methylmalonyl-CoA that would be converted to the (2S)- isomer by a methylmalonyl-CoA epimerase.

Our analysis demonstrated that the several genes coding for the component enzymes of the tricarboxylic acid cycle (TCA) were mutated in Px strain. In particular, pyc gene (SACE_6118) coding for pyruvate carboxylase (PC) was inactivated by a frameshift mutation. This result is noteworthy, because PC serves a major anaerobic role for the TCA by catalyzing the ATP-dependent carboxylation of pyruvate to oxalacetate, and there is evidence that in Corynebacterium glutamicum PC contributes about 90% to C(3) carboxylation at the anaerobic node (Petersen, 2000). Moreover, pyc has no obvious paralog(s) in the genome of S. erythraea. In
### Table 7 Functional classification of cell cycle-regulated genes in both strains

| COG    | Functional categories                                                                 | Strain NRRL2338 | Px |
|--------|----------------------------------------------------------------------------------------|------------------|----|
|        | % COG probesets                                                                         | % Probesets with q-value ≤ 0.001 | % Probesets with q-value ≤ 0.001 |
| I      | Cellular processes and signaling                                                        |                  |    |
| I.1    | Cell cycle control, cell division, chromosome partitioning                             | 0.35             | 0.99 | 0.69 |
| I.2    | Cell motility                                                                          | 0.01             | 0    | 0    |
| I.3    | Cell wall/membrane/envelope biogenesis                                                 | 2.05             | 3.22 | 3.81 |
| I.4    | Defense mechanisms                                                                     | 0.89             | 0.00 | 0.87 |
| I.5    | Intracellular trafficking, secretion, and vesicular transport                          | 0.24             | 0.25 | 0.69 |
| I.6    | Posttranslational modification, protein turnover, chaperones                           | 1.90             | 5.94 | 3.99 |
| I.7    | Signal transduction mechanisms                                                         | 1.91             | 0.99 | 1.91 |
| II     | Information storage and processing                                                     |                  |    |
| II.1   | Amino acid transport and metabolism                                                     | 0.01             | 0    | 0    |
| II.2   | Cell wall/membrane/envelope biogenesis                                                 | 0.01             | 0    | 0    |
| II.3   | Chromatin structure and dynamics                                                        | 0.01             | 0    | 0    |
| II.4   | General function prediction only                                                        | 0.30             | 0.50 | 0.35 |
| II.5   | Nucleotide transport and metabolism                                                    | 0.07             | 0    | 0    |
| II.6   | Posttranslational modification, protein turnover, chaperones                           | 0.04             | 0    | 0    |
| II.7   | Replication, recombination and repair                                                   | 2.44             | 0.74 | 2.43 |
| II.8   | RNA processing and modification                                                        | 0.01             | 0    | 0    |
| II.9   | Secondary metabolites biosynthesis, transport catabolism                               | 0.04             | 0    | 0    |
| II.10  | Signal transduction mechanisms                                                         | 0.71             | 0.50 | 0.69 |
| II.11  | Transcription                                                                         | 7.20             | 2.23 | 3.29 |
| II.12  | Translation, ribosomal structure and biogenesis                                        | 2.48             | 4.21 | 10.75 |
| III    | Metabolism                                                                             |                  |    |
| III.1  | Amino acid transport and metabolism                                                     | 6.60             | 0    | 8.15 |
| III.2  | Carbohydrate transport and metabolism                                                  | 6.08             | 7.67 | 6.59 |
| III.3  | Cell wall/membrane/envelope biogenesis                                                 | 0.38             | 0    | 0.17 |
| III.4  | Coenzyme transport and metabolism                                                      | 2.63             | 5.20 | 4.33 |
| III.5  | Energy production and conversion                                                       | 4.79             | 9.90 | 8.15 |
| III.6  | Inorganic ion transport and metabolism                                                 | 2.68             | 4.21 | 2.25 |
| III.7  | Lipid transport and metabolism                                                        | 4.24             | 2.72 | 5.03 |
| III.8  | Nucleotide transport and metabolism                                                    | 1.27             | 2.97 | 3.64 |
| III.9  | Posttranslational modification, protein turnover, chaperones                           | 0.08             | 0    | 0.17 |
| III.10 | Secondary metabolites biosynthesis, transport catabolism                               | 2.90             | 2.48 | 2.77 |
| III.11 | Signal transduction mechanisms                                                         | 0.16             | 0    | 0    |
| IV     | Poorly characterized                                                                   |                  |    |
| IV.1   | Function unknown                                                                       | 32.11            | 27.23| 22.70|
| IV.2   | General function prediction only                                                       | 7.37             | 8.17 | 6.59 |
|        | Total probesets                                                                        | 7060             | 404  | 577  |

*The categories over-represented and under-represented are evidenced, respectively, by BOLD and oblique number*
addition to pyc, non-conservative missense mutations affected both the E1 and the E2 component-encoding genes (sucA [SACE_6385] and sucB [SACE_1638], respectively) of the 2-oxoglutarate dehydrogenase complex. 2-oxoglutarate dehydrogenase is a key enzyme in the TCA cycle, converting 2-oxoglutarate, coenzyme A and NAD(+) to succinyl-CoA, NADH and carbon dioxide. This activity is tightly regulated and it is a

![Figure 5](image1.png)

**Figure 5** Visualization of the LAP algorithm results on the two chromosome strands independently analyzed. Comparison between the transcriptional profiling of all the 6494 *S. erythraea* genes in phase A versus phase B in the NRRL2338 strain (A) and in the Px strain (B) is shown. A q value of 0.01 and a fold change of 0.5 were chosen as filtering parameters. The transcriptionally up-modulated regions are shown in RED while the down-modulated are in GREEN. The OriC is indicated as O and the resolution of the chromosome is 1 Mb. The core region is evidenced in blue and the non-core region in orange; the position of clusters involved in the secondary metabolism, erythromycin production and coding for ribosomal proteins are outlined by arrows.

