Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics

Stefano Donadio,* Paolo Monciardini and Margherita Sosio

Received (in Cambridge, UK) 27th February 2007
First published as an Advance Article on the web 10th May 2007
DOI: 10.1039/b514050c

Covering: bacterial genome sequences to 2005 and post-genomic literature to June 2006

A total of 223 complete bacterial genomes are analyzed, with 281 citations, for the presence of genes encoding modular polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS). We report on the distribution of these systems in different bacterial taxa and, whenever known, the metabolites they synthesize. We also highlight, in the different bacterial lineages, the PKS and NRPS genes and, whenever known, the corresponding products.

1 Introduction

During the last two decades, enormous progress has been made in elucidating the biosynthesis of hundreds of secondary metabolites, mostly from microorganisms. There is little doubt that the major contribution to this wealth of knowledge has resulted from the application of DNA sequencing to secondary metabolism, facilitated by the fact that microorganisms usually carry all the relevant genes in a contiguous DNA segment known as a gene cluster. These studies were therefore chemistry-driven, *i.e.* genes were characterized because they participated in the synthesis of known natural products. The data obtained have confirmed that the biosynthesis of a large number of natural products requires the participation of sophisticated molecular machines known as polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS).

There is also little doubt that, around the turn of the millennium, we have fully witnessed the impact of the genomic revolution in our understanding of biology. One of the most outstanding advances from the genomic revolution has been in prokaryotic biology, with over 250 complete bacterial genomes publicly available. It should be noted that bacterial genomes were initially sequenced during the ground work necessary for bigger projects (*i.e.* the human genome). However, in an era when antibiotic resistance has become a serious medical concern, it was soon realized that an inventory of all the genes present in a bacterial species, as provided by bacterial genomics, would provide all possible targets for the search of new antibiotics,[1] all candidate proteins for vaccine development,[2] or a better understanding of pathogens' biology. For these reasons, the choice of sequenced strains is heavily biased towards those which are pathogenic to humans, plants or animals.

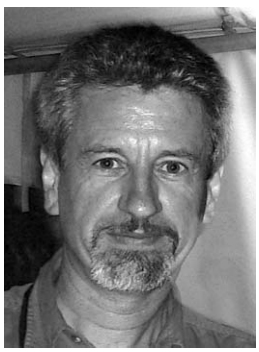
This top-down approach of sequencing entire bacterial genomes has also led to the unexpected outcome that many strains harbor genes highly related to those involved in natural product formation. With few exceptions, bacterial genomes were not specifically analyzed for their potential to synthesize natural products. Here, we try to merge the two worlds of natural product biosynthesis and of bacterial genomes, by reporting the occurrence of typical genes for secondary metabolism in bacterial genomes. We chose to limit our analysis to PKSs and NRPSs, since these two classes participate in the synthesis of many diverse secondary metabolites. In addition, they usually encode easily recognizable large multimodular polypeptides that often comprise a large

KtedoGen, via Fantoli 16/15, 20132 Milan, Italy. E-mail: stefano.donadio@ktedogen.com

Stefano Donadio received his doctoral degree in Chemistry / Biochemistry from the University of Naples, Italy. He then did postdoctoral work at the Johns Hopkins University, at the International Institute of Genetics and Biophysics and at the University of Wisconsin–Madison. In 1988 he started his career in the pharmaceutical industry, first at Abbott Laboratories and then at the Lepetit Research Center. A cofounder of Biosearch Italia (1996), he left the company in 2004 to start KtedoGen. For over 20 years, his research interests have been antibiotics and antibiotic-producing microorganisms, especially the characterization and manipulation of antibiotic biosynthesis pathways, novel screening tools for antibiotic discovery, and the discovery of novel antibiotic producing actinomycetes.

Paolo Monciardini received his education at the University of Parma, Italy, studying Biology and receiving his Ph.D. in 1998 with research on stress-responsive promoters in transgenic plants at the Department of Environmental Sciences. In 1999 he joined the Microbial Technology department at Biosearch Italia, and has followed Biosearch's fate till 2007, when he joined Ktedogen. For the past 7 years, his main research interests have been on filamentous actinomycetes, particularly on bacterial diversity and taxonomy, and on the identification and characterization of new actinomycete lineages as potential antibiotic producers.

Margherita Sosio received her doctoral degree in Agricultural Sciences at the University of Milan, Italy. She then did postdoctoral work with Prof. R. Huetter at the ETH in Zuerich on actinomycete genetics, before joining the Biotechnology Department at Farmitalia Carlo Erba. In 1989, she moved to the Lepetit Research Center. In 1997 she joined the newly formed Biosearch Italia as the Microbial Genetics group leader, following Biosearch's fate before joining Ktedogen in 2007. For over 20 years, her research interests have focussed on the characterization of antibiotic biosynthesis and resistance genes, on the identification and validation of new targets and assays for antibiotic discovery, on developing novel tools for manipulation of antibiotic pathways, and on the discovery of novel taxa of antibiotic producers.



Stefano Donadio

Paolo Monciardini

Margherita Sosio

fraction of gene clusters. This report is also limited to bacteria, which represent the largest set of sequenced microbes. As the reader will soon appreciate, the top-down approach of whole-genome sequencing has provided plenty of gene sequences and relatively modest chemical information. Nonetheless, the emerging view presented in this report may stimulate further analyses.

2 Thiotemplate modular systems

PKSs and NRPSs are key players in the synthesis of natural products, since they carry out the oligomerization of small building blocks into often complex structures. Excellent reviews have been published since 2005 on PKSs[3] and NRPSs,[4] and the reader is referred to these works and to previous literature cited therein for further details. For the purpose of this review, only modular PKSs will be covered and the term "PKS" will refer to modular PKSs only. Other types of PKSs will be explicitly indicated.

PKSs and NRPSs are molecular assembly lines that direct product formation on a protein template. Both systems accomplish their task by maintaining reaction intermediates covalently bound as thioesters on the same phosphopantetheine prosthetic group. Usually, a 4′-phosphopantetheinyltransferase (PPTase) specific for modular enzymes modifies an active site serine residue in the thiolation (T) and acyl carrier protein (ACP) domains of NRPSs and PKSs, respectively, to generate the corresponding holoenzymes.[5] In addition, each monomer is handled by a separate set of enzymatic domains known as a module, and usually there are as many modules as monomers incorporated in the final product. For these reasons, PKSs and NRPSs are thiotemplate modular systems (TMS), and we will use the generic term TMS genes (enzymes) to indicate NRPSs, PKSs or both. We will briefly review the *modus operandi* of TMS enzymes below, to help the reader in understanding the following sections.

NRPSs use amino or hydroxy acids as building blocks, catalyzing the formation of amide or ester bonds, respectively. As shown in Fig. 1, each NRPS module consists of three core domains: an adenylation (A) domain, which selects the cognate amino acid, activates it as an amino acyl adenylate and transfers it to the T domain (also known as peptidyl carrier protein, or PCP) where a thioester bond is formed, a condensation (C) domain, responsible for peptide bond formation between the amino acid present on the T domain of the same module and the peptidyl intermediate bound to the T domain of the preceding module, and the T domain itself. Usually, all elongation modules present these core domains. A dedicated loading module (carrying just A and T domains) and a termination module, containing a thioesterase (TE) domain, usually complete the NRPS assembly line. Additional reactions may be carried out by specialized domains within a module, such

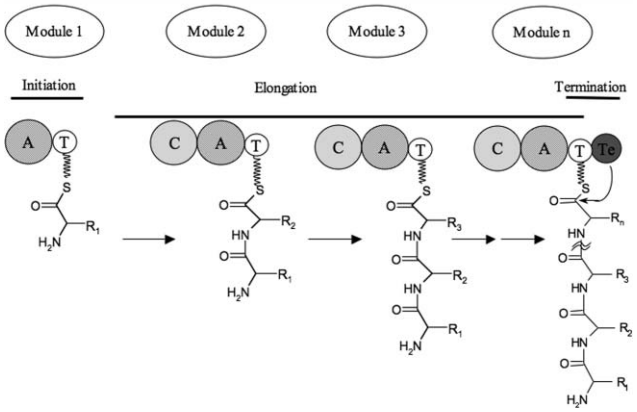


Fig. 1 Basic steps during nonribosomal synthesis of peptides.

as amino acid epimerization (E), methylation (M) and reduction (R) activities. Additional variations include the presence of a heterocyclization domain (Cy) in place of a C domain, or the occurrence of C domains capable of epimerization.

PKSs generate polyketide chains through the oligomerization of small carboxylic acids. Each PKS module consists of three core domains: an acyltransferase (AT) domain, which selects the appropriate extender unit (usually malonyl-CoA or methylmalonyl-CoA) and transfers it to the ACP domain where a thioester bond is formed, and a ketosynthase (KS) domain, responsible for decarboxylative condensation between the extender unit present on the ACP domain of the same module and the polyketide intermediate bound to the ACP domain of the preceding module. All elongating modules present these core domains, while the loading module lacks a functional KS domain and the last module contains an additional TE domain, for release of the finished polyketide from the PKS. Most PKS modules contain additional domains for processing the newly formed β-keto: the β-ketoreductase (KR), the dehydratase (DH) and the enoylreductase (ER) domains carry out the reactions depicted in Fig. 2. Occasionally, M domains are present in PKSs.

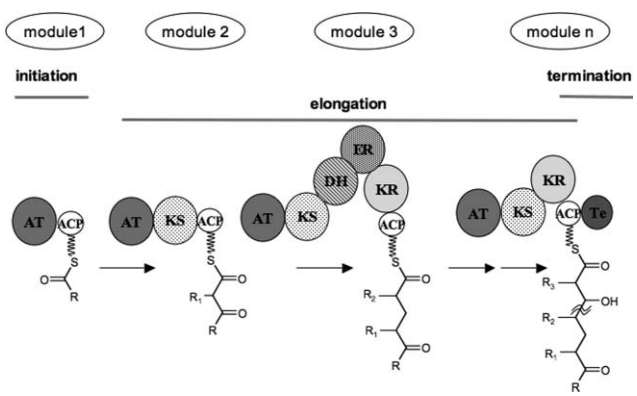


Fig. 2 Basic steps during synthesis of polyketides.

Each TMS domain carries out the same basic reaction in different modules, but different domains may act on different substrates or produce different stereoisomers. Thus, one or more domains must be responsible for substrate specificity or stereoselectivity. The large number of sequenced genes involved in the synthesis of natural products of defined structure and stereochemistry has led to the proposal of *in silico* rules for predicting substrate specificity.[6] Thus, the amino acids recognized by A domains,[7] the dicarboxylic acids loaded by AT domains[8] and the stereoselectivity of KR domains[9] are strongly associated with defined sequence motifs in the corresponding domains. These motifs can thus be used to predict the substrate specificity of TMS modules uncovered by genome sequencing. In addition, the domain composition of each module can be used to predict the ancillary reactions performed by that module and the number of modules directly indicates the length of the oligomer. There are, however, exceptions to template:product co-linearity, as modules are now known to be both skipped and iterated during the normal biosynthetic processes.[3,10] These exceptions notwithstanding, the chemical structure of a natural product synthesized by TMS enzymes can be predicted with reasonable approximation by the number of modules, the domain composition of each module and by the *in silico* specificity of relevant domains.

3 A global view of bacterial genomes

At the time of this writing, there were over 250 bacterial strains for which finished genome sequences had been generated.[11] We chose to limit this report to finished genomes described in the literature, referring occasionally to relevant TMS genes from unpublished genomes. It should be noted that the annotations of individual genes, as they appear in the database, are often complemented by informative comments in the published literature. However, the majority of bacterial genomes were not analyzed for their genetic repertoire in natural product formation, and different criteria are often used for gene annotation. Thus, relying solely on available annotations or published information does not provide a consistent picture on occurrence and role of TMS genes in the published bacterial genome sequences. For these reasons, we surveyed bacterial genomes for the presence of TMS genes using either the authors' annotations or Blast[12] searches. Then, we catalogued all those genes that either encoded at least two TMS domains or were located in close proximity to other TMS genes. Accordingly, about 50% of the 223 analyzed bacterial strains harbor TMS genes, which extend for almost 4.5 Mb, or 0.6% of the cumulative 738 Mb of available genome sequences (Table 1). In some instances, the sequenced bacterial strain or a close relative produced a known natural product, and the corresponding gene cluster was identified. In other cases, a TMS cluster uncovered from the genome sequence was highly correlated to an orthologous cluster directing the synthesis of a known compound, which was then actually detected. Finally, in a few cases, *in silico* predictive tools were used to propose a chemical structure from a newly discovered TMS cluster. However, the number of TMS clusters without an associated natural product far outweighs the number of chemical structures. Because of the limited chemical information, TMS genes are reported from a taxonomic perspective. However, there are occurrences of distantly related bacteria harboring TMS clusters for identical (or very similar) natural products, and these will be discussed in a separate chapter of this report. Finally, it should be noted that most of the information derived from genome sequences refers to *putative* enzymes. However, we decided to omit

Table 1 Analyzed genomes by phylum

Phylum	Genomes[a]	Size[b]	TMS genes[c]	Density[d]
Actinobacteria	18	70 044 365	452 991	1.940
Aquificae	1	1 590 791	0	0.000
Bacteroidetes	5	22 776 893	2628	0.035
Chlamydiae	9	11 601 785	0	0.000
Chlorobi	1	2 154 946	0	0.000
Cyanobacteria	8	26 666 055	51 826	0.583
Deinococci	2	5 411 638	0	0.000
Firmicutes	63	159 832 396	133 511	0.251
Fusobacteria	1	21 74 500	0	0.000
Planctomycetes	1	7 145 576	11 197	0.470
α-Proteobacteria	25	77 768 614	39 940	0.154
β-Proteobacteria	13	59 473 882	236 438	1.193
γ-Proteobacteria	59	247 393 752	540 029	0.655
δ-Proteobacteria	4	15 226 925	0	0.000
ε-Proteobacteria	6	10 640 511	0	0.000
Spirochaetes	6	15 806 532	0	0.000
Thermotogae	1	1 860 725	0	0.000
Total	223	737 569 886	1 468 560	0.597

[a] Different strains with published sequenced genomes. [b] Cumulative genome size, in bp. [c] Cumulative size of TMS polypeptides, in aa. [d] Percentage of TMS genes, obtained dividing cumulative size of TMS genes by cumulative genome size.

the repetitive use of the term "putative", pointing out specifically those cases where experimental proof has been obtained.

The distribution of genome sequences and TMS gene content in bacterial phyla is reported in Table 1. For convenience, the Proteobacteria, which include almost half of the analyzed genomes, were split into the five subphyla. The most represented (sub)phyla are the Firmicutes and the γ-Proteobacteria, with 63 and 59 strains, respectively, reflecting the high number of pathogens in these two phyla. The publicly available genomes are biased towards pathogens, which represent a small fraction of bacterial diversity, and do not include representatives from the myxobacteria (δ-Proteobacteria), a group well known for their richness in TMS genes.[13] With these caveats in mind, it can be observed that the presence of TMS genes is not uniformly distributed within the various taxa, with the γ-Proteobacteria, Actinobacteria and β-Proteobacteria contributing most TMS genes (Table 1). This reflects the long-known view that not all bacteria are capable of producing secondary metabolites.[14]

It has been observed that bacterial genomes contain a roughly constant coding density, which implies that, with few exceptions (*i.e.* obligate parasites or decaying genomes), the number of genes is proportional to genome size and this relation holds also for specific classes (*e.g.* regulatory genes[15]). However, it does not apply to TMS genes (Fig. 3), as expected from their uneven distribution in bacterial taxa. Below a threshold of *ca.* 3 Mb, TMS genes are either rare or absent. Above 5 Mb, there appears to be a linear correlation between genome size and content of TMS genes, but several strains lie above or below a putative regression line. Clearly, additional genome sequences must be obtained from a more representative set of bacteria before we can understand if any correlation exists between genome size and content of TMS genes.