![Figure 6](image2.png)

**Figure 6** Overview of carbon and nitrogen metabolic pathways and their relationships with the biosynthesis of erythromycin. Genes affected by missense mutations are indicated in blue; genes affected by missense/nonsense mutations are shown in red. Single-copy genes are underlined.
major determinant of the metabolic flux through the TCA cycle. Also in this case, obvious paralogs could not be found in the genome of this microorganism. Although it is difficult to predict the effect of the observed mutations on the activity of the 2-oxoglutarate dehydrogenase complex, altogether our results are consistent with the hypothesis that the carbon flux through the TCA cycle may be greatly reduced in the Px strain, and that, as a consequence, more acetyl-CoA is funneled into alternative routes including propanoate metabolism leading to the 6DEB precursors (Figure 6).

This view is supported by evidence that glutamine synthetase/glutamate synthase (GS-GOGAT) and urea cycles, which are linked to TCA by 2-oxoglutarate and oxaloacetate/fumarate, respectively, may also be affected in the Px strain. Indeed, in this strain, *gltD*, coding for small subunit of glutamate synthase (SACE_5741), a component of the high-affinity ammonium assimilation system, is inactivated by a frameshift mutation (Table 1 and Figure 6). However, in this case, a paralogous gene (SACE_3997) is predicted to exist in the genome, to alleviate the deleterious effect of such a mutation. A possible deficiency in the high-affinity ammonium assimilation system may also account for the different behavior of *S. erythraea* NRRL 2338 and Px strain with respect to ammonium content in the growth medium. Growth, pigment and antibiotic production was strongly inhibited in the wild type strain in the presence of high ammonium content, while, in contrast, the same parameters were not affected by ammonium in the erythromycin over-producing strain (Figure 7). Frameshift mutations inactivate SACE_6330 (*argH* domain) and SACE_0635 (*ureB*). These CDSs may be involved in the urea cycle as they code, respectively, for a fusion protein with both ligase/carboxylase and argininosuccinate lyase domains, and urease beta subunit. In the latter case, no obvious *ureB* paralog could be detected in the genome.

In addition to the urea cycle, one of the two aspartate aminotransferase-encoding paralog (*aspB*) (SACE_4319), whose activity is linked to the urea cycle and is essential for growth on L-glutamate, was also affected by a non-conservative missense mutation.

In the absence of sulfate and cysteine, several bacteria can use aliphatic sulfonates as a source of sulfur for growth. *tauABCD* genes coding for an ABC-type transport system required for uptake of aliphatic sulfonates and a desulfonation enzyme accomplish this property. In Px strain *tauC* (SACE_4434) coding for permease component of taurine transport system, and *tauD* (SACE_4651) coding for 2-oxoglutarate-dependent taurine dioxygenase are affected by missense mutations, along with *cdo2* (SACE_6133) encoding a cysteine dioxygenase, which catalyzes the conversion of L-cysteine to cysteine sulfinic acid, a compound that lies at a branch-
point in cysteine catabolism, where it can follow two pathway resulting in the formation of taurine or sulfate (Table 1). Down-modulation or inactivation of this metabolic pathway involving the 2-oxoglutarate-dependent taurine dioxygenase may be functional to alleviate the reduced carbon flux through the Krebs cycle by increasing the pool of 2-oxoglutarate.

Several genes involved in nucleotide and thiamine metabolism were also affected by non-conservative missense mutations in the Px strain, including \textit{prsA} (SACE\textsubscript{2398}) coding for phosphoribosylpyrophosphate (PRPP) synthase (which catalyzes the synthesis of the common precursor for biosynthesis of histidine and purine and pyrimidine nucleotides), \textit{purF} (SACE\textsubscript{7125}) and \textit{purH} (SACE\textsubscript{6664}) whose products code for glutamine and pyrimidine nucleotides, \textit{thiO} (SACE\textsubscript{0506}) and \textit{thiC} (SACE\textsubscript{0511}) specifically involved in thiamine biosynthesis, and \textit{pyrC} (SACE\textsubscript{2080}) encoding dihydroorotase, a key enzyme in pyrimidine nucleotide biosynthesis (Table 1 and Figure 6). Mutations in the above-mentioned genes are expected to reduce the carbon flux toward these biosynthetic pathways (purine/thiamine, pyrimidine), and increase the flux toward the pentose phosphate (\textit{prsA} mutation) and glycine, serine and threonine metabolic pathway (\textit{purF, purH, thiO, thiC} mutations). The last pathway is a source of precursors for 6DEB biosynthesis.

The above-mentioned genetic defects in metabolic genes are well correlated with the slow-growth phenotype of the Px strain in mineral medium MM-101, caused by nutritional requirements which are indicative of TCA cycle precursors and PRPP limitations (Additional file 6). At the same time, our results emphasize the balance that must be reached between pathways competing for the same substrate to maintain robustness of the metabolic network.

**Secondary metabolism in \textit{S. Erythraea} px**

In addition to the gene clusters for erythromycin (\textit{ery}), for a second modular polyketide synthase (PKS) of unknown function (\textit{pke}) and for a type III PKS (\textit{rppA}), which generates a reddish pigment, the genome sequencing of \textit{S. erythraea} NRRL 2338 has revealed additional 22 clusters for the biosynthesis of polyketides, terpenes and non-ribosomally synthesized peptides. In Px strain a total of 19 missense/nonsense/frameshift mutations affected genes related to biosynthesis of secondary metabolites (Table 3).