Within the set of available genomes, NRPSs seem to be more common than PKSs, which in turn are more abundant than mixed systems. However, given the relatively small number

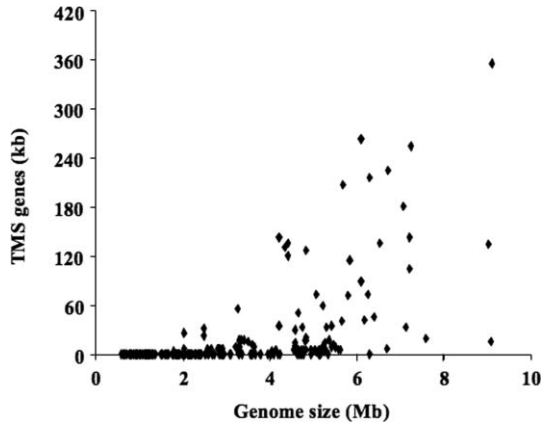


Fig. 3 Correlation between genome size (Mb) and TMS genes (kb) in published bacterial genomes.

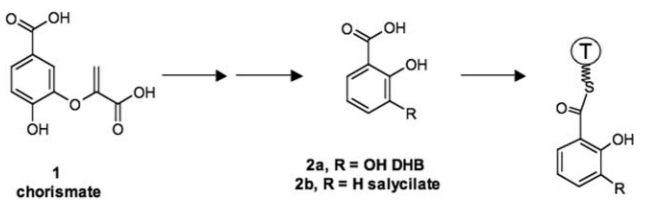
of TMS-encoding genomes, we have not attempted to analyze the individual distribution of NRPS, PKS or mixed systems. Furthermore, no analysis was performed on PPTases, and we only report relevant observations made by the authors on these enzymes.

4 Commonly encountered natural products

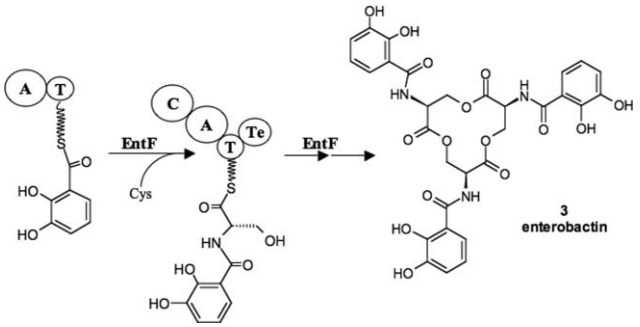
Some relatively similar TMS clusters are found in rather unrelated strains. In most instances, these clusters direct the synthesis of siderophores, compounds that are probably essential for growth in nature, where free iron is not abundant. Thus, unrelated bacteria have adopted similar strategies for scavenging iron, but the occurrence of a defined siderophore is not a taxon-uniform trait. The pathways to these common compounds will be reviewed before describing other TMS genes.

4.1 Cathechol-based iron chelating compounds

Cathechol siderophores are widespread natural products produced by many unrelated bacteria. The biosynthetic routes to cathechol siderophores have been reviewed in detail,[16] and are just summarized here. A common step is the activation of an aryl moiety (Scheme 1). Chorismate **1** is transformed into 2,3-dihydroxybenzoate (DHB) **2a** by the action of EntC, EntB and EntA (or their orthologs). EntB also carries a T domain, which is loaded with DHB **2a** by the aryl-activating adenylating protein EntE. The activated DHB then follows different routes to the various siderophores. In enterobactin **3** formation (Scheme 2), the monomodular NRPS EntF activates L-serine, forming the dipeptidyl DHB-seryl intermediate, which then cyclotrimerizes, possibly through the formation of a TE-bound ester intermediate to yield enterobactin **3**.

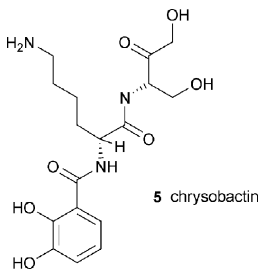


Scheme 1

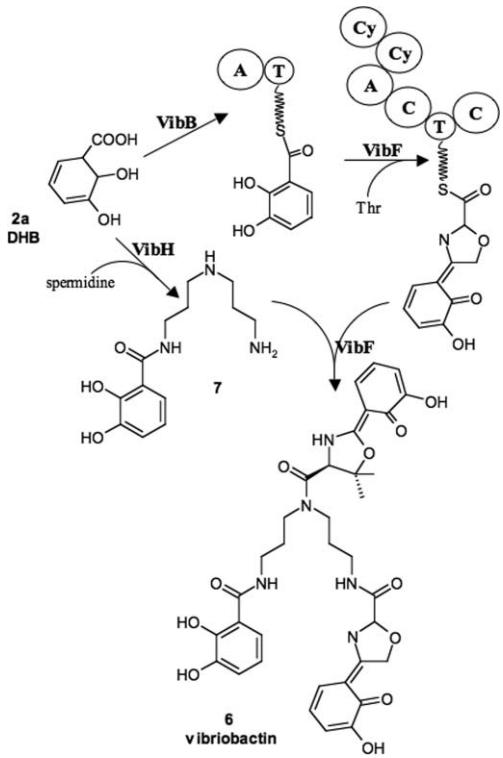


Scheme 2

A similar route is followed for the production of bacillibactin **4**, except that the dimodular NRPS DhbF incorporates glycine and L-threonine prior to trimerization and cyclization (Scheme 3).[17] A dimodular NRPS is encoded also by the cluster responsible for the synthesis of chrysobactin **5** in *Erwinia chrisantemi*, but no trimerization occurs in this case.[18]



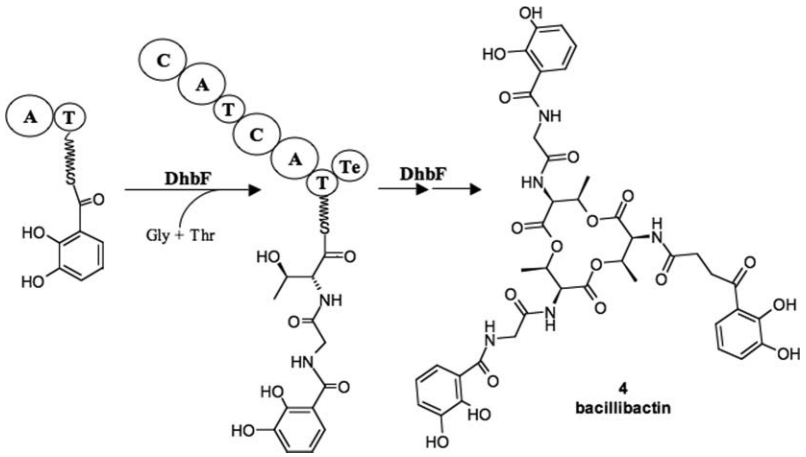
A slightly different situation is found for vibriobactin **6**. In this case (Scheme 4), two different enzymes utilize the VibE-activated DHB. VibH, a free-standing C domain, links DHB **2a** to N6 of norspermidine generating the DHB-norspermidine intermediate **7**, while VibF fuses DHB **2a** to threonine followed by generation of an oxazoline ring by condensation and dehydration. At this stage, VibF performs two cycles, fusing the DHB-oxazolyl intermediate first to N1 and then to N3 of **7**, leading to vibriobactin **6**. VibF carries two Cy and two C domains, and is encoded by a gene physically unlinked to the cluster which includes *vibBECAH* (*Vc* in Fig. 4).[19]



Scheme 4

4.2 Prodiginines

Prodiginines **8** are a large family of red-pigmented antibiotics produced by actinomycetes and other eubacteria, with a characteristic tripyrrole moiety. The biosynthetic route has been described for the actinobacterium *Streptomyces coelicolor*[20] and has been recently reviewed for *Serratia*.[21] It proceeds in both cases through the separate formation of monopyrrole and bipyrrole moieties, which are then fused to give the tripyrrole end product. Biosynthetic clusters (Fig. 5) have been described for *S. coelicolor*[20] and from the γ-proteobacteria (two species of *Serratia*[22] and *Hahella chejuensis*[23]). All *H. chejuensis* genes (except for two) are conserved and arranged in the same order as in the *Serratia* clusters (*pig*). The Proteobacteria clusters are smaller than the *S. coelicolor*



Scheme 3

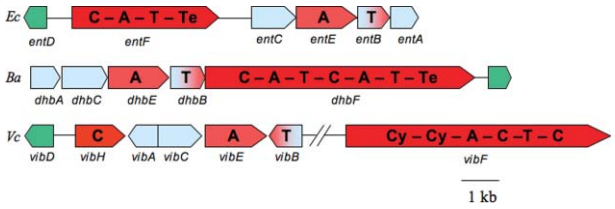
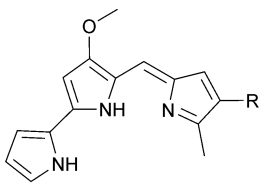


Fig. 4 Clusters for different cathechol-based siderophores: the enterobactin cluster from *E. coli* K12 (*Ec*), the bacillibactin cluster from *B. anthracis* Ames (*Ba*) and the vibriobactin cluster from *V. cholerae* O395 (*Vc*). Biosynthetic genes are indicated by arrows, lines represent other genes (not to scale). Colour codes correspond to different functions: red, NRPS domains; light blue, DHB biosynthesis; green, PPTase. Letters in arrows indicate NRPS domains. The *dhb* cluster is virtually identical in all published genomes from *B. anthracis*, *B. cereus*, *O. iheyensis*, *B. subtilis* and *B. licheniformis*, with the exception that the PPTase gene is absent in the latter two species. The *vibF* gene is unlinked to the vibriobactin cluster. Clusters 1 and 3 also include genes for siderophore transport and uptake (not shown).

red cluster, and they apparently lack some enzymes needed for prodigiosin biosynthesis, which are supposed to be mutuated from other pathways.[22] All clusters also encode a PPTase.



8 R = C_5H_{11} prodigiosin
R = $C_{11}H_{23}$ undecylprodigiosin

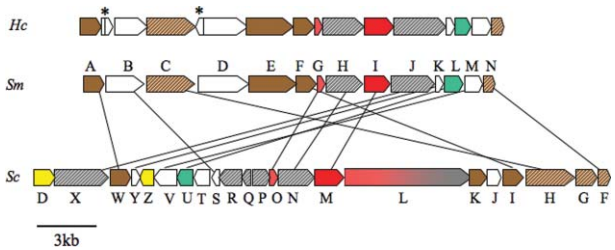
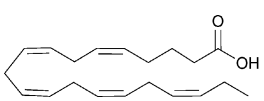


Fig. 5 The prodigiosin clusters from *H. chejuensis* (*Hc*), *Serratia marcescens* ATCC 274 (*Sm*) and *S. coelicolor* (*Sc*). Color codes are: red, NRPS; grey, PKS; green, PPTase; yellow, regulatory; brown, biosynthesis; white, unknown. Striped and filled (except green and yellow) arrows denote genes for the synthesis of the monopyrrole and bipyrrole moieties, respectively. Lines indicate pairs of ortholog genes in the *Sm* and *Sc* clusters. Asterisks indicate the two genes in *Hc* without orthologs in *Sm*.

4.3 Polyunsaturated fatty acids

Polyunsaturated fatty acids (PUFAs) **9**, essential components of membrane lipids or hormone precursors in eukaryotes, have also been identified in marine bacteria where, in contrast to eukaryotes, they are synthesized by PKS-like systems.[24] PKS genes resembling the prototype PUFA synthase are found in the γ-Proteobacteria



9 eicosanpentanoic acid (PUFA)

Colwellia psychrerythraea,[25] *Photobacterium profundum*[26] and *Shewanella denitrificans*,[11] as well as in *S. coelicolor*[27] (Fig. 6). Cluster organization is extremely conserved in the three γ-Proteobacteria.

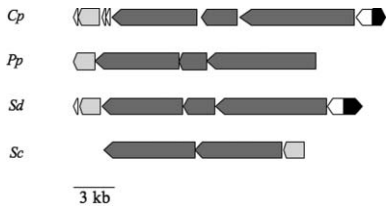


Fig. 6 The polyunsaturated fatty acid loci from *Colwellia psychrerythrea* (*Cp*), *Photobacterium profundum* (*Pp*), *Shewanella denitrificans* (*Sd*) and *Streptomyces coelicolor* (*Sc*). Color codes are: light grey, dioxygenase; dark grey, PKS; black, PPTase; white, unknown.

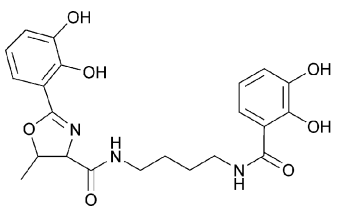
5 Phylum proteobacteria

This phylum comprises the largest number of sequenced genomes (107), and the five different classes of α-, β-, γ-, δ- and ε-Proteobacteria will be analyzed separately. Orthologous TMS clusters involved in the synthesis of iron-chelating compounds are found in diverse representatives of this phylum, and will be illustrated before proceeding with a taxon-based description of other TMS genes.

5.1 Common natural products

5.1.1 Cathechol siderophores

Clusters for the cathechol siderophores brucebactin, enterobactin **3**, chrysobactin **5**, photobactin **10** and vibriobactin **6** are found in members of the α-, β- and γ-subdivisions of the Proteobacteria, while they have not been found in the few sequenced genomes from the δ- and ε-subdivisions (Table 1).



10 photobactin

All sequenced *Escherichia coli* strains possess identical clusters for enterobactin **3** synthesis, including the *feuB* and *fepBCDEG* genes (for Fe-siderophore transport) and the enterochelin esterase *fes* (for intracellular iron release). Enterobactin synthesis and transport genes are subject to positive selection in uropathogenic *E. coli* and *entF* is upregulated during urinary tract infection.[28] Substantially identical clusters are found in the *Salmonella* and *Shigella* genomes, although some genes are inactivated by point mutations (Fig. 7).