Twelve of them affected genes coding for PKS or related biosynthetic proteins. In particular, a nonsense mutation in SACE\textsubscript{0019} (coding for an ACP), a frameshift mutation in SACE\textsubscript{0022} (\textit{pfaB}) (coding for a modular PKS) and two non-conservative missense mutations in SACE\textsubscript{0023} (\textit{pfaC}) (coding for a modular PKS) inactivated the \textit{pfa} cluster, which appears to govern the biosynthesis of a polyunsaturated fatty acid such as eicosapentaenoic acid. The \textit{pks3} cluster was inactivated by a frameshift mutation in SACE\textsubscript{2875} (coding for a modular PKS) and a non-conservative missense mutations in SACE\textsubscript{2876} (\textit{gdmF}) (coding for 3-amino-5-hydroxybenzoic acid synthase). Missense mutations affected the \textit{pks3}-associated SACE\textsubscript{2888} (coding for an aromatic-L-amino acid decarboxylase). Non-conservative missense mutations were mapped in SACE\textsubscript{2630} (\textit{pks2-1}) (coding for a modular PKS) of the biosynthetic cluster \textit{pks2, SACE\textsubscript{4140} (pkeA2)} (coding for a modular PKS) in the \textit{pke} cluster, and SACE\textsubscript{5308} coding for a multifunctional single-module PKS enzymes apparently related to the iterative PKSs involved in enediyne or methylsalicylic acid synthesis. It is conceivable that some of these pathways may compete with that of erythromycin for the same substrates and that their inactivation/down-modulation may be beneficial to 6DEB biosynthesis.

While it is difficult to predict the effects of the missense mutations affecting \textit{eryCVI} and \textit{eryBVI} (Table 3) coding for the enzymes catalyzing the last steps of the biosynthesis of dTDP-D-desosamine and dTDP-L-mycarose biosynthesis, the activated sugars that decorate the erythronolide B, the missense mutation affecting SACE\textsubscript{6416} (\textit{rmlC}) coding for dTDP-4-dehydroharmnose 3,5-epimerase (Table 1 and Figure 6) may cause an increase in production of both dTDP-D-desosamine and dTDP-L-mycarose precursors. In fact, the \textit{rmlC, eryCVI} and \textit{eryBVI} gene products compete for the same substrate (dTDP-4-oxo-6-deoxy-D-glucose). Additional work is required to clarify these aspects, as well as the possible effects of the inactivation/down-modulation of \textit{nrps4} and \textit{nrps7} biosynthetic clusters, coding for unknown non-ribosomally synthesized peptides, and of \textit{tpc1/geo1} and \textit{tpc3/geo2} showing substantial similarity to terpene cyclase-encoding clusters, which in other microorganisms are known to produce geosmin, the sesquiterpene that provide the soil with its characteristic smell.

**Global and pathway-specific control of gene expression, and DNA repair mechanisms in \textit{S. Erythraea} px**

Antibiotic production is under global and local pathway-specific control. At the same time, there is evidence that genetic manipulation of RNA polymerase and/or ribosome may influence the control of secondary metabolism [31,32]. Mutations affecting the basic transcription and translation machineries were thus expected in the
erythromycin-overproducing strain (Table 4). However, unexpectedly, several inactivating mutations mapped in a number of genes, whose products were thought to play fundamental roles in these processes including nusB (SACE_2076) coding for NusB antiterminator factor and efp (SACE_2075) encoding the elongation factor P (EF-P). Both these genes, which are organized in a putative conserved operon also including pepQ gene coding for Xaa-Pro dipeptidase, were inactivated by frameshift mutations (Figure 8A).

NusB participates together with NusE/S10 protein in processive transcription antitermination. NusB and NusE, bind to form a heterodimer, which interacts with a specific boxA site on the RNA. The NusB/NusE/boxA RNA ternary complex interacts with the RNA polymerase transcription complex, stabilizing it and allowing transcription past premature termination points [37–40]. NusB is essential to suppress transcription termination in the ribosomal RNA (rrn) operons. Compared to wild type Escherichia coli, about two-fold decreased RNA polymerase density was observed by electron microscopy over 23S genes in a nusB mutant [41]. As a consequence, the fraction of total RNA polymerase engaged in transcribing the rrn operons is significantly reduced in a nusB mutant. Similar reduction is observed during the stringent response, when (p)ppGpp binds RNA polymerase and changes the global transcriptional profile, decreasing the synthesis of translational machinery and increasing the transcription of biosynthetic genes including those coding for antibiotics in actinomycetes [32,42–48].

EF-P is a highly conserved protein that is essential for protein synthesis in several bacteria including E. coli [49]. It has been suggested that EF-P plays a role in translational fidelity, prevents entry of fMet-tRNA into the A-site enabling it to bind to the 50S P-site, and promotes a ribosome-dependent accommodation of fMet-tRNA into the 70S P-site [50,51]. In Bacillus subtilis genetic inactivation of EF-P abolished spore formation without affecting growth [52]. Although it is difficult to predict the effects of EF-P inactivation on erythromycin production, it is relevant to note that erythromycin production in S. erythraea is stimulated by streptomycin-resistance mutations [31] and that EF-P was shown to protect 16S rRNA near the G526 streptomycin and the S12 and mRNA binding sites (30S T-site) [50].

In addition to EF-P, other genetic changes occurs in the Px strains affecting the translational machinery, including missense mutations affecting SACE_5926 (infB) coding for the translation initiation factor 2, (SACE_0799) (metS) coding form methionyl-tRNA synthetase, SACE_0443 (cysS paralog) encoding a cysteinyl-tRNA synthetase, SACE_3403 (gatA) coding for A subunit of Asp-tRNA Asn/Glu-tRNA Gln amidotransferase, and a frameshift mutation inactivating a paralogous gene (SACE_5919) coding for pseudouridine synthase. These findings emphasize the importance of the translational machinery as a potential target for improvement of antibiotic-producing strains.

The mutations affecting the basic transcription and translation machineries may in turn affect pathway-specific control accounting for the global changes in transcriptional profile, which was observed in Px strain with respect to NRRL 2338 (Figures 4 and 5). As previously mentioned, a total of 459 and 220 genes were up-regulated during growth phase a in Px and in NRRL 2338 respectively. Among these DEGs, 138 were up-regulated in both strains, while 335 and 114 were specific for Px
and NRRL 2338, respectively (Additional file 3, 4, 5). Within the list of genes specifically up-regulated in the Px strain, we found most of the genes belonging to the ery biosynthetic gene cluster. Regulation of the ery cluster was found to be altered also in another classically improved S. erythraea strain, which exhibited prolonged expression of antibiotic biosynthetic genes compared to the wild type during fermentation [34].

In addition to the ery cluster, SACE_1456 (mmsA2) coding for methylmalonate semialdehyde dehydrogenase was also found to be up-regulated in Px. This enzyme provides 6DEB biosynthesis with propionyl-CoA and (2S)-methylmalonyl-CoA precursors (Figure 6) and was found to be up-regulated in a rifampicin-resistant erythromycin-overproducing mutant of S. erythraea [32]. Moreover, the expression of bldD (SACE_2077), coding for a key developmental regulator that seems to regulate the ery cluster positively, was also up-regulated during phase a in the Px strain (Figure 8B). Intriguingly, this gene maps immediately downstream of the pepQ-epg-nusB locus in the genome of S. erythraea (Figure 8A). In addition to the ery cluster and associated CDSs, the genes belonging to the pks6 and tpc5 (geo3) biosynthetic clusters were also found to be up-regulated in this strain along with nusA (SACE_5927 coding for transcriptional termination/antitermination factor NusA), ribosomal protein operons with the associated genes infA (SACE_6808 coding for translation initiation factor IF1), tufA (SACE_6838 coding for elongation factor EF1A), fusA (SACE_6839 coding for translation elongation factor G), and F0F1 ATPase. In contrast, in the list of genes specifically up-regulated in the NRRL 2338, we found only genes related to the TCA cycle including fumB (SACE_1784 coding for class I fumarate hydratase), fumC (SACE_0930 coding for class I fumarate hydratase), korA (SACE_3927 coding for 2-oxoglutarate ferredoxin oxidoreductase alpha subunit), the urea cycle (ureA) (SACE_0634 coding for urease gamma subunit) or the PRPP biosynthesis (prsA) (SACE_0816). In addition to changes in basic transcription and translation machineries, mutations in genes coding for transcriptional factors and signal transduction proteins may contribute to alter pathway-specific control. In particular, 13 genes coding for transcriptional factors and 5 genes encoding two-component system proteins were affected by missense or frameshift mutations in Px (Table 4). However, in the absence of functional data, the effects of such mutations on secondary metabolism may be only hypothesized on the basis of proximity of the mutated regulatory genes to the biosynthetic cluster for secondary metabolites. For instance, SACE_3079, coding for a LysR family transcriptional regulator and affected by a frame shift mutation in the same strain, is genetically linked to a hydrogenase operon (also containing gmhA genes, which were shown to be inactivated by frameshift mutations in Px) and may be involved in its regulation. Similarly, SACE_4500, encoding a GntR family transcriptional regulator that is inactivated by a frameshift mutation in Px, maps close to sarcosine oxidase operon. Sarcosine oxidase links glycero-phospholipid metabolism to erythromycin biosynthesis catalyzing the oxidative demethylation of sarcosine to glycine, whose metabolism provides the biosynthesis of 6DEB with precursors (Figure 6).

Px strain was obtained by multiple cycle of mutagenesis by chemical (N-methyl-N-nitro-N-nitrosoguanidine) or radiation (X-ray and UV) agents. The sequence data suggests that the mutagenesis/selection process to erythromycin-overproduction phenotype might have been accelerated by selection of mutator phenotype. This hypothesis is supported by missense and frameshift mutations affecting genes involved in DNA replication and repair. In particular, SACE_0826 (mfd) coding for transcription/repair coupling factor, SACE_4427 (mug) coding for G/U mismatch-specific DNA glycosidase and SACE_5437 (polA) were inactivated by frameshift mutations, while missense mutations affected other genes involved in these processes (Table 5).

Conclusions

Overall our findings demonstrate that the phenotypes of the erythromycin-overproducing strain S. erythraea Px are associated with a large number of genetic changes with respect to the reference strain NRRL 2338. Mutations affect 227 CDSs, corresponding to about 3% of CDSs of S. erythraea genome. Although certain mutations may be neutral in term of improved antibiotic production, and may have been favored by the mutator background of Px, a considerable number of them map within genes coding for key enzymes involved in central carbon and nitrogen metabolism, and biosynthesis of secondary metabolites, redirecting common precursors toward erythromycin biosynthesis. Several mutations inactivate genes coding for proteins that play fundamental roles in basic transcription and translation machineries including the transcription anti-termination factor NusB and the transcription elongation factor Efp, and genes coding for pleiotropic or pathway-specific regulators, with dramatic effects on global expression profile. The comparison of Px and NRRL 2338 at both genomic and transcriptomic levels not only contributed to elucidate the molecular mechanism underlying the overproduction of erythromycin, but also revealed new possible targets suitable for rationale improvement of industrial antibiotic-producing strains. However, as most of mutated genes are not directly related to the erythromycin over-production, next effort will be to test the genome-assisted predictions by experimental verification.
Methods

Bacterial strains and media

*S. erythraea* wild type strain NRRL2338 was a gift of S. Donadio (KtedoGen, Milan, Italy). This strain has been deposited at the American Type Culture Collection. *S. erythraea* Px is an erythromycin-overproducing strain that was obtained by the traditional mutate-and-screen method over a period of about 10 years. The strains were stored in 1-ml cryotubes at -80°C as frozen mycelium in YS medium containing 15% glycerol at a biomass concentration of approximately 0.25 g dry cell weight (DCW) ml⁻¹, or at -20°C as spores in 20% glycerol (in distilled water) at a title of approximately 5 × 10⁸ ml⁻¹. The composition (per liter) of the complete media used in this study is reported in Table 8. When requested all media were agarized at a concentration of 1.8%. The composition (per liter) of the nutrient broth agar in the microbiological assays with *Micrococcus luteus* tester strain was: 3 g beef extract, 5 g tryptone, 15 g NaCl, 15 g agar.