Chromobacterium violaceum[29] and *H. chejuensis*[23] each contain two putative clusters for cathechol siderophores (Fig 7). In *C. violaceum*, one cluster (*Cv1* in Fig. 7) is reported as being involved in the synthesis of enterobactin **3**, and actually contains all the genes required for DHB **2a** synthesis and an *entF* ortholog, together with genes for the receptor and uptake of the iron–siderophore complex. The second cluster (*Cv2* in Fig. 7) encodes proteins for ferric siderophore receptor and uptake, a homolog (54% identity) of the *E. chrisantemi* bimodular NRPS CbsF,[18]

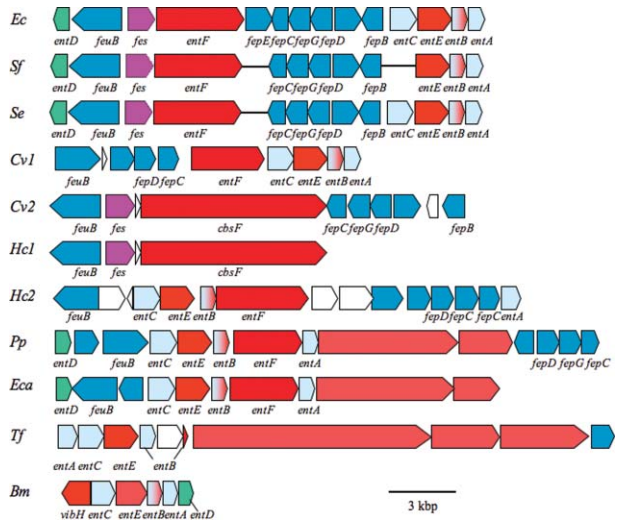


Fig. 7 Clusters for cathechol siderophores from different strains. Enterobactin clusters from: *Ec*, *E. coli* K12 (virtually identical in the other *E. coli* strains from Table 5 and in *S. typhimurium*, *S. enterica choleraesuis* and *paratyphi*); *Sf*, *Shigella flexnerii* 2457T (virtually identical in strain 301); *Se*, *S. enterica* typhi Ty2 (virtually identical in strain CT18). Other siderophore clusters from: *C. violaceum* (*Cv1* and *Cv2*), *H. chejuensis* (*Hc1* and *Hc2*), *P. profundum* (*Pp*), *E. carotovora* (*Eca*), *T. fusca* (*Tf*). The brucebactin cluster from *B. melitensis* (*Bm*) is identical in the other sequenced *Brucella*, with the exception of *B. abortus* 9-941 in which the PPTase is inactivated by a point mutation. Orthologs are indicated using the *E. coli* nomenclature (except for *cbsF*). Colour codes are: red, NRPS; light blue, DHB biosynthesis; green, PPTase; dark blue, Fe-siderophore transport; purple, Fe-siderophore esterases; white, unknown. Black lines in *Sf* and *Se* indicate genes inactivated by point mutations.

responsible for chrysobactin **5** formation, but it lacks genes for DHB **2a** synthesis. If the two clusters are coordinately expressed, then the DHB **2a** formed by the enzymes specified by cluster *Cv1* may well serve as a substrate for the CbsF homolog. A similar situation also occurs in *H. chejuensis* (*Hc1* and *Hc2* in Fig. 7).

Different *Erwinia* strains produce chrysobactin **5**.[18,30] In *Erwinia carotovora* subsp. *atroseptica*,[31] however (*Eca* in Fig. 7), in addition to orthologs of the *entABCDE* genes, a gene is present encoding a homolog of the monomodular EntF component, and not the dimodular CbsF required for chrysobactin **5** synthesis. This cluster also includes genes for two additional NRPSs (containing C-A-T-E-C and A domains). An orthologous cluster is present in *P. profundum*[26] (*Pp* in Fig. 7).

The cluster directing the synthesis of brucebactin, a structurally uncharacterized siderophore from *Brucella* spp.[32] (*Bm* in Fig. 7), is similar to that for vibriobactin **6** from *Vibrio cholerae* O395[19] (*Vc* in Fig. 4), but it lacks component F. In this case, the VibH homolog is supposed to fuse a small unidentified nucleophile to the DHB moiety as the final step in brucebactin synthesis.[32]

5.1.2 Pyochelin

This siderophore, first detected in *Pseudomonas aeruginosa*, is synthesized by the successive addition and cyclization of two cysteine molecules to salicylate **2b**.[16] The biosynthetic route is summarized in Scheme 5. The overall strategy resembles that for enterobactin **3** and related compounds. The aryl cap at the N-terminus of pyochelin **11** derives from salycilate **2b**, which is made from chorismate **1** by the action of PchA and PchB (Scheme 1). Salycilate **2b** is then adenylated by PchD, transferred to the first T domain of PchE, and condensed with two cysteines by the consecutive action of the monomodular NRPSs PchE and PchF, with concomitant formation of thiazoline rings. The C-terminal thiazoline, once reduced to thiazolidine by PchG, is methylated by the M domain embedded in PchF. The thioesterase PchC may be involved in release of the final product.[16] The *P. aeruginosa* pyochelin **11** cluster includes genes for an ABC transporter (*pchH* and *pchI*), the ferric pyochelin receptor FtpA and the regulator PchR (Fig. 8). *Pseudomonas fluorescens* Pf-5[33] harbors an orthologous cluster with the same genes, except that *pchG* is replaced by an unrelated dehydrogenase gene (Fig. 8). Pyochelin **11** production has also been reported from clinical isolates belonging

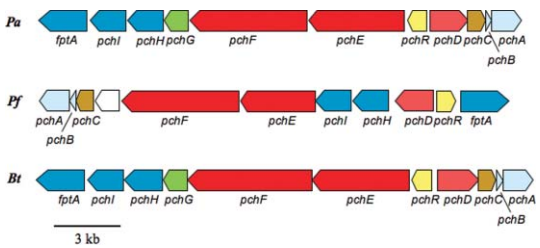
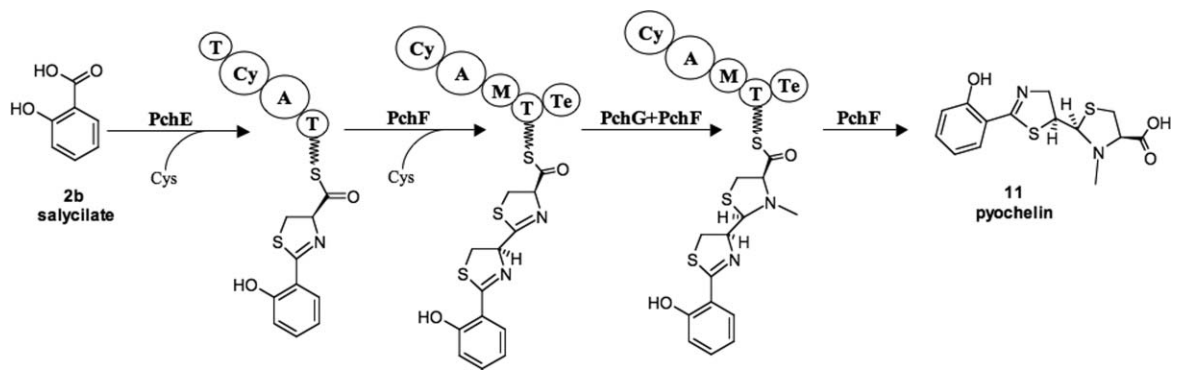


Fig. 8 The pyochelin clusters from: *Pa*, *P. aeruginosa* PA01; *Pf*, *P. fluorescens* Pf-5; and *Bt*, *B. thailandensis* E264 (*B. pseudomallei* K96243 and *Burkholderia* sp. 383 possess substantially identical clusters). Orthologs are indicated using the *P. aeruginosa* nomenclature. Colour codes are: red, NRPS; light blue, salycilate biosynthesis; yellow, regulatory protein; brown, thioesterase; dark blue, Fe-siderophore transport; green, NADPH-dependent oxidoreductase; white, unknown.



Scheme 5

to the *Burkholderia cepacia* complex.[34] Orthologous clusters with the same gene order and composition are also present in *B. pseudomallei*,[35] *B. thailandensis*,[36] as well as in the unpublished genome of *Burkholderia* sp. 383[11] (Fig. 8).

5.1.3 Yersiniabactin

The siderophore yersiniabactin **12** is one of the iron transport systems of *Yersinia*, and acts as a virulence factor for pathogenic strains. Its formation[16] (Scheme 6), which involves a hybrid PKS–NRPS system, resembles the pyochelin **11** pathway (Scheme 5). Salicylate **2b**, formed from chorismate **1** by the action of YbtS, is activated by YbtE and tethered to the first T domain of HMWP2 (encoded by *irp1*). Two cysteines are then incorporated and heterocyclized by HMWP2 to yield the tripeptidyl intermediate hydroxyphenyl-thiazolinyl-thiazolinyl-*S*-enzyme. The growing chain is then transferred to the PKS HMWP1 (encoded by *irp2*), which elongates the chain by a two-carbon unit through the incorporation of one malonyl extender unit. One thiazoline is reduced, likely by YbtU, followed by two *C*-methylations and β-ketoreduction by the M and KR domains of HMWP1, respectively. The growing chain is then translocated to the C-terminal NRPS module within HMWP1, where one additional thiazoline is added and methylated. Final hydrolytic release by the TE domain yields yersiniabactin **12**.

Yersiniabactin **12** is also found in other γ-Proteobacteria, as depicted in Fig. 9. Each cluster contains orthologs of *ybtE* (salicylate adenylation), of *irp1* and *irp2*, of *ybtT* (encoding a TE) and of *ybtU* (reductase). Other common genes code for the AraC-type regulator YbtA (which is duplicated in *P. profundum*), for a siderophore receptor (duplicated in *Pseudomonas* spp.), and for the two ABC transporters needed for iron uptake. Salicylate formation requires a single protein (YbtS) in *Yersinia* spp. and in *P. profundum*, while two proteins (PchAB) are encoded by the *Pseudomonas* clusters.

5.1.4 The *rkpA* locus

Several Proteobacteria contain a TMS gene, designated *rkpA* or *wcbR* in different strains, often linked to genes involved in capsule production. This gene encodes a protein with KS-AT-KR-ACP

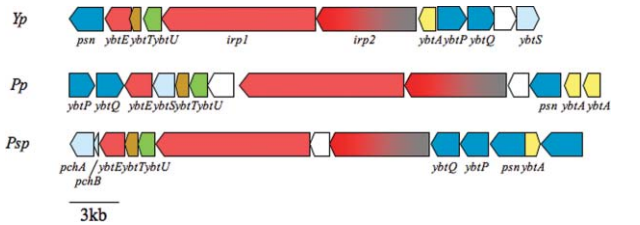


Fig. 9 The yersiniabactin clusters from: *Yp*, *Y. pestis* KIM5 (substantially identical cluster in strain CO-92); *Pp*, *P. profundum*; and *Psp*, *P. syringae* 1448A (substantially identical cluster in *P. syringae* DC3000). Colour codes are: red, NRPS; grey, PKS; dark blue, siderophore receptors and transporters; yellow, regulatory proteins; light blue, salicylate biosynthesis; brown, thioesterase; green, reductase; white, unknown. Orthologs are indicated according to the *Y. pestis* nomenclature.

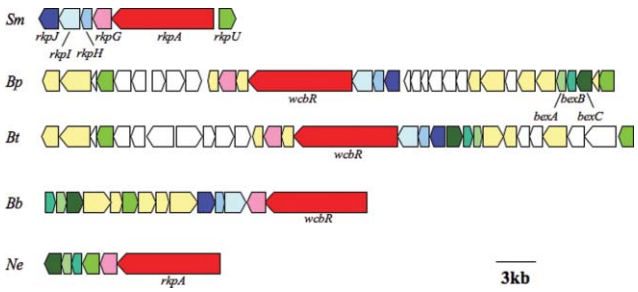
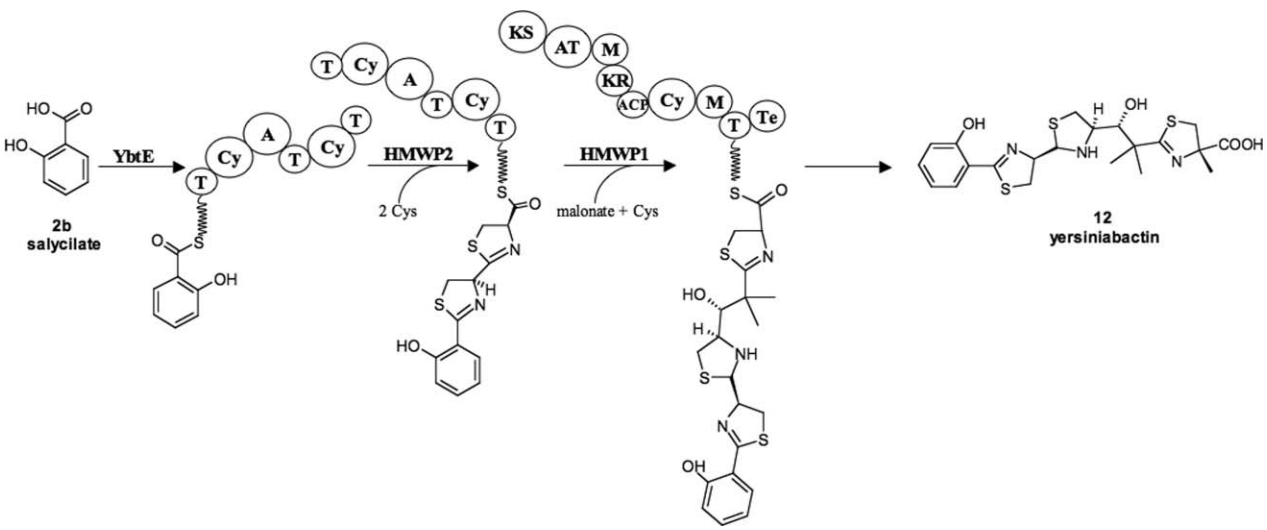


Fig. 10 The *rkpA* loci from *S. meliloti* (*Sm*), *Burkholderia pseudomallei* K96243 (*Bp*, highly related to that of *B. mallei* ATCC 23344), *B. thailandensis* E264 (*Bt*, highly related to that of *B.* sp. 383), *Bordetella bronchiseptica* RB50 (*Bb*, substantially identical to that of *B. parapertussis* 12822) and *N. europaea* (*Ne*). Color codes are: red, TMS; pink, aminotransferase; green, capsule polysaccharide export; blue, capsule polysaccharide synthesis; yellow, capsule synthesis and transport without orthologs in the other reported clusters; white, unknown or unrelated functions.

domains. In *Sinorhizobium meliloti* (*Sm* in Fig. 10), *rkpA* has been postulated to be involved in the synthesis of a lipid anchor (or a lipid carrier) for an extracellular polysaccharide.[37] Next to *rkpA* are genes encoding an aminotransferase (*rkpG*), three proteins involved in capsular polysaccharide biosynthesis (*rkpHIJ*) and a



Scheme 6

capsule polysaccharide exporter (*rkpU*). Orthologous clusters are present in other genomes (Fig. 10). In several *Burkholderia* species, the *rkpA* orthologs (*wcbR*) are linked to orthologs of *rkpGHI* and of *rkpU*, as well as to other genes (*bexABC*) involved in export of a capsular polysaccharide. An *rkpA* ortholog is also found in *Nitrosomonas europaea*,[38] but the cluster (Fig. 10, *Ne*) does not include other *rkp* orthologs. Instead, it harbors orthologs of the *bexABC* genes. *rkpA* orthologs are also found in *Bordetella bronchiseptica*[39] and *Bordetella parapertussis*,[39] linked to *rkpHIJ* and *bexABC* orthologs (Fig. 10, *Bb*). The similar organization of this locus in different strains suggests an equivalent role for this PKS, even if the protein similarity between members of different lineages is just above 50%.

5.2 Class α-Proteobacteria

Twenty five genomes have been analyzed, belonging to 23 distinct species (Table 2). TMS genes are found in only 10 strains, however, 14 genomes are 2 Mb or smaller, thus less likely to harbor secondary metabolism genes. With the exception of *Agrobacterium tumefaciens*, the density of these genes is modest (Table 2).