Preparation of spores

Concentrated spore suspensions (5 × 10⁸ ml⁻¹) are crucial for purposes like starting reproducible cultures for physiological or fermentation studies. To prepare spores adapted to the conditions of liquid medium, spores were spread on the same medium with agar. Mycelium with spores was strongly attached to the surface agar, thus making impossible to collect spores without agar traces. Therefore, strains have been grown on cellophane discs, as described in [53]. The cellophane discs were sterilized in distilled water and then placed on agar, and the inoculum was spread on cellophane using a glass stick. After two weeks, spores (control in microscope) were easily scraped from cellophane and stored in 20% glycerol at -20°C.

Growth conditions

For shake-flask experiments, spores in frozen aliquots were collected by centrifugation, re-suspended in medium 707 (for rehydration), and readily separated by vortexing. Individual aliquots (about 5 × 10⁸ spores) were used to inoculate each 500 ml baffled Erlenmeyer flask containing 50 ml of the liquid media described above. Cultures were incubated at 30°C with shaking at 250 rpm. Bioreactor cultures were carried out on Mini-fermenters (Infors AG, Bottmingen, CH) that operated with a working volume of 1.5 l. Stirring was provided by Rushton-type impellers rotating at 250 rpm. Sterile air was supplied through a sparger. The bioreactors were equipped with pH electrode, PO2 electrode (polarographic), antifoam probe and Pt-100.

Erythromycin assays

Erythromycin production in solid media was assayed by bioassay. To this purpose, *S. erythraea* strains were grown in solid media (30 ml) in Petri dishes (8.5 cm). After desired time of cultivation, agar discs of 1.6 cm in diameter (with mycelium on the surface) were removed and placed into empty Petri dishes (diameter 8.5 cm). Petri dishes were then filled with soft nutrient agar seeded with *Micrococcus luteus*. Diameters of the zone of inhibition were measured after 2 days of incubation at 37°C. Agar discs containing defined amounts of > 95% pure erythromycin A (Sigma) were used as a reference. In liquid media, erythromycin was extracted as described [54]. Five hundred μL of n-butylacetate were added to 500 μL of broth fermentation, samples were vortexed for 5 min and centrifuged at 9500 g for 15 min. Then the organic phase was mixed with 500 μL buffer 25 mM K₂PO₄ pH 5 and centrifuged at 800 g for 5 min. Ten μL of the aqueous phase was injected into HPLC.

HPLC analyses were carried out using an Agilent 1100 Series HPLC system equipped with security guard Cartridges (C18 ODS 4×3 mm) and a Phenomenex-luna 5 μC18 (2) 100 Å column (250×4.6 mm). An isocratic elution mode was used according to [55] Tsuji and Goetz

Table 8 Composition of the media used in this study

| Medium                          | Composition (per liter)                                                                 | pH   |
|--------------------------------|----------------------------------------------------------------------------------------|------|
| **Complex**                    |                                                                                        |      |
| Seed medium (SM)               | 4 g peptone, 4 g yeast extract, 2 g KH₂PO₄, 4 g KH₂PO₄, 0.5 g MgSO₄·7H₂O, 10 g glucose | 7.2  |
| Yeast starch (YS)              | 2 g yeast extract, 10 g soluble starch                                                 | 7.3  |
| Oat meal yeast (OMY)           | 40 g oatmeal, 1 g yeast extract                                                       | 68-70|
| Soluble complete medium (SCM)  | 20 g soytone, 15 g soluble starch, 105 g morpholinepropanesulfonic acid, 1.5 g yeast extract, 0.1 g CaCl₂ | 7.2  |
| **Chemically defined**         |                                                                                        |      |
| MM-101                         | 7 g NH₄Cl, 3 g KH₂PO₄, 7 g KH₂PO₄, 0.25 g MgSO₄·7H₂O, 0.0138 g CaCl₂, 2 H₂O, 10 g glucose, 2 ml trace solution element (TSE) | 6.9  |

*When indicated 1.0 g casamino acids (Difco, Detroit, Mich.) and/or 2 ml (per liter) trace solution element (TSE) were added. The TSE solution composition (per liter) was: 40 mg ZnCl₂, 200 mg FeCl₃·6 H₂O, 10 mg CuCl₂·2 H₂O, 10 mg MgCl₂·6 H₂O, 4 H₂O, 10 mg Na₂B₄O₇·10 H₂O, 10 mg (NH₄)₆Mo₇O₂₄·4 H₂O*
(1978). Mobile phase was as follows: acetonitrile-methanol-0.2 M ammonium acetate-water (45:10:10:35) at pH 7.8, at flow rate of 1 ml min⁻¹ at 25°C using UV detector at 215 nm. Quantification of the erythromycin was achieved using calibration curves of peak area against injected concentration of the various erythromycin standards A, B, C (European Pharmacopoeia HPLC assay).

DNA procedures
High molecular weight genomic DNA was extracted from *S. erythraea* strains grown in 50 ml of SM medium with shaking at 28°C for 5 days (120 h). After centrifugation, the mycelium was re-suspended in 10 ml SET buffer (75 mM NaCl, 25 mM EDTA, 20 mM Tris-Cl pH 7.5) and incubated in the presence of 5 mg ml⁻¹ lysozyme for 30' at 37°C. Samples were sonicated (Sonifier sonicator Model 250/240, Brain Ultrasonic Corporation) 3 times for 30 sec, and incubated in the presence of 20 mg ml⁻¹ Proteinase K and 1.2% sodium dodecyl sulfate (SDS) for 2 h at 55°C. Nucleic acids were extracted by fenol-chloroform:isoamlylic alcohol (24:1) according to standard procedure (Sambrook and Russell, 2001) and RNA removed using 15 µg ml⁻¹ ribonuclease A. After fenol-chloroform:isoamlylic alcohol (24:1) extraction and ethanol-precipitation, high molecular-weight DNA was collected by spooling high Shepherd’s crooks [56,57].

Sequencing and assembly of the *S. Erythraea* px genome
Whole-genome shotgun DNA sequencing of *S. erythraea* Px genome was performed by MWG Biotech (Eurofins MWG Operon, Ebersberg, Germany) using frequently cutting restriction enzymes and 2- to 10-kbp fragments cloned into plasmid vectors. Cosmids (32-46 kbp inserts) were also generated from genomic DNA and end-sequenced to provide additional read-pair information, increase coverage of selected regions, and fill the gaps. Remaining gaps and ambiguities were closed using PCR products from specifically designed oligonucleotide primers. Sequence assembly was done using the Phrap assembler43 and editing was done using consed version 14. Repeats were resolved by doing a mini-assembly for the individual sections of the genome and the resulting consensus was integrated into the main genome assembly.

RNA extraction and RNA microarray experiments
The *S. erythraea* Custom GeneChip used for the NRRL2338 strain gene expression profiling by Peano et al. [24] was used here to analyze the time course gene expression profiling of the Px strain. The sequences of all probes present on the GeneChip were compared with the genomic sequence of the Px strain using BLASTN [58], finding a perfect match for all of them, both for similarity (100%) and length (25 bp).

For each time point, RNA was extracted from mycelium pellets deriving from 1-ml culture samples using the GeneElute™ total RNA Purification Kit (SIGMA), recovering it in 50 µl of Elution Solution. After extraction RNAs were quantified with a NanoDrop spectrophotometer (NanoDrop Technologies) and analyzed by capillary electrophoresis on a Agilent Bioanalyzer (Agilent). The RNA samples showing an RIN (RNA Integrity Number, a quality parameter calculated by the instrument software) value higher than 7 were processed for microarray hybridization, following the instructions for “Prokaryotic Target Preparation” (Affymetrix GeneChip® Expression Analysis Technical Manual). The protocol consists in cDNA synthesis by reverse transcription (starting with 10 µg RNA), followed by cDNA fragmentation with DNase I and labeling with Terminal Deoxynucleotidyl Transferase. The labeled cDNAs were then hybridized for 16 h at 50°C on individual GeneChips. After hybridization, GeneChips were washed and stained with streptavidin-conjugated phycoerythrin by using the Fluidic Station FS450 (Affymetrix) following the FS450_0005 Protocol. Fluorescent images of the microarrays were acquired using a GeneChip Scanner 3000 (Affymetrix). All raw data files are available in Gene Expression Omnibus under accession number GSE30600.

Genomic comparison and variation detection
The software Nucmer from the MUMmer package [59] (http://mummer.sourceforge.net) was used to investigate the presence of possibly large genomic rearrangement. All other comparative analyses have been carried out using the BLAST suite from NCBI (http://www.ncbi.nlm.nih.gov BLAST ftp://ftp.ncbi.nih.gov/blast), in particular the BLASTx and BLASTn programs, whose results were then parsed using home-made scripts. The circular plots were generated by an home-made Python script. The sequences of the 227 genes (SACX_) affected by mutations in the Px strain as compared to corresponding genes (SACE_) in the parental NRRL2338 strain have been submitted to GeneBank (accession numbers: from JN392509 to JN392716 in Additional file 7).

Microarray data analysis
The quality of the raw data obtained from microarray hybridization was assessed considering the MAS5.0 (Microarray Suite/Software, Affymetrix) control parameters after a global scaling at a target intensity of 100. Quality and control parameters as well as box plots of raw intensities highlighted the overall high quality of the data set and the absence of any outlying sample. Probe
level data was converted to expression values using both the Robust Multi-array Average (RMA) procedure [60] and the MAS5.0 algorithms. In the former case, PM values (Perfect Match) were background-adjusted, normalized using invariant set normalization, and log transformed. In the latter case, intensity levels were normalized using the Global Scaling option to target value (i.e. TGT = 100).

Genes characterized by a statistically significant modulation of the expression level during the growth time course (within-class temporal differential expression) were identified using the EDGE software package, which is based on the Optimal Discovery Procedure [61] and allows identifying genes that are differentially expressed between two or more different biological conditions or to perform significance analysis on time course experiments [62]. Whereas other methods employ statistics essentially designed for testing one gene at a time (e.g. t-statistics and F-statistics), the ODP uses all relevant information from all genes to test each gene for differential expression, thus improving the power of the test. In the particular case of a time course, ODP takes into account the ordering and spacing information provided by the time points.