5.2.1 Order Rhizobiales

(a) Agrobacterium tumefaciens. *A. tumefaciens* is the etiological agent of the plant disease crown gall, but the main interest in this species is due to its capability to transfer DNA to plant cells, which makes it an ideal tool for the generation of transgenic plants. The 5.7 Mb genome of strain C58 has been independently sequenced by two groups.[40] Two TMS loci of unknown function are located at two loci on the 2.1 Mb linear chromosome (none are found on the circular chromosome or on the two plasmids). One cluster consists of two divergently transcribed portions: one encodes three NRPSs for a total of five modules and a monomodular PKS, the other two monomodular NRPSs and a monomodular PKS. It also encodes a receptor involved in iron transport and an ABC transporter. The other locus encodes a monomodular NRPS.

(b) Sinorhizobium meliloti. The 6.7 Mb genome of *S. meliloti*,[60] the nitrogen fixing symbiont of alfalfa, does not contain TMS genes other than *rkpA* (Fig. 10).

(c) Bradyrhizobium japonicum. Two NRPS genes of unknown function are found in the 9.1 Mb genome of the nitrogen-fixing soybean symbiont *B. japonicum*.[43] The largest encodes a three-module enzyme, while the other codes for a bimodular NRPS, and is located adjacent to the gene for a PPTase.

(d) Mesorhizobium loti. In the nitrogen-fixing symbiotic bacterium *M. loti* TMS genes are distributed on the three replicons constituting its 7.6 Mb genome.[52] Two NRPS genes are adjacent on the chromosome, next to genes involved in lipopolysaccharide synthesis. One plasmid contains an NRPS gene, while the other plasmid encodes a cluster comprising an NRPS loading module and a PKS which terminates in a TE.

(e) Rhodopseudomonas palustris. *R. palustris* is a metabolically versatile species, capable of growth using light, inorganic, or organic compounds as energy sources and carbon dioxide or organic compounds as carbon sources. Its 5.5 Mb genome[54] contains a single TMS cluster of unknown function encoding a hybrid PKS–NRPS (one module each), an NRPS loading module and an isolated C domain. The cluster also encodes a PPTase. Expression of this cluster is proposed to be regulated by a quorum sensing mechanism.[54]

(f) Genus Brucella. *Brucella* spp. are facultative intracellular pathogens and each of the four completed genomes[44–47] contains a cluster for the synthesis of the cathechol siderophore brucebactin.[32,64] Experimental evidence indicates that brucebactin is essential for wild-type virulence in the natural host.[32] However,

Table 2 Analyzed genomes from the α-Proteobacteria

Strain	Order	Genome[a]	TMS[b]	Reference
Agrobacterium tumefaciens C58	Rhizobiales	5.7	39[c]	40
Anaplasma marginale StMaries	Rickettsiales	1.2	0	41
Bartonella henselae str. Houston-1	Rhizobiales	1.9	0	42
Bartonella quintana str. Toulouse	Rhizobiales	1.6	0	42
Bradyrhizobium japonicum USDA 110	Rhizobiales	9.1	16.5	43
Brucella abortus 9-941	Rhizobiales	3.3	3	44
Brucella melitensis 16M	Rhizobiales	3.3	3	45
Brucella melitensis bv. *abortus* 2308	Rhizobiales	3.3	3	46
Brucella suis 1330	Rhizobiales	3.3	3	47
Caulobacter crescentus CB15	Caulobacterales	4	0	48
Ehrlichia ruminantum Gardel	Rickettsiales	1.5	0	49
Ehrlichia ruminantum Welgevonden	Rickettsiales	1.5	0	50
Gluconobacter oxydans 621H	Rhodospirillales	2.9	0	51
Mesorhizobium loti MAFF 303099	Rhizobiales	7.6	19.7	52
"*Pelagibacter ubique*" HTCC1062	Rickettsiales	1.3	0	53
Rhodopseudomonas palustris CGA009	Rhizobiales	5.5	12	54
Rickettsia conorii Malish 7	Rickettsiales	1.3	0	55
Rickettsia felis URRWXCal2	Rickettsiales	1.6	0	56
Rickettsia prowazekii Madrid E	Rickettsiales	1.1	0	57
Rickettsia typhi Wilmington	Rickettsiales	1.1	0	58
Silicibacter pomeroyi DSS-3	Rhodobacterales	4.6	11	59
Sinorhizobium meliloti 1021	Rhizobiales	6.7	7.5	60
Wolbachia pipientis wMel	Rickettsiales	1.3	0	61
Wolbachia endosymbiont strain TRS	Rickettsiales	1.1	0	62
Zymomonas mobilis ZM14	Sphingomonadales	2	0	63

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb. [c] Average of the different gene sizes from ref. 40.

the PPTase gene is inactivated by a frameshift mutation in the sequenced field isolate strain 9-941.

5.2.2 Order Rhodobacterales

(a) Silicibacter pomeroyi. The genome of the marine litho-heterotrophic bacterium *S. pomeroyi* DSS3[59] contains a TMS cluster of unknown function encoding a monomodular NRPS (including a formyl transferase associated to A and T domains), a monomodular PKS (terminating with a TE domain) and a PPTase. It should be noted that a very similar cluster is also present in the draft genomes of two related species, *Silicibacter* sp TM1040 and *Paracoccus denitrificans* PD1222.[11]

5.2.3 Other orders

Members of the order Rickettsiales (10 strains with genomes ranging from 1.1 to 1.6 Mb) are mainly obligate intracellular symbionts or pathogens, with the notable exception of Candidatus *Pelagibacter ubique*, a highly abundant marine bacterium with the smallest genome among free-living bacteria identified so far. They harbor no TMS genes, nor do *Caulobacter crescentus* and *Zymomonas mobilis*, the only analyzed representatives of the orders *Caulobacterales* and *Sphingomonadales*, respectively.

5.3 Class β-Proteobacteria

Eight of the 13 analyzed genomes, representing 12 species, contain TMS genes (Table 3). The density of TMS genes is highly variable, and large differences can be found even within a single genus, such as *Burkholderia*.

5.3.1 Order Burkholderiales

(a) Genus Burkholderia. The genomes of three *Burkholderia* species have been analyzed. In addition, two unpublished genomes are available from *Burkholderia* sp. 383 (*Bc*) and *B. xenovorans*.[11] *B. pseudomallei* (*Bp*), normally found in terrestrial environments, is an opportunistic pathogen that causes melioidosis, a serious health hazard. *B. mallei* (*Bm*), the causative agent of glanders, is thought to have evolved from *Bp* through a process of genome size reduction to become an obligate parasite with a narrow host range. Albeit highly related to *Bp* and *Bm*, *B. thailandesis* (*Bt*) is non-pathogenic. *Bm*,[66] *Bp*[35] and *Bt*[36] contain a large number of TMS genes (ranging from 1.6 to 3.5% of the genome). *Bm* also contains a relevant number of TMS pseudogenes, which probably resulted from the reduced selective pressure to produce secondary metabolites due to its narrow host range. The TMS clusters present in the three genomes are summarized in Table 4. *Bc*, *Bm*, *Bp* and *Bt* encode a homolog of RkpA (Fig. 10), while *pch* clusters for pyochelin **11** formation are found in *Bc*, *Bp* and *Bt* (Fig. 8).

All five *Burkholderia* strains share one cluster, encoding two NRPSs supposedly involved in the synthesis of a hydroxamate siderophore. The cluster also encodes other siderophore biosynthesis enzymes and a siderophore receptor. The larger NRPS has a frameshift and may be nonfunctional in *Bm*. This cluster represents the only occurrence of TMS genes in the unpublished genome of *B. xenovorans*.[11]

Three clusters of unknown function (A, B and C in Table 4) are shared by *Bm*, *Bp* and *Bt*: the first encodes an NRPS loading module and a bimodular NRPS (inactivated by a frameshift in *Bm*); the second a bimodular PKS and a bimodular hybrid NRPS–PKS; and the third an NRPS loading module and a monomodular PKS, as well as a PPTase.

Four TMS clusters of unknown function present in *Bp* are partially deleted or inactivated in *Bm*. In three clusters (D, E and F in Table 4) some of the *Bp* orthologs are absent or show extensive

Table 4 Orthologous TMS clusters in *Burkholderia*[a]

Cluster/compound[b]	*Bp*	*Bm*	*Bt*
Hydroxamate siderophore	Yes	Pseudo	Yes
Pyochelin	Yes	—	Yes
rkpA locus	Yes	Yes	Yes
A	Yes	Pseudo	Yes
B	Yes	Yes	Yes
C	Yes	Yes	Yes
D	Yes	Pseudo	—
E	Yes	Pseudo	—
F	Yes	Pseudo	—
G	Yes	Pseudo	—
H	Yes	—	Yes
I	Yes	—	Yes
J	Yes	—	Yes
K	Yes	—	Yes
L	—	—	Yes
M	—	—	Yes

[a] 'Yes' indicates that an ortholog is present, 'Pseudo' that one or more genes are inactive or deleted. [b] See text for cluster abbreviations.

Table 3 Analyzed genomes from the β-Proteobacteria

Strain	Order	Genome[a]	TMS[b]	Reference
Azoarcus sp. EbN1	Rhodocyclales	4.7	0	65
Bordetella bronchiseptica RB50	Burkholderiales	5.3	7.6	39
Bordetella parapertussis 12822	Burkholderiales	4.8	7.6	39
Bordetella pertussis Tohama I	Burkholderiales	4.0	0	39
Burkholderia mallei ATCC 23344	Burkholderiales	5.8	96.1[c]	66
Burkholderia pseudomallei K96243	Burkholderiales	7.2	255.2	34
Burkholderia thailandensis E264	Burkholderiales	6.7	226.5	35
Chromobacterium violaceum ATCC 12472	Neisseriales	4.8	33.8	29
Neisseria meningitidis MC58	Neisseriales	2.2	0	67
Neisseria meningitidis Z2491	Neisseriales	2.3	0	68
Nitrosomonas europaea ATCC 19718	Nitrosomonadales	2.8	7.6	38
Ralstonia solanacearum GMI1000	Burkholderiales	5.8	71.5	69
Thiobacillus denitrificans ATCC 25259	Hydrogenophilales	2.9	0	70

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb. [c] 147.3 kb also considering pseudogenes.

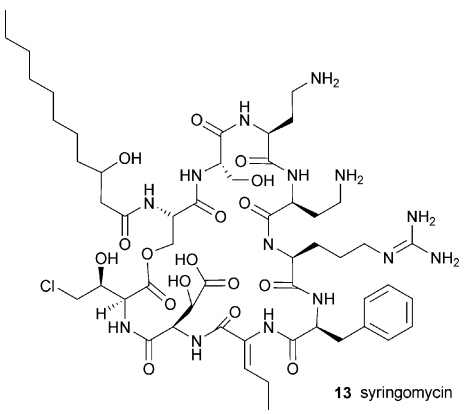
deletions in *Bm*. The fourth cluster (G, Table 4) contains four hybrid NRPS–PKS genes in *Bp* and is inactivated by several point mutations in *Bm*.

A large TMS cluster is shared by *Bp* and *Bt* (H, Table 4). It codes for a bimodular NRPS–PKS and two monomodular NRPSs, a stand-alone A domain (transcribed in the opposite direction), three monomodular PKSs and two monomodular NRPSs. The two genomes also share three other loci: two encode a monomodular NRPS each (I and J, Table 4), while the third (K) contains a stand-alone A domain possibly involved in the synthesis of a nonproteinogenic amino acid.

Finally, two large TMS clusters of unknown function (L and M, Table 4) are present exclusively in *Bt*: the first specifies five NRPS and three PKS modules, the second one NRPS and 13 PKS modules.

(b) Genus Bordetella. The genus *Bordetella* includes pathogens causing severe diseases of the respiratory tract. An *rkpA* orthologous cluster is present in *B. bronchiseptica* and *B. parapertussis* (Fig. 10), but not in the related *B. pertussis* Tohama I strain, which is reported to have several deletions in the capsule locus.[39]

(c) Genus Ralstonia. Over 1.2% of the 5.8 Mb genome of the plant pathogen *Ralstonia solanacearum*[69] encode TMS enzymes, while no TMS genes are present in the unpublished genome sequence of *R. eutropha* (*Cupriavidus necator*) JMP134.[11] A TMS cluster on the *R. solanacearum* 2.1 Mb megaplasmid encodes a hybrid enzyme with one PKS and five NRPS modules and a five-module NRPS. These genes have been implicated in pathogenesis[69] due to their similarity to the syringomycin **13** synthetase and to other NRPSs from several plant pathogens. An additional cluster, also on the megaplasmid, encodes two monomodular NRPSs. The megaplasmid contains another region potentially coding for a PKS, but presenting a frameshift, so likely representing a mutated gene that may be active in different strains. The 3.7 Mb chromosome also contains a TMS cluster encoding a hybrid protein including two NRPS and one PKS module, a polypeptide consisting of TE and PPTase domains, a monomodular PKS and a monomodular NRPS. These genes are annotated as involved in siderophore synthesis because of their association with genes encoding a siderophore receptor and an ABC transporter.



13 syringomycin

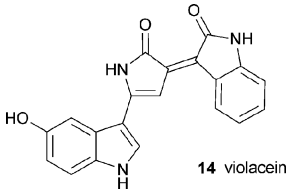
5.3.2 Order Nitrosomonadales, genus *Nitrosomonas*

N. europaea is a chemolithoautotroph nitrifying bacterium with an obligate requirement for ammonia oxidation and CO_2 fixation for growth. It contains a single TMS gene encoding an RkpA ortholog[38] (Fig. 10).

5.3.3 Order Neisseriales

TMS genes are not present in the pathogens *Neisseria meningitidis* and *N. gonorrhoeae*, but only in *C. violaceum*.

(a) Chromobacterium violaceum. This bacterium, which lives in soil and water, is highly abundant in the Brazilian Amazon and it produces the bactericidal purple pigment violacein **14**. It can occasionally be pathogenic, mainly in immunocompromized individuals or children. Its genome[29] harbors orthologous clusters for enterobactin- and chrysobactin-related siderophores (Fig. 7). Two other clusters of unknown function are present: one encodes two NRPSs for a total of 6 modules, while the other a small PKS annotated as involved in antibiotic synthesis,[29] although the flanking genes do not suggest such a role.



14 violacein

5.3.4 Other orders

No TMS genes have been identified in *Thiobacillus denitrificans* and in *Azoarcus* sp. EbN1, the only representatives of the orders *Hydrogenophilales* and *Rhodocyclales*, respectively.

5.4 Class γ-Proteobacteria

This represents the largest bacterial (sub)division in terms of cumulative genome size (247 Mb) and total length of TMS genes (1.62 Mb), with published genomes available from 26 distinct genera (Table 5). TMS genes, which account for 0.66% of the overall genome lengths, are found in 40 out of 59 strains.

5.4.1 Order Pseudomonadales

(a) Family Moraxellaceae, genus Acinetobacter. *Acinetobacter* spp. are non-motile, strictly aerobic, nutritionally versatile chemoheterotrophs that parallel *Pseudomonas* spp. in the range of substrates used as sole carbon and energy sources. They are also attracting notoriety as opportunistic pathogens in nosocomial infections. The sequenced strain ADP1[71] was derived from a soil isolate. Within the single 3.6 Mb chromosome, there are two NRPS loci. One cluster consists of three distinct genes, for a total of three modules, flanked by homologs of genes involved in photobactin **10** biosynthesis in *Photorabdus luminescens*,[119] suggesting that strain ADP1 may synthesize a similar siderophore. The other locus contains a single gene encoding an A domain as part of a larger polypeptide.