Briefly, we have tested each gene by first fitting a model (e.g. natural cubic sp-lines) under the null hypothesis that there is no differential expression, and then under the alternative hypothesis that there is differential expression. A statistic is calculated to compare the goodness of fit of the two models under the two different hypotheses. The statistic is a quantification of evidence for transcriptional modulation, and the larger it is the more differentially expressed the gene appears to be. Once the statistic is calculated for each gene, a significance cut-off is applied using a false discovery rate criterion. This process, based on the calculation of the null distribution of the statistics when there is no differential expression, is accomplished through a data re-sampling technique and results in the q-value. Modulated genes are finally selected based on the q-value threshold and, eventually on a fold change limit. Differentially expressed genes were selected considering each time point replicate as an independent entry and setting the confidence threshold at q-value \( \leq 0.001 \). Hierarchical clustering and Eisen’s maps were used to group modulated genes and samples in the software package dChip. Before clustering, the expression values for a gene across all samples were standardized (linearly scaled) to have mean 0 and standard deviation 1, and these standardized values were used to calculated correlations between genes and samples and served as the basis for merging nodes. Hierarchical agglomerative clustering was carried out using Pearson correlation coefficient as distance metric and centroid as linkage method.

Chromosomal regions presenting a between-class temporal differential expression were identified using PREDA [35,63,64]. PREDA is a bioinformatic tool developed under R statistical environment for the identification of differentially expressed chromosomal regions, which accounts for variations in gene distance and density. PREDA consists of three main steps: i) computation of a statistic for ranking probes in order of strength of evidence for an expression feature; ii) adaptive bandwidth smoothing of the statistic after sorting the statistical scores according to the chromosomal position of the corresponding genes; and iii) application of a permutation test to identify differentially expressed chromosomal regions with a q-value correction for multiple tests. Transcriptional and structural information are locally integrated smoothing, along the chromosomal coordinate, an expression statistic. The smoothing procedure is approached as a non-parametric regression problem using a local variable bandwidth kernel estimator. A permutation scheme is used to identify differentially expressed regions under the assumption that each gene has a unique neighborhood and that the corresponding smoothed statistic is not comparable with any statistic smoothed in other regions of the genome. The permutation process over B random assignments allows defining a null smoothed statistic for any chromosomal position. The significance of the differentially expressed regions (i.e. the p-value) is computed as the probability that the random null statistic exceeds the observed statistic over B permutations. Once the distribution of empirical p-values has been generated, the q-value is used to identify differentially expressed chromosomal regions according to Storey and Tibshirani [65].

Additional material

Additional file 1: Cell cycle-regulated genes in S. erythraea NRRL 2338.
Additional file 2: Cell cycle-regulated genes in S. erythraea Px.
Additional file 3: Genes up-regulated in both S. erythraea NRRL 2338 and S. erythraea Px during phase a.
Additional file 4: Genes specifically up-regulated during phase a in S. erythraea Px.
Additional file 5: Genes specifically up-regulated during phase a in S. erythraea NRRL 2338.
Additional file 6: Growth of S. erythraea NRRL 2338 and S. erythraea Px on SCM or MM-101 agar with or without amino acid and/or adenine supplement.
Additional file 7: Accession numbers.

Acknowledgements

We are grateful to Prof. S. Donadio for gift of S. erythraea wild type strain NRRL2338, and Dr. S. M. Tredici and Dr. C. Marcuccio for technical assistance. This work was partially supported by grant from the Italian Ministry for University and Research (MIUR): MIUR Cofin 2008 - prot. 200823294H to PA,
FIRB-MIUR NG-LAB Research Program, Project RBLA03ER38_004, to GDB, and from the Lombardy Region: “Accordo Quadro CNR-Regione Lombardia - Sottoprogetto 4, Piattaforma 2” for Antibiotics development. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author details
1Institute of Biomedical Technologies, National Research Council, Via Fratelli Cervi 93, 00090 Segrato, Italy. 2Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, Via Monteroni, 73100 Lecce, Italy. 3Operative Unit of Lecce, CNR- Institute of Sciences of Food Production (ISPA), via Provinciale Lecce-Monteroni, 73100 Lecce, Italy. 4Center for Genome Research, Dept. of Biomedical Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41100 Modena, Italy.

Authors’ contributions
Conceived and designed the experiments: GDB PA. Performed the experiments: CP AT GC DP SB. Analyzed the data: CP, SB, GDB, PA. Wrote the paper: CP, PA. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 8 September 2011 Accepted: 8 March 2012 Published: 8 March 2012

References
1. Champness WC, Chater KF: The regulation and integration of antibiotic production and morphological differentiation in Streptomyces spp. Washington, DC: American Society for Microbiology; 1994.

2. Adrio JL, Demain AL: Genetic improvement of processes yielding microbial products. FEBS Microbiol Rev 2006, 30(2):187-214.

3. Baltz RH: Genetic methods and strategies for secondary metabolite yield improvement in actinomycetes. Antonie Van Leeuwenhoek 2001, 79(3-4):251-259.

4. Vinc WA, Byng G: In Manual of Industrial Microbiology and Biotechnology. 2 edition. Edited by: Demain AL, Davies JE, Atlas RM, Cohen G, Hershberger CL, Hu W-S, Sherman DH, Wilson RC, Wu JHD. Washington, DC: American Society for Microbiology; 1999:103-113.

5. Blanchard AP, Hood L: Sequence to array: probing the genome’s secrets. Nat Biotechnol 1996, 14(13):1649.

6. Bro C, Nielsen J: Impact of ‘ome’ analyses on inverse metabolic engineering. Metab Eng 2004, 6(3):204-211.

7. Lander ES: The new genomics: global views of biology. Science 1996, 274(5287):536-539.

8. Lander ES: Array of hope. Nat Genet 1999, 21(1 Suppl):3-4.

9. Lee SY, Lee DY, Kim TY: Systems biotechnology for strain improvement. Trends Biotechnol 2005, 23(7):349-358.

10. Stafford DE, Stephanopoulos G: Metabolic engineering as an integrating platform for strain development. Curr Opin Microbiol 2001, 4(3):336-340.