(b) Family Pseudomonadaceae, genus Pseudomonas. *Pseudomonas* spp. are ubiquitous inhabitants of soil and water. They often live in a commensal relationship on plant surfaces, where they can exert a profound effect on the eukaryotic host. *P. aeruginosa* is also of major medical concern as an opportunistic pathogen. Six genomes were analyzed: one from *P. aeuriginosa* (Pa)[14]; one from *P. fluorescens* (Pf)[33]; one from *P. putida* (Pp)[92]; and three from *P. syringae*,[93–95] namely *P. syringae* pv. *syringae* strain B728a (Pss), *P. syringae* pv. tomato strain DC3000 (Pst) and *P. syringae* pv. *phaseolicola* strain 1448A (Psp). *Pseudomonas* spp. are known to produce a wide variety of secondary metabolites,

Table 5 Analyzed genomes from γ-Proteobacteria

Strain	Order	Genome[a]	TMS[b]	Reference
Acinetobacter sp. ADP1	Pseudomonadales	3.6	12.9	71
Blochmannia floridanus	Enterobacteriales	0.7	0	72
Blochmannia pennsylvanicus BPEN	Enterobacteriales	0.8	0	73
Buchnera aphidicola APS	Enterobacteriales	0.7	0	74
Buchnera aphidicola Bp	Enterobacteriales	0.6	0	75
Buchnera aphidicola Sg	Enterobacteriales	0.7	0	76
Colwellia psychrerythraea 34H	Alteromonadales	5.4	17	25
Coxiella burnetii RSA 493	Legionellales	2.0	6.2	77
Erwinia carotovora subsp. *atroseptica* SCRI104	Enterobacteriales	5.1	73.2	31
Escherichia coli K12	Enterobacteriales	4.7	5.5	78
Escherichia coli O157:H7 EDL933	Enterobacteriales	5.6	5.5	79
Escherichia coli O6 CFT073	Enterobacteriales	5.2	58.9	80
Escherichia coli Sakai O157:H7	Enterobacteriales	5.6	5.5	81
Francisella tularensis SCHU S4	Thiotrichales	1.9	0	82
Haemophilus influenzae 86-028NP	Pasteurellales	1.9	0	83
Haemophilus influenzae Rd KW20	Pasteurellales	1.8	0	84
Hahella chejuensis KCTC 2396	Oceanospirillales	7.2	144	22
Idiomarina loihiensis L2TR	Alteromonadales	2.8	0	85
Legionella pneumophila Lens	Legionellales	3.3	18.3	86
Legionella pneumophila Paris	Legionellales	3.5	15.7	86
Legionella pneumophila Philadelphia 1	Legionellales	3.4	17.2	87
Mannheimia succiniciproducens MBEL55E	Pasteurellales	2.3	0	88
Methylococcus capsulatus Bath	Methylococcales	3.3	17	89
Pasteurella multocida subsp. *multocida* Pm70	Pasteurellales	2.3	0	90
Photobacterium profundum SS9	Vibrionales	6.4	45.9	26
Photorhabdus luminescens laumondii TTO1	Enterobacteriales	5.7	207.9	91
Pseudomonas aeruginosa PAO1	Pseudomonadales	6.3	73.8	14
Pseudomonas fluorescens Pf-5	Pseudomonadales	7.1	182.8	33
Pseudomonas putida KT2440	Pseudomonadales	6.2	42.1	92
Pseudomonas syringae B728a	Pseudomonadales	6.1	263.5	93
Pseudomonas syringae phaseolicola 1448A	Pseudomonadales	6.1	89.1	94
Pseudomonas syringae pv. *tomato* DC3000	Pseudomonadales	6.5	136.6	93, 95
Salmonella enterica choleraesuis SC-B67	Enterobacteriales	4.9	5.5	96
Salmonella enterica Paratyphi-A SARB42	Enterobacteriales	4.6	14.7	97
Salmonella enterica Typhi CT18	Enterobacteriales	5.1	5.5	98, 99
Salmonella enterica Typhi Ty2	Enterobacteriales	4.8	5.5	99
Salmonella typhimurium LT2	Enterobacteriales	5.0	5.5	100
Shewanella oneidensis MR-1	Alteromonadales	5.1	0	101
Shigella boydii sv4 Sb227	Enterobacteriales	4.6	5.5	102
Shigella dysenteriae sv1 Sd197	Enterobacteriales	4.6	5.6	102
Shigella flexneri 2a 2457T	Enterobacteriales	4.6	5.5	103
Shigella flexneri 2a 301	Enterobacteriales	4.8	5.5	104
Shigella sonnei Ss046	Enterobacteriales	5.0	5.5	102
Vibrio cholerae O1 biovar eltor N16961	Vibrionales	4.0	2.9	105
Vibrio fischeri ES114	Vibrionales	4.2	0	106
Vibrio parahaemolyticus RIMD 2210633	Vibrionales	5.2	0	107
Vibrio vulnificus YJ016	Vibrionales	5.3	7.6	108
Wigglesworthia glossinidia endosymbiont	Enterobacteriales	0.7	0	109
Xanthomonas axonopodis citri 306	Xanthomonadales	5.3	14.7	110
Xanthomonas campestris campestris 8004	Xanthomonadales	5.1	4	111
Xanthomonas campestris campestris ATCC 33913	Xanthomonadales	5.1	4	110
Xanthomonas campestris pv. *vesicatoria* 85-10	Xanthomonadales	5.2	0	112
Xanthomonas oryzae KACC10331	Xanthomonadales	4.9	0	113
Xylella fastidiosa 9a5c	Xanthomonadales	2.7	0	114
Xylella fastidiosa Temecula1	Xanthomonadales	2.5	0	115
Yersinia pestis biovar Medievalis 91001	Enterobacteriales	4.8	17.4	116
Yersinia pestis CO92	Enterobacteriales	4.8	20.2	116
Yersinia pestis KIM	Enterobacteriales	4.6	29.3	116, 117
Yersinia pseudotuberculosis IP 32953	Enterobacteriales	4.8	17.3	118

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb.

including pyoverdines **15**, a diverse class of siderophores containing a chromophore linked to a nonribosomally synthesized peptide varying in length and composition.[120] All strains contain a similar gene set for the synthesis of pyoverdine **15** as illustrated in Fig. 9. However, in some strains the pyoverdine genes are present as two or three separate clusters (Fig. 11). Despite some deviations in gene placement and orientation, the general organization of pyoverdine biosynthesis and uptake genes is similar to that found in other pseudomonads.[121] In addition, a TMS cluster involved in pyochelin **11** formation is found in *Pa* and *Pf* (Fig. 8). The TMS genes found in these six *Pseudomonas* are summarized in Table 6.

Table 6 Orthologous TMS clusters from *Pseudomonas* spp.[a]

Cluster/compound[b]	*Pa*	*Pf*	*Pp*	*Pss*	*Psp*	*Pst*
Coronafacic acid	—	—	—	—	—	Yes
Pyochelin	Yes	Yes	—	—	—	—
Pyoluteorin	—	Yes	—	—	—	—
Pyoverdine	Yes	Yes	Yes	Yes	Yes	Yes
Syringolin	—	—	—	Yes	—	—
Syringomycin	—	—	—	Yes	—	—
Syringopeptin	—	—	—	Yes	—	—
Yersiniabactin	—	—	—	—	Yes	Yes
A (polyketide)	—	Yes	—	—	—	—
B (cyclic lipopeptide)	—	Yes	—	—	—	—
C (2 Mod NRPS)	—	Yes	—	—	—	—
D (6 Mod NRPS + 1 PKS)	—	—	—	Yes	Yes	—
E (8 Mod NRPS)	—	—	—	Yes	—	Yes
F (6 Mod NRPS)	—	—	—	Yes	—	Yes[c]
G (A-T-Red)	—	—	—	Yes	Yes	Yes
H (A-T-AT + PPTase)	—	—	—	Yes	Yes	—
I (5 Mod NRPS)	—	—	—	—	—	Yes
J (polyketide)	—	—	—	Yes	—	—
K (A-T)	—	—	—	—	Yes	—
L (A + KS)	—	—	—	—	Yes	—
M (3 Mod NRPS)	Yes	—	—	—	—	—
N (2 Mod NRPS)	Yes	—	—	—	—	—
O (1 Mod NRPS)	Yes	—	—	—	—	—

[a] 'Yes' indicates that an ortholog is present. [b] See text for details. [c] Only 4 modules in Pst.

• *Pseudomonas fluorescens*. *P. fluorescens* Pf-5 had been reported to produce the polyketide pyoluteorin **16** and 2,4-diacetylphloroglucinol **17**.[122,123] Analyzing its 7.1 Mb genome, Paulsen *et al.* [33] performed a detailed analysis of the genetic potential for secondary metabolism in *Pf*, identifying nine distinct clusters, for approximately 400 kb, or 5.7% of the *Pf* genome. In addition to the clusters for pyochelin **11** (Fig. 8) and pyoverdine **15** (Fig. 11) formation, *Pf* contains four additional TMS loci. The pyoluteorin **16** cluster from the *Pf* genome is similar to that previously reported.[123] Another cluster (A in Table 6) is likely to catalyze the formation of a novel polyketide: it encodes four predicted PKSs, a hybrid NRPS–PKS, a cytochrome P450 enzyme and a methyltransferase. The product synthesized by these enzymes is predicted to start with a serine residue followed by 14 unknown carboxylic acids. This cluster is unique to *Pf* (Table 6).

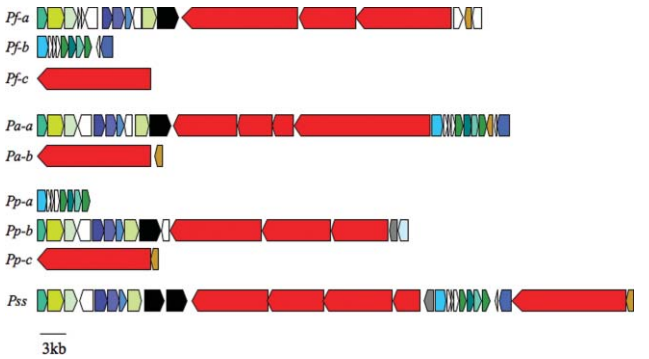
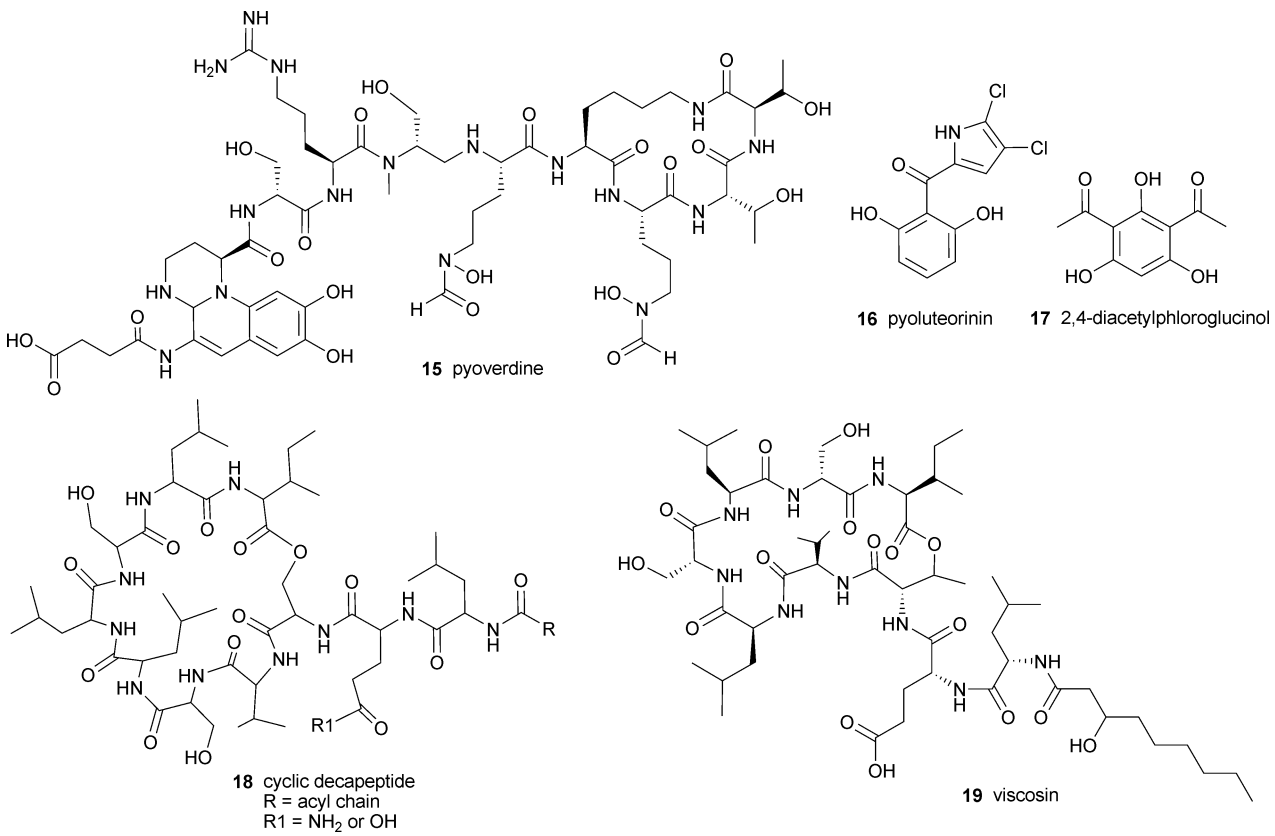


Fig. 11 The pyoverdine clusters from *P. fluorescens* Pf-5 (*Pf*), *P. aeruginosa* PA01 (*Pa*), *P. putida* (*Pp*) and *P. syringae* B728a (*Pss*, substantially identical clusters in strains DC3000 and 1448). Letters indicate non-contiguous regions. Colour codes are: red, NRPS; green, transport; gray, regulatory; black, siderophore receptors; brown, thioesterase; blue, other conserved functions; white, unknown. Orthologs have identical colours, except for white.

Despite the lack of sequence data on clusters synthesizing similar compounds, an *in silico* analysis of cluster B (Table 6)



suggested a hypothetical structure for the predicted cyclic decapeptide product **18**, which resembles cyclic lipopeptides of the viscosin **19** group. The first NRPS lacks the characteristic loading module, however, a gene encoding a protein with an AMP-binding domain may supply this function in *trans*. Two TE domains are present at the C-terminus of the last NRPS, and one of them could be responsible for decapeptide cyclization, while the other would catalyze lipid side chain addition, in analogy to SyrC during syringomycin **13** biosynthesis. The authors could detect surfactant activity in *Pf* and observed expression of selected genes from cluster B by RT-PCR under conditions in which the surfactant activity was produced. The structure predicted from *in silico* studies has substantially been confirmed by the actual purification of the compound, dubbed orfamide.[123b] A highly similar cluster, bearing an additional elongation module, is present in the unpublished genome sequence of *P. fluorescens* strain PfO-1.[11] Finally, cluster C (Table 6) encodes two small NRPSs carrying A-T and A-C domains. No counterparts are found in the other *Pseudomonas* strains.

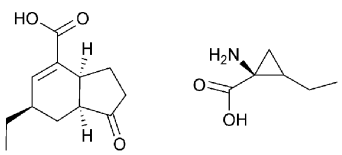
• *Pseudomonas syringae*. *P. syringae* is a widespread pathogen of many plant species, subdivided into pathovars based on pathogenicity and host range. *P. syringae* pv. *syringae* (*Pss*) can achieve and maintain large populations epiphytically on healthy plants, which serve as inocula for subsequent plant invasion and disease. *P. syringae* pv. *tomato* (*Pst*), instead, poorly colonizes the exterior of plants, and multiplies endophytically within the plant. *P. syringae* pv. *phaseolicola* (*Psp*) is the causative agent of halo blight disease in the common bean (*Phaseolus vulgaris*). Diseased bean leaves present a yellow halo produced by the release of phaseolotoxin **20**. *Psp* is virulent on all *P. vulgaris* varieties examined and is closely related to *Pst*. The three genomes are similarly sized (6.1–6.5 Mb, Table 5).

Some natural products had been previously identified from the sequenced strains. *Pss* is known to produce two classes of lipodepsipeptides: the syringopeptins **21** and the lipodepsinonapeptides. Usually each strain secretes a single type of syringopeptin **21** and one or two lipodepsinonapeptides.[124] *Pss* is known to synthesize two syringopeptins **21** and syringomycin **13**.[125] *Pss* strains are also capable of producing a family of peptide derivatives called syringolins **22**. *Pst* is known to produce coronatine **23**.

The *Psp*, *Pss* and *Pst* genomes each present a cluster for pyoverdin **15** biosynthesis (Fig. 11). In addition, *Psp* and *Pst* contain a cluster resembling the yersiniabactin **12** cluster (Fig. 9). Yersiniabactin **12** production was detected by HPLC in all *P. syringae* strains that were PCR-positive to the *ybt* genes.[126] In *Psp* and *Pst*, the *ybt* locus lies between two genes that are adjacent in *Pss*, a strain lacking the *ybt* cluster.[126]

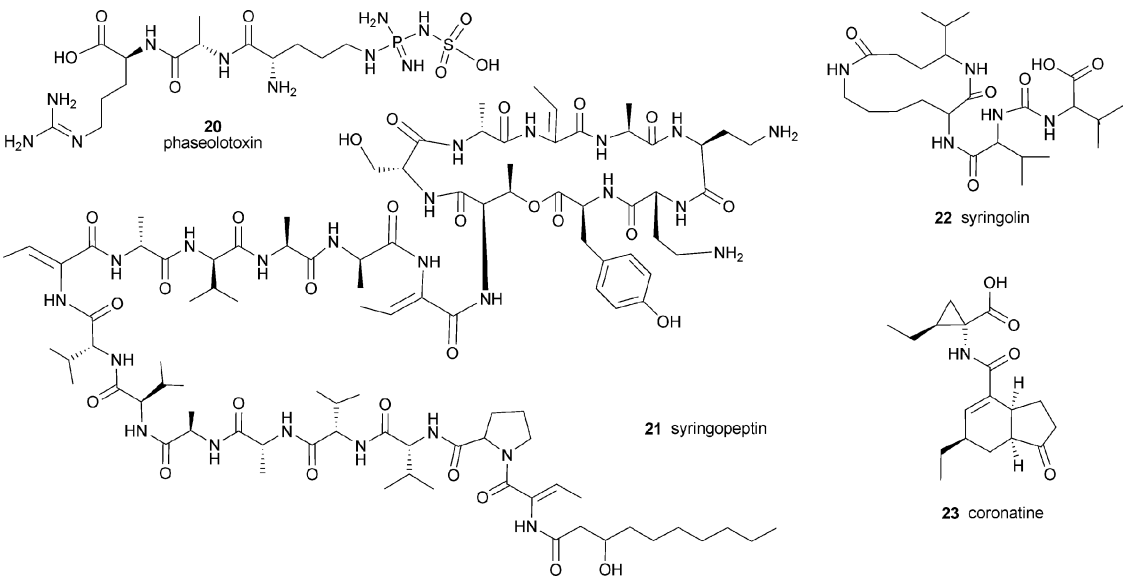
The syringomycin **13** and syringopeptin **21** clusters lie in close proximity and are divergently transcribed in the *Pss* chromosome, where they are part of a larger cluster that includes an ABC transporter for export of both metabolites as well as genes for siderophore biosynthesis and uptake, streptomycin resistance, arginine degradation, and genes for lipid A modification.

The *cfa* genes are responsible for the synthesis of the polyketide coronafacic acid **24**, which is joined through an amide bond with coronamic acid **25** (made by the *cma* genes) to form the phytotoxin coronatine **23**. In many *P. syringae* pathovars the coronatine genes are clustered and plasmid-encoded. However, the *cfa* and *cma* genes are separated by 26 kb in the *Pst* chromosome.



24 coronafacic acid **25** coronamic acid

P. syringae strains produce many phytotoxins. However, clusters for the production of known phytotoxins (in addition to coronatine **23**) were not found in *Psp* and *Pst*, but TMS clusters of unknown function are present in two or more of the three *P. syringae* genomes (Table 6). *Pss* and *Pst* possess one cluster (D, three TMS genes) encoding a six-module NRPS, a PKS module (lacking an AT domain) and a dioxygenase (an orthologous cluster is also present in the unpublished genome sequence of *P. fluorescens* Pf0-1[11]). *Pss* and *Pst* also share two additional TMS clusters: one (E) encodes two NRPSs for a total of 8 modules,



20 phaseolotoxin

21 syringopeptin

22 syringolin

23 coronatine

the other (F) a single NRPS with 84% identity between the two strains, but consisting of 6 modules in *Pss* and 4 in *Pst* (in the latter strain, the NRPS gene is placed between the *cfa* and *cma* genes, in a region that also includes many virulence factors). Finally, one locus (G) encoding a polypeptide with A, T and reductase domains is present in a similar genetic context in the three *P. syringae* strains, while another locus (H) encoding a PPTase and a polypeptide with A, T and AT domains is shared by *Psp* and *Pss* (Table 6). TMS genes or clusters unique to one strain are also present (Table 6): *Pst* encodes two contiguous NRPSs for a total of 5 modules (I); an atypical PKS (for a total of 6 KS domains and a separate AT-containing polypeptide) is specified by the *Pss* chromosome (J); and *Psp* presents two distinct loci, one (K) encoding a polypeptide with A and T domains and the other (N) separate A- and KS-containing polypeptides (Table 6).

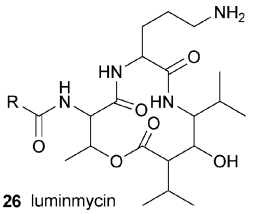
• *P. aeruginosa and P. putida*. When the genome sequence of *P. aeruginosa* PAO1 was reported,[14] it was not analyzed for the presence of TMS genes. In addition to the cluster for pyoverdine **15** and pyochelin **11**, this strain contains three regions encoding NRPSs: two polypeptides for a total of three modules; a bimodular enzyme; and a polypeptide containing A-T-TE domains (O–Q, Table 6). The *P. putida* KT2440 genome[93] is notable for its lack of genes for most known plant-related virulence traits. However, it contains two close clusters that may direct formation of pyoverdine **15** (Fig. 11).

5.4.2 Order Enterobacteriales, family Enterobacteriaceae

(a) Genus Erwinia. In addition to many human pathogens, the family Enterobacteriaceae includes several plant pathogens, among which *Erwinia carotovora* subsp. *atroseptica*—recently reclassified as *Pectobacterium atrosepticum*[127]—is a commercially relevant pathogen restricted to potato in temperate regions, in contrast to other *Erwinia* strains, which can infect a broader range of plants. Its 5.1 Mb chromosome[31] contains several putative horizontally acquired genomic islands. Two of them, HAI2 and HAI6, are relevant for TMS genes. HAI2 shows a high level of conservation of sequence and gene order to the SPI-7 pathogenicity island in *S. enterica* serovar Typhi. However, instead of containing the *viaB* operon (which encodes the exopolysaccharide pathogenicity determinant) as does *S. enterica*, the equivalent position in HAI2 is occupied by a segment highly similar to the *cfa* gene cluster in *P. syringae*. However, the *E. carotovora* genome does not contain the *cma* genes and polyketide phytotoxins have not previously been identified in enterobacterial plant pathogens. Nonetheless, knockouts of the *cfa6* and *cfa7* genes resulted in decreased virulence, suggesting that the *E. carotovora cfa* cluster produces a compound important for virulence, as coronatine **23** is in *P. syringae*.[94] The other island, HAI6, contains two genes of unknown function encoding 13 NRPS modules.

(b) Genera Escherichia, Salmonella and Shigella. Four distinct *E. coli* strains, and five strains each from *Salmonella* and *Shigella* were analyzed (Table 5). All strains possess a cluster (Fig. 7) likely to direct synthesis of enterobactin **3**. However, in both *Shigella flexneri* strains, *entC* and *fepE* are inactivated by point mutations, while *fepE* is also inactive in the *Salmonella enterica* Typhi strains Ty2 and CT18 (Fig. 7). The *ent* cluster accounts for all NRPSs in this group of strains, except for *E.coli* CFT073. This strain contains a region highly related to the yersiniabactin **12** cluster (Fig. 9), but the *irp1*, *irp2* and *irp5/ybtE* orthologs appear to be inactivated by IS elements. In contrast, other *E. coli* strains have been shown to contain intact *ybt* genes and to produce yersiniabactin **12**.[126]

(c) Genus Photorabdus. *P. luminescens* is symbiotic with soil nematodes and pathogenic to a wide range of insects. When a nematode carrying *P. luminescens* in its gut attacks a prey insect larva, the bacterium degrades the insect polymers while producing a wide range of antibiotics to ward off microbial competitors. Within the 5.7 Mb chromosome of strain TT01, 33 TMS genes were identified as part of 20 different loci.[91] In addition to the *ybt* cluster (Fig. 9), the NPbiogene site[128] reports a cluster predicted to direct the synthesis of luminmycin **26**. An additional cluster encodes one of the largest prokaryotic proteins, a 16367-aa NRPS, encompassing a loading module, 14 elongation modules and a TE domain. The authors observed that, as in strain W14,[129] ten TT01 genes are similar to genes for the biosynthesis of syringomycin **13**; however, these genes are not clustered in a single region. The sequenced genome does not contain a photobactin **10** cluster, detected in another strain.[119] It should be noted that many *P. luminescens* TMS genes share 60–72% identity with each other, so they may have originated through duplication events.



26 luminmycin

(d) Genus Yersinia. This genus includes some notorious pathogens. *Y. pestis*, the causative agent of plague, is primarily a rodent pathogen, usually transmitted to humans by the bite of an infected flea. *Y. pestis* has been proposed to be a clone evolved 1500–20 000 years ago from the closely related gastrointestinal pathogen *Y. pseudotuberculosis*.[130] Three *Y. pestis* isolates and one *Y. pseudotuberculosis* strain were analyzed (Table 5). The *Y. pestis* strains KIM and CO-92 harbor a virtually identical yersiniabactin **12** cluster (Fig. 9), which is however absent in *Y. pestis* bv. Medievalis and in *Y. pseudotuberculosis*. The latter strains contain instead a TMS cluster encoding A-T, KS-T, $(C\text{-}A\text{-}T)_3$ domains and a C-terminal TE. This cluster is located next to a region encoding a TonB-like receptor. This region is also present in the *Y. pestis* strains KIM and CO-92, but the TMS genes are interrupted by IS elements in strain KIM, while no TMS genes are left in strain CO-92, although IS sequences are present.

5.4.3 Order Alteromonadales, genus *Colwellia*

All characterized members of the genus *Colwellia* have been obtained from stably cold marine environments and are strictly psychrophilic. *C. psychrerythraea* 34H, which was isolated from Arctic marine sediments, grows in heterotrophic media over a temperature range of approximately −1 °C to 10 °C. Its 5.4 Mb chromosome[25] contains a single TMS locus with a PUFA-type organization (Fig. 6).

5.4.4 Order Legionellales

(a) Family Coxiellaceae, genus Coxiella. *Coxiella burnetii*, the etiological agent of Q fever, is highly infective to humans and livestock. *C. burnetii* is an obligate intracellular acidophile highly adapted to thrive within the phagolysosome of the

eukaryotic phagocyte. Its 2.0 Mb chromosome[77] encodes several predicted drug-efflux systems, which are suggested to provide resistance to host-produced antimicrobial defensins or to have had a role in secondary metabolite secretion. In fact, most PKS genes are likely pseudogenes, and the only apparently complete cluster encodes a single module PKS and an isolated A domain.

(b) Family Legionellaceae, genus Legionella. *Legionella pneumophila* is the causative agent of Legionnaires' disease. It replicates as an intracellular parasite of amoebae and can also persist in the environment as a free-living bacterium. The three *L. pneumophila* isolates (Table 5) encode a similarly-sized PKS, containing three KS domains and ending with a C domain. In each case, *ca.* 5 kb 5′ to the PKS gene is another TMS gene, which is identical in strains Paris and Philadelphia (it encodes a formyltransferase followed by A-T-TE domains) but not in strain Lens, where it encodes a hybrid polypeptide consisting of an NRPS loading module followed by a PKS module and a TE domain. The existence or function of these regions in the three strains has not been commented on by the authors.[86,87]

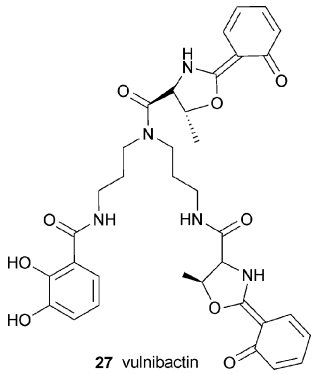
5.4.5 Order Methylococcales, genus *Methylococcus*

Methylococcus capsulatus is an obligate methanotroph that oxidizes methane to formaldehyde, which is then assimilated into cellular biomass or further oxidized to formate and CO_2 for energy production. The 3.3 Mb *M. capsulatus* chromosome[89] encodes a bi-modular NRPS that comprises a loading module containing an A domain likely to recognize 5-hydroxy ornithine (or another ornithine derivative), a T domain, and an unusual AT domain. The next module contains a C domain and a terminal TE. The authors hypothesize that this NRPS may synthesize a heavily charged peptide that could be involved in binding/scavenging of copper or other metals, since *Methylosinus*, another methylotroph, excretes copper-binding compounds.[131] The genome also encodes a single PKS polypeptide presenting an atypical two-module, six-domain organization, also containing domains of unknown functions. The presence of a PPTase gene has also been noticed.[131]

5.4.6 Order Vibrionales

(a) Genus Photobacterium. *P. profundum* strain SS9, which was isolated at a depth of 2500 m, is a piezophile that can grow over a 90 MPa pressure range. Its genome[26] consists of two chromosomes of 4.1 and 2.2 Mb. The 4.1 Mb chromosome contains a likely *ybt* cluster (Fig. 9) and a PUFA-like locus (Fig. 6). The smaller chromosome contains a TMS cluster, with a putative transposase gene on one side, highly related to the *Erwinia* locus, annotated as responsible for enterobactin **3** production (Fig. 7).

(b) Genus Vibrio. *Vibrio* spp. represent a significant portion of the culturable heterotrophic bacteria of marine environments. This genus also includes serious pathogens for finfish, shellfish and mammals. *V. cholerae* is the aetiological agent of the severe diarrhoeal disease cholera. *V. vulnificus* causes, in at risk patients, fatal and rapidly-progressing septicemia with high mortality associated with the consumption of contaminated raw seafood. *V. vulnificus* YJ016,[108] which has two circular chromosomes of 3.4 and 1.9 Mb, contains one cluster encoding two NRPSs highly related to VibF, two aryl-activating enzymes, proteins for DHB synthesis, as well as a PPTase. A virtually identical cluster is present in the unpublished sequence of *V. vulnificus* strain CMCP6.[11] This cluster is likely to participate in vulnibactin **27** formation, a siderophore containing both DHB and salycilate produced by another *V. vulnificus* strain.[132]



27 vulnibactin

The sequenced *V. cholerae* strain[105] encodes only part of the machinery for vibriobactin **6** synthesis, since *vibF* is inactivated by a frameshift mutation. The other biosynthetic genes are intact, so this strain might synthesize a moiety with iron-chelating activity similar to DHB-norspermidine **7**. No TMS have been found in the genome of the other two sequenced *Vibrio* strains.

5.4.7 Order Xanthomonadales, genus *Xanthomonas*

This genus includes diverse and economically important phytopathogens. *X. axonopodis* pv. *citri* (*Xac*) causes citrus canker in most commercial citrus cultivars, with significant losses worldwide. *X. campestris* pv. *campestris* (*Xcc*), which causes black rot in crucifers, is grown commercially to produce the exopolysaccharide xanthan gum. The genomes of one *Xac* strain, of two *Xcc* strains, of one *X. campestris* pv. *vecsicatoria* and of *X. oryzae* (the causal agent of bacterial blight on rice) were available (Table 5). *Xac* and the two *Xcc* strains share a cluster encoding a C-A-T-TE NRPS as well as arginine decarboxylase, glycosyltransferase and acyltranferase as separate polypeptides. *Xac* also contains one cluster encoding a trimodular NRPS of unknown function, flanked by a transposase gene and a tRNA gene.

5.4.8 Order Oceanospirillales, genus *Hahella*

The heterotroph *H. chejuensis*, isolated from coastal marine sediments in the southern part of Korea, produces an algicidal agent active against problematic red-tide dinoflagellates. The authors found that the algicidal activity was associated with fractions containing the red pigment produced by *H. chejuensis*.[22] When searching its 7.2 Mb circular chromosome for genes possibly involved in red pigment synthesis, genes similar to the *S. coelicolor red* genes were identified (Fig. 5). The *H. chejuensis* cluster, when expressed in *E. coli*, enabled production of a red pigment, while transposon insertions abolished it. Using LC-ESI-MS/MS analysis and NMR, the compound was identified as prodigiosin **8**. In addition to the prodigiosin cluster, this strain harbors seven other TMS clusters often associated with horizontally-acquired genomic islands. One cluster is highly related to those involved in enterobactin **3** formation, and another is similar to the *C. violaceum* chrysobactin **5** locus (Fig. 7). The other clusters encode: a hybrid system consisting of a trimodular PKS and an 8-module NRPS; two PKS modules and seven NRPS modules possibly involved in siderophore formation; an NRPS module (C-A-T), five KS and six ACP domains, as well as a separate AT protein; a monomodular NRPS; and a trimodular NRPS.

Table 7 Analyzed genomes from the δ- and ε-Proteobacteria[a]

Strain	Order	Genome[b]	Reference
Bdellovibrio bacteriovorus HD100	Bdellovibrionales	3.8	133
Desulfotalea psychrophila LSv54	Desulfobacterales	3.7	134
Desulfovibrio vulgaris Hildenborough	Desulfovibrionales	3.8	135
Geobacter sulfurreducens PCA	Desulfuromonadales	3.8	136
Campylobacter jejuni NCTC 1168	Campylobacterales	1.6	137
Campylobacter jejuni RM1221	Campylobacterales	1.8	138
Helicobacter hepaticus ATCC 51449	Campylobacterales	1.8	139
Helicobacter pylori 26695	Campylobacterales	1.7	140
Helicobacter pylori J99	Campylobacterales	1.6	141
Wolinella succinogenes DSM 1740	Campylobacterales	2.1	142

[a] These strains do not harbor TMS genes. [b] Cumulative size of chromosomes and plasmids, in Mb.

5.5 Classes δ- and ε-Proteobacteria

None of published genomes from the δ- and ε-Proteobacteria contains TMS genes (Table 7). As mentioned before, no myxobacterial genome is publicly available.

6 Phylum Firmicutes

Genome sequences were available for representatives of the classes Bacilli (45 strains from 24 species; Table 8), Clostridia and Mollicutes (5 and 11 species, respectively; Table 9). None of the Mollicutes, which lack the cell wall and are characterized by a marked genome reduction, contains TMS genes.

6.1 Class Bacilli, order Bacillales

Twenty four genomes were published: 10 from the family Bacillaceae, 3 from the family Listeriaceae, which does not contain TMS genes; and 11 from the family Staphylococcaceae.

6.1.1 Family Bacillaceae

The production of bacillibactin-related siderophores seems a common characteristic for this family, since the *dhb* cluster (Fig. 4) is present in 9 published genomes. The DhbE proteins from *Bacillus licheniformis*, from strains of the *Bacillus cereus* group and from *Oceanobacillus iheyensis* share 75, 71 and 60% identity, respectively, with the *Bacillus subtilis* counterparts. Identity scores for the DhbF proteins are 67, 64 and 47%, respectively. Within the *B. cereus* group, the DhbE and DhbF sequences are over 91% identical. Cluster organization is practically identical in all 9 genomes, except that the PPTase gene is not cluster-linked in *B. subtilis* and *B. licheniformis.*

(a) Genus Bacillus. *Bacillus* spp. are soil bacteria that, under starvation conditions, can initiate a pathway that leads to formation of highly resistant spores. All the complete genome sequences of members of this genus contain TMS genes, with the exception of *B. halodurans* C-125.

• *Bacillus subtilis. B. subtilis* strain 168 has long been a model system for prokaryote differentiation and was the first sequenced Gram positive.[149] The sequenced *B. subtilis* strain contains an inactive allele of the PPTase gene *sfp*.[181,182] This strain is therefore unable to convert TMS enzymes into the active holoforms. However, lipopeptide synthesis can be restored by transformation with an active *sfp* gene.[181] In addition to the bacillibactin **4** cluster (Fig. 4), this strain contains three TMS clusters, some of which direct the synthesis of compounds known in other *B. subtilis* strains: the *pps* operon encodes five NRPSs and is responsible for the synthesis of the lipodecapeptide plipastatin **28**[181]; the *srf* cluster consists of three NRPS genes and directs formation of the lipoheptapeptide surfactin **29a**[183]; the *pks* cluster[184] consists of nine PKS genes and probably directs the synthesis of bacillaene or difficidin **30**,[185] produced by the related strain *B. subtilis* A1/3.[186] Different *B. subtilis* strains are known to synthesize the highly related lipopeptides iturin **31**, mycosubtilin **32** and bacillomycin **33**. However, the corresponding genes are not present in strain 168, consistent with the fact that *B. subtilis* strains seem to be able to produce only a subset of the compounds known for the species.[185]

• *Bacillus licheniformis. B. licheniformis*, used in the industrial production of enzymes and metabolites, is a close relative of *B. subtilis* and approximately 80% of the *B. licheniformis* genes have an ortholog in *B. subtilis*.[148] With respect to secondary metabolism genes, *B. licheniformis* shares with *B. subtilis* a bacillibactin **4** cluster (67–75% identity) and a cluster for the synthesis of the surfactin-related lipopeptide lichenysin **29b**, with the seven-module NRPSs 93% identical to those from the lichenysin cluster of *B. licheniformis* ATCC 10716.[187] While strain ATCC 10716 also contains the cluster for bacitracin **34** synthesis,[188] no other TMS genes are present in strain ATCC 14580.

• *Bacillus cereus group*. *B. cereus* is a soil-dwelling opportunistic pathogen that can cause food poisoning. It is very similar to *B. anthracis*, the causative agent of anthrax. Within the *B. cereus* group, published genomes are available for three *B. cereus* strains and three *B. anthracis* strains. All six genomes show the presence of a *dhb* cluster (Fig. 4), which was shown to direct siderophore production in *B. anthracis* Sterne[189] and *B. cereus*.[190] In *B. cereus* ATCC 14579, the *dhbF* gene is interrupted by a stop codon.[146] *B. cereus* E33L contains an additional cluster encoding three bimodular NRPSs and a PPTase. The NRPSs are annotated as involved in production of mycosubtilin **32**, but the cluster does not encode enzymes for the synthesis of the mycosubtilin lipid moiety[191] and the amino acids predicted from some A domains are not consistent with mycosubtilin structure. *B. cereus* ATCC 14579 contains two TMS regions of unknown function, one consisting of a single NRPS gene and the other comprising several TMS genes.

Notably, none of the analyzed *B. cereus* strains harbors TMS genes for the NRPS-synthesized emetic toxin cereulide **35** (the

Table 8 Analyzed genomes from the Firmicutes, class *Bacilli*

Strain	Order	Genome[a]	TMS[b]	Reference
Bacillus anthracis Ames	Bacillales	5.2	9.0	143
Bacillus anthracis "Ames Ancestor"	Bacillales	5.5	9.0	144
Bacillus anthracis Sterne	Bacillales	5.2	9.0	144
Bacillus cereus ATCC 10987	Bacillales	5.4	9.0	145
Bacillus cereus ATCC 14579	Bacillales	5.4	34.9	146
Bacillus cereus E33L	Bacillales	5.3	32.6	144
Bacillus halodurans C 125	Bacillales	4.2	0	147
Bacillus licheniformis ATCC 14580	Bacillales	4.2	35.0	148
Bacillus subtilis str.168	Bacillales	4.2	142.9	149
Enterococcus faecalis V583	Lactobacillales	3.4	0	150
Geobacillus kaustophilus HTA426	Bacillales	3.6	0	151
Lactobacillus acidophilus NCFM	Lactobacillales	2.0	0	152
Lactobacillus johnsonii NCC 533	Lactobacillales	2.0	0	153
Lactobacillus plantarum WCFS1	Lactobacillales	3.3	18.3	154
Lactobacillus sakei subsp. *sakei* 23K	Lactobacillales	1.8	0	155
Lactococcus lactis subsp. *lactis* Il1403	Lactobacillales	2.4	0	156
Listeria innocua Clip 11262	Bacillales	3.1	0	157
Listeria monocytogenes EGD-e	Bacillales	2.9	0	157
Listeria monocytogenes str.4b F2365	Bacillales	2.9	0	158
Oceanobacillus iheyensis HTE831	Bacillales	3.6	9.0	159
Staphylococcus aureus subsp. *aureus* COL	Bacillales	2.8	7.2	160
Staphylococcus aureus subsp. *aureus* MRSA252	Bacillales	2.9	7.2	161
Staphylococcus aureus subsp. *aureus* MSSA476	Bacillales	2.8	7.2	161
Staphylococcus aureus subsp. *aureus* MW2	Bacillales	2.8	7.2	162
Staphylococcus aureus subsp. *aureus* Mu50	Bacillales	2.9	7.2	163
Staphylococcus aureus subsp. *aureus* N315	Bacillales	2.8	7.2	163
Staphylococcus epidermidis ATCC 12228	Bacillales	2.6	7.2	164
Staphylococcus epidermidis RP62A	Bacillales	2.6	7.2	160
Staphylococcus haemolyticus JCSC1435	Bacillales	2.5	0	165
Staphylococcus saprophyticus ATCC 15305	Bacillales	2.7	0	166
Streptococcus agalactiae 2603V/R	Lactobacillales	2.2	0	167
Streptococcus agalactiae A909	Lactobacillales	2.1	0	168
Streptococcus agalactiae NEM316	Lactobacillales	2.2	0	169
Streptococcus mutans UA159	Lactobacillales	2.0	26.3	170
Streptococcus pneumoniae R6	Lactobacillales	2.0	0	171
Streptococcus pneumoniae TIGR4	Lactobacillales	2.2	0	172
Streptococcus pyogenes M1 GAS	Lactobacillales	1.9	0	173
Streptococcus pyogenes MGAS5005	Lactobacillales	1.8	0	174
Streptococcus pyogenes M28 MGAS6180	Lactobacillales	1.9	0	175
Streptococcus pyogenes MGAS10394	Lactobacillales	1.9	0	176
Streptococcus pyogenes MGAS315	Lactobacillales	1.9	0	177
Streptococcus pyogenes MGAS8232	Lactobacillales	1.9	0	178
Streptococcus pyogenes SSI-1	Lactobacillales	1.9	0	179
Streptococcus thermophilus CNRZ1066	Lactobacillales	1.8	0	180
Streptococcus thermophilus LMG 18311	Lactobacillales	1.8	2.9	180

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb.

corresponding genes were detected in 30 cereulide-producing *B. cereus* strains[192]) or for the NRPS/PKS-made zwittermicin A **36**.[193]

(b) Other Bacillaceae. The deep sea extremophile *O. iheyensis* contains only an ortholog of the *dhb* cluster (Fig. 4), while no TMS genes are present in the genome of the moderate deep sea thermophile *Geobacillus kaustophilus*.

6.1.2 Family Staphylococcaceae, genus *Staphylococcus*

Staphylococci, generally found inhabiting the skin and mucous membranes of mammals and birds, can become opportunistic pathogens. No TMS genes are present in the genomes of *S. haemolyticus* JCSC1435 and *S. saprophyticus* subsp. *saprophyticus* ATCC 15305. All six *S. aureus* strains and two *S. epidermidis* strains contain a region of unknown function encoding a bimodular NRPS and a PPTase. Sequence identity for the NRPS is 98% intraspecies and 52% interspecies.

6.2 Class Bacillales, Order Lactobacillales

Fifteen and five genomes have been published from the families *Streptococcaceae* and *Lactobacillaceae*, respectively. Within the *Enterococcaceae*, the genome of the opportunistic pathogen *Enterococcus faecalis* V583 does not contain TMS genes.

6.2.1 Family Streptococcaceae, genus *Streptococcus*

Streptococci are facultative or obligate anaerobes that vary widely in pathogenic potential. For *S. mutans*, the main cause of tooth decay, the genome of strain UA159 contains a single TMS cluster, which encodes a hybrid polypeptide (consisting of KS-T-C domains), five NRPSs (for a total of 7 modules) and a PPTase. The identity of the corresponding metabolite has not been reported. The only other *Streptococcus* strain displaying TMS genes is *S. thermophilus* LMG 18311 (a lactic acid bacterium often used as a starter culture in yogurt and cheese), which encodes a single monomodular NRPS.

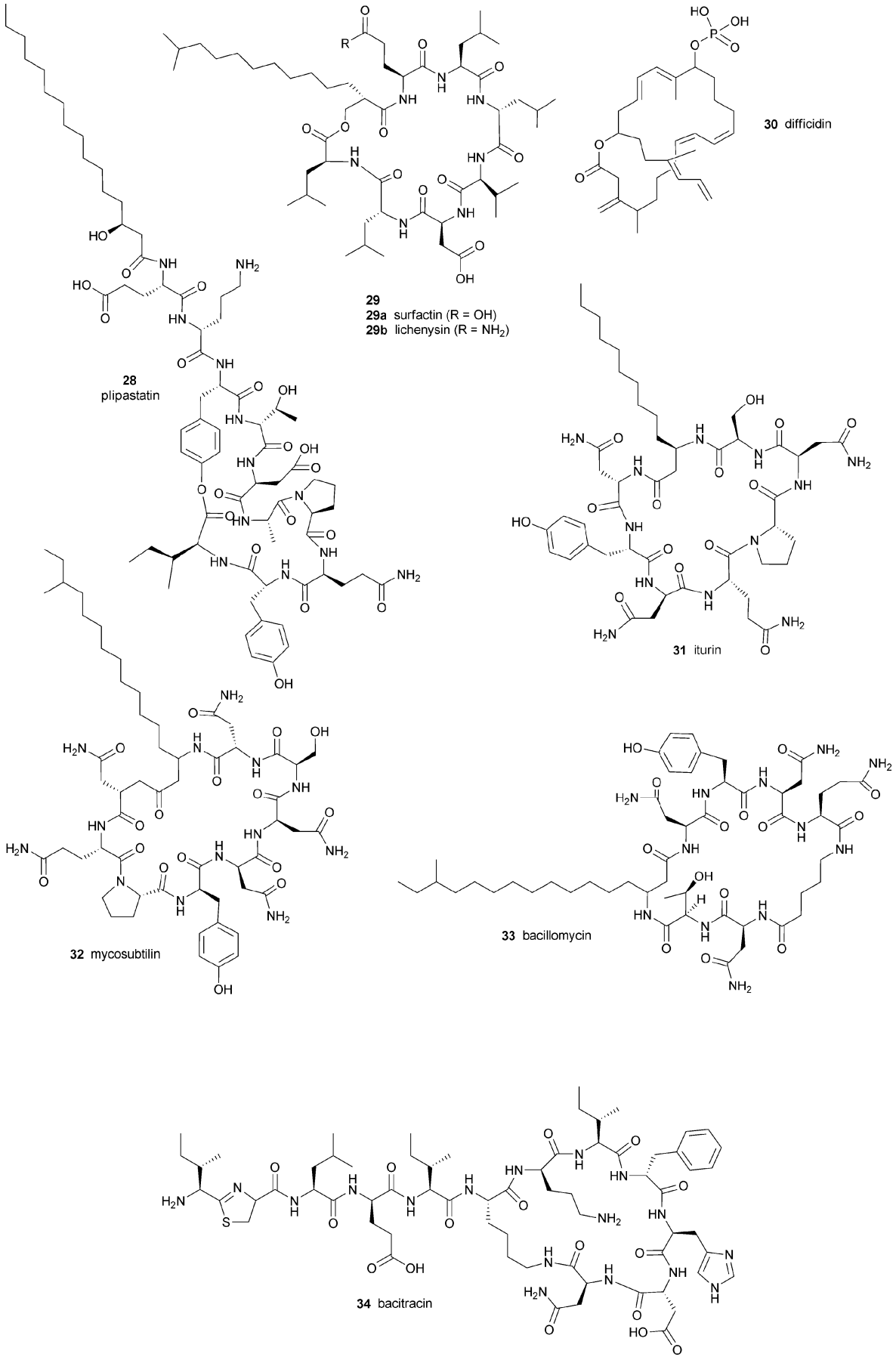
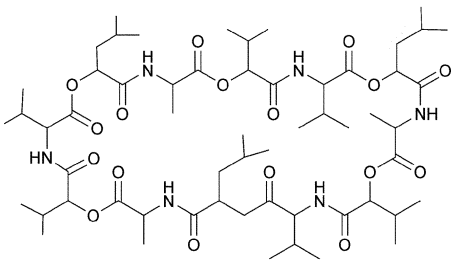

28
plipastatin
29
29a surfactin (R = OH)
29b lichenysin (R = NH2)
30 difficidin
31 iturin
32 mycosubtilin
33 bacillomycin
34 bacitracin

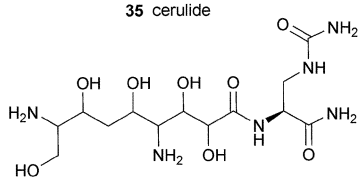
Table 9 Analyzed genomes from the Firmicutes, classes Clostridia and Mollicutes

Strain	Order	Genome[a]	TMS[b]	Reference
Carboxydothermus hydrogenoformans Z-2901	Clostridiales	2.4	0	194
Clostridium acetobutylicum ATCC 824	Clostridiales	4.1	5.4	195
Clostridium perfringens 13	Clostridiales	3.1	0	196
Clostridium tetani E88	Clostridiales	2.9	0	197
Thermoanaerobacter tengcongensis MB4	Thermoanaerobacterales	2.7	0	198
Mycoplasma gallisepticum R	Mycoplasmatales	1.0	0	200
Mycoplasma genitalium G37	Mycoplasmatales	0.6	0	201
Mycoplasma hyopneumoniae 232	Mycoplasmatales	0.9	0	202
Mycoplasma hyopneumoniae 7448	Mycoplasmatales	0.9	0	203
Mycoplasma hyopneumoniae J	Mycoplasmatales	0.9	0	203
Mycoplasma mobile 163K	Mycoplasmatales	0.8	0	204
Mycoplasma mycoides subsp *mycoides* SC str. PG1	Mycoplasmatales	1.2	0	205
Mycoplasma penetrans HF2	Mycoplasmatales	1.4	0	206
Mycoplasma pneumoniae M129	Mycoplasmatales	0.8	0	207
Mycoplasma pulmonis UAB-CTIP	Mycoplasmatales	1.0	0	208
Mycoplasma synoviae 53	Mycoplasmatales	0.8	0	203
Onion yellow phytoplasma OY-M	Acholeplasmatales	0.9	0	199
Ureaplasma parvum sv 3 ATCC 700970	Mycoplasmatales	0.8	0	209

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb.



35 cerulide



36 zwittermicin

6.2.2 Family Lactobacillaceae

All the completed genomes from this family belong to members of the genus *Lactobacillus*, which comprises several species of great importance for the food industry. Only the genome of *L. plantarum* WCFS1,[154] one of the largest known among lactic acid bacteria (3.3 Mb; see Table 8), contains a cluster encoding two NRPSs (for a total of six modules) and a PPTase. In addition, the cluster also encodes proteins supposedly involved in regulation, transport and precursor synthesis.

6.3 Class Clostridia

Of the five complete genomes in this class, only the largest (from *C. acetobutylicum* ATCC 824) contains a single gene for a monomodular PKS of unknown function. *C. acetobutylicum* is a saccharolytic and proteolytic soil bacterium capable of producing a number of organic solvents through fermentation of various organic compounds.

7 Phylum Actinobacteria

The 17 published genomes from this phylum are listed in Table 10. All strains with a genome larger than 4 Mb contain a considerable number of TMS genes, while these systems are not present in *Bifidobacterium*, *Leifsonia*, *Symbiobacter* and *Tropheryma*.

7.1 Common metabolites

Some TMS genes are found in different genera and will be described first. In addition, the unpublished genome of *Thermobifida fusca*[11] presents a cluster similar to those for catechol-based siderophores (Fig. 7), representing the first occurrence of this system in *Actinobacteria* genomes.

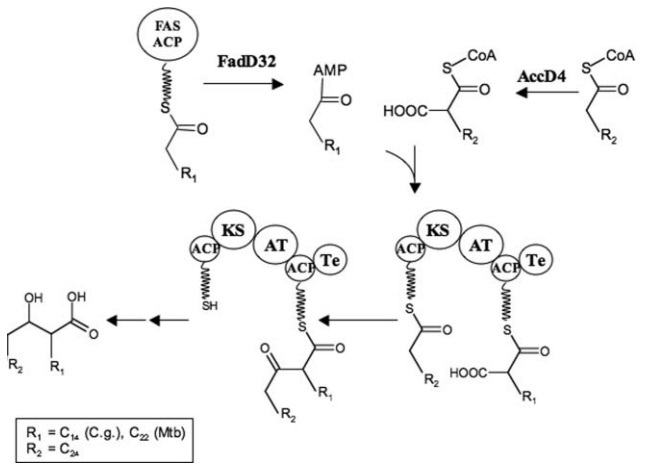
7.1.1 The *pks13* locus

Mycolic acids are long chain branched β-hydroxyl fatty acids that exist either covalently attached to the cell wall or as trehalose dimycolates. They are key components of the cell envelope of mycobacteria and corynebacteria. A cluster involved in the final step of mycolic acid formation is conserved in all *Corynebacterium* and *Mycobacterium* spp. (Fig. 12). It includes six core genes transcribed in the same direction, encoding the acyl-AMP ligase FadD32, Pks13 (consisting of ACP-KS-AT-ACP-TE domains), the acyl-CoA carboxylase AccD4, the mycolyltransferase FbpA and its paralog FbpD, as well as a variable number of acyltransferase and transporter genes. This locus has been shown to participate in the final assembly of mycolic acids in corynebacteria and mycobacteria[228,229] as illustrated in Scheme 7. FadD32 activates a long chain acid into its acyl adenylate and transfers it to the first ACP domain of Pks13. AccD4 carboxylates the CoA thioester of the acyl chain to yield its 2-carboxyl derivative, which is then transferred to the second ACP domain of Pks13, presumably by its AT domain. A Claisen-type condensation between the two fatty acyl groups leads to the formation of the 3-oxo-mycolyl intermediate bound to the C-terminal ACP domain. Reduction of the 3-oxo-intermediate by an unidentified reductase yields the mature mycolic acid. The other genes of the *pks13* locus (Fig. 12)

Table 10 Analysed genomes from the class *Actinobacteria*

Strain	Order	Genome[a]	TMS[b]	Reference
Bifidobacterium longum NCC2705	Bifidobacteriales	2.3	0	210
Corynebacterium diphtheriae NCTC 13129	Actinomycetales	2.5	23.0	211
Corynebacterium efficiens YS-314	Actinomycetales	3.1	8.8	212
Corynebacterium glutamicum ATCC 13032	Actinomycetales	3.3	8.7	213
Corynebacterium jeikeium K411	Actinomycetales	2.4	28.0	214
Leifsonia xyli subsp. *xyli* CTCB07	Actinomycetales	2.6	0	215
Mycobacterium avium subsp. *paratubercolosis* k10	Actinomycetales	4.8	127.0	216
Mycobacterium bovis AF2122/97	Actinomycetales	4.3	130.0	217
Mycobacterium leprae TN	Actinomycetales	3.2	55.0	218
Mycobacterium tuberculosis CDC1551	Actinomycetales	4.4	121.0	219
Mycobacterium tuberculosis H37RV	Actinomycetales	4.4	136.0	220
Nocardia farcinica IFM 10152	Actinomycetales	6.3	204.0	221
Propionibacterium acnes KPA171202	Actinomycetales	2.6	5.0	222
Streptomyces avermitilis MA-4680	Actinomycetales	9.2	356.0	223
Streptomyces coelicolor A3(2)	Actinomycetales	8.6	135.0	28
Symbiobacterium thermophilum[c] IAM 14863	Unclassified	3.5	0	224
Tropheryma whipplei Twist	Actinomycetales	0.9	0	225
Tropheryma whipplei TW08/27	Actinomycetales	0.9	0	226

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb. [c] Recently reclassified to an outer branch of the Actinobacteria, no order assigned yet.[227]



Scheme 7

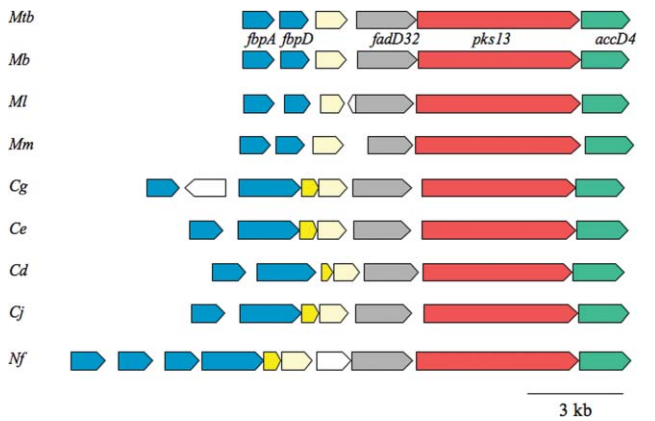


Fig. 12 The *pks13* locus from mycolate-containing bacteria: *M. tuberculosis* (*Mtb*), *M. bovis* (*Mb*), *M. leprae* (*Ml*), *M. marinum* (*Mm*), *C. glutamicum* (*Cg*), *C. efficiens* (*Ce*), *C. diphtheriae* (*Cd*), *C. jeikeium* (*Cj*), *N. farcinica* (*Nf*). Color codes are: red, PKS; green, acyl-CoA-carboxylase; grey, acyl-CoA synthase; blue, mycolyltransferase; yellow, other conserved genes; white, unknown. Except for white, identical colours indicate orthologs.

are mycolyltransferases likely involved in export and translocation of mature mycolic acids to the cell wall and outer lipid layer. *Nocardia farcinica* also possesses an orthologous *pks13* cluster (Fig. 12), suggesting its involvement in mycolic acid synthesis.

7.1.2 Mycobactin-related siderophores

The *Mtb* genome also contains a cluster involved in the formation of the siderophore mycobactin **37**.[230] The six proteins MtbA to MtbF constitute a 20-domain assembly line with an NRPS–PKS–NRPS order that activates and elongates the monomers needed for mycobactin. MbtA activates salicylate **2b** and attaches it to the first T domain of MbtB, which condenses it with serine (threonine) followed by (methyl)oxazoline ring formation by the Cy domain. MbtE and MbtF likely activate the two lysine residues incorporated into mycobactin, but the order in which they act has not been determined. The β-hydroxy acyl moiety present in mycobactin could arise from MbtD, a PKS module with two ACP domains. Finally, the cyclization of the seven-membered ring and release of mycobactin can be assigned to a variant TE or C domain. Orthologous *mbt* clusters are present in all mycobacteria strains characterized so far[231] (Fig. 13, Table 11), except for *M. avium paratuberculosis*. This strain lacks *mbtI* (required for salycilate formation) and *mbtJ* (thioesterase) is replaced by an *nbtA* ortholog. In addition, an extra gene (encoding A and T domains) is found between *mbtE* and *mbtF*. The *N. farcinica* orthologs, designated *nbtABCDEFGH*, are supposed to govern

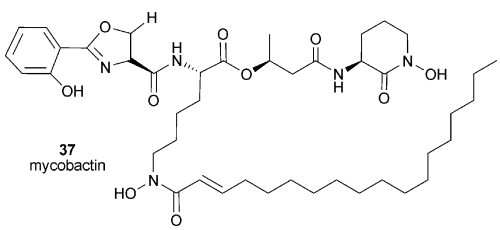


Table 11 Orthologous TMS clusters from *Mycobacterium*[a]

TMS[b]	*Mtb* H37Rv	*Mtb* CDC1551	*Ml*	*Mav*	*Mb*	Product
mas	Yes	Yes	—	—	Yes	Mycocerosic acid
mbt	Yes	Yes	—	Yes	Yes	Mycobactin
msl1	Yes	—	Yes	—	Yes	Unknown
msl2	Yes	Yes	—	Yes	Yes	Hydroxyphthioceranic acids
msl3	Yes	Yes	—	—	Yes	Unknown
msl4	Yes	Yes	—	Yes	Yes	Phthiocerol dimycocerosate esters
msl5	Yes	Yes	—	Yes	Yes	Phthiocerol dimycocerosate esters
msl6	Yes	Yes	—	Yes	Yes	Mannosyl-β-1-phosphomycoketide
msl7	Yes	Yes	Yes	—	Yes	Phenolphthiocerol
nrps[c]	Yes	Yes	Yes	—	Yes	Unknown
pks6	Yes	Yes	—	—	Yes	Unknown
pks9	Yes	Yes	—	Yes	Yes	Phthiocerol dimycocerosate esters
pks13	Yes	Yes	Yes	Yes	Yes	Mycolic acids
pks16	Yes	Yes	Yes	Yes	Yes	Unknown
pps	Yes	