11. van der Werf MJ, Jellema RH, Hankemeier T: Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. J Ind Microbiol Biotechnol 2005, 32(6):234-252.

12. Donadio S, Monciardi P, Sasso M: Polyketide syntheses and nonribosomal peptide synthetases: the emerging view from bacterial genomes. Nat Prod Rep 2007, 24(5):1073-1109.

13. Katz L, Donadio S: Macrolides. Biotechnology 1995, 28:385-420.

14. Leadlay PF: Combinatorial approaches to polyketide biosynthesis. Curr Opin Chem Biol 1997, 1(2):162-168.

15. Hutchinson CR, Decker H, Madduri K, Otten SL, Tang L: Genetic control of antibiotic overproduction by rpsL and rsmG mutants of various actinomycetes. Appi Environ Microbiol 2009, 75(14):4919-4922.

16. McDaniel R, Licari P, Khosla C: Genetic improvement of processes yielding microbial products. Nat Microbiol Biol 2003, 16(3):187-214.

17. Corcoran JW: Biochemical mechanisms in the biosynthesis of the erythromycins. In Antibiotics IV Biosynthesis. Edited by: Corcoran JW. New York: Springer-Verlag edn, 1981:112-174.

18. Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF: An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. Nature 1990, 348(6297):176-178.

19. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L: Modular organization of genes required for complex polyketide biosynthesis. Science 1991, 252(5006):675-679.

20. Donadio S, Staver MJ: S11136, an insertion element in the erythromycin gene cluster of Saccharopolyspora erythraea. Gene 1993, 126(1):147-151.

21. Weber JM, Leung JO, Mainie GT, Potenz L, Paulus T, Devitt JP: Organization of a cluster of erythromycin genes in Saccharopolyspora erythraea. J Bacteriol 1990, 172(5):2357-2363.

22. Reeves AR, English RS, Lampel JS, Post DA, Vanden Boom TJ: Transcriptional organization of the erythromycin biosynthetic gene cluster of Saccharopolyspora erythraea. J Bacteriol 1999, 181(22):7095-7106.

23. Ohyink M, Samborskyy M, Leskan JB, Schotz N, Dickens S, Haydock SF, Leadlay PF: Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. Nat Biotechnol 2007, 25(4):443-447.

24. Peano C, Bicciato S, Corti G, Ferrari F, Rizzi E, Bonnal RJ, Bordoni R, Albertini A, Bernardi LR, Donadio S, et al: Complete gene expression profiling of Saccharopolyspora erythraea using GeneChip DNA microarrays. Microb Cell Fact 2007, 6:57.

25. Yang C, Liu S, Yu YT, Li YX, Li YY: Identifying modules of coexpressed transcript units and their organization of Saccharopolyspora erythraea from time series gene expression profiles. PLoS One 8(5) e2126.

26. Elliott MA, Dalmy F, Passantino R, Chater K, Leskiw B: The bldD gene of Streptomyces coelicolor A3(2): a regulatory gene involved in morphogenesis and antibiotic production. J Bacteriol 1998, 180(6):1549-1555.

27. Elliott MA, Bubik MJ, Burren MJ, Leskiw BK: BldD is a direct regulator of key developmental genes in Streptomyces coelicolor A3(2). Mol Microbiol 2001, 40(1):257-269.

28. Chng C, Lunn AM, Vroom JA, Koo CM: A key developmental regulator controls the synthesis of the antibiotic erythromycin in Saccharopolyspora erythraea. Proc Natl Acad Sci USA 2008, 105(32):11346-11351.

29. Reeves AR, Cernota WH, Brikun IA, Wesley RK, Weber JM: Engineering precursor flow for increased erythromycin production in Aeromicrobium erythreum. Metab Eng 2006, 8(4):300-312.

30. Reeves AR, Brikun IA, Cernota WH, Leach BI, Gonzalez MC, Weber JM: Effects of methylmalonyl-CoA mutase gene knockouts on erythromycin production in carbohydrate-based and oil-based fermentations of Saccharopolyspora erythraea. J Ind Microbiol Biotechnol 2006, 33(7):600-609.

31. Tanaka Y, Komatsu M, Okamoto S, Tokuyama S, Kaji A, Ikeda H, Ochi K: Antibiotic overproduction by rpsL and rsmG mutants of various actinomycetes. Appl Environ Microbiol 2009, 75(14):4919-4922.

32. Carata E, Peano C, Tredici SM, Ferrari F, Tala A, Corti G, Bicciato S, De Bellis G, Alfano P: Phenotypes and gene expression profiles of Saccharopolyspora erythraea rifampicin-resistant (rif) mutants affected in erythromycin production. Microb Cell Fact 2009, 8:18.

33. Zhang H, Wang Y, Wu J, Skalina K, Pfeifer BA: Complete biosynthesis of erythromycin A and designed analogs using E. coli as a heterologous host. Chem Biol 2010, 17(11):1232-1240.

34. Lum AM, Huang J, Hutchinson CR, Kao CM: Reverse engineering of industrial pharmaceutical-producing actinomycete strains using DNA microarrays. Metab Eng 2004, 6(3):186-196.

35. Callegaro A, Basso D, Bicciato S: A locally adaptive statistical procedure (LAP) to identify differentially expressed chromosomal regions. Bioinformatics 2006, 22(21):2658-2666.

36. Katz L, Khosla C: Antibiotic production from the ground up. Nat Biotechnol 2007, 25(4):428-429.

37. Gopal B, Haire LF, Cox RA, John Colston M, Smerdon SD, Dodson G: The crystal structure of NusB from Mycobacterium tuberculosis. Nat Struct Biol 2003, 10(6):475-478.

38. Doherty GP, Meredith DH, Lewis PJ: Subcellular partitioning of transcription factors in Bacillus subtilis. J Bacteriol 2006, 188(11):4101-4110.
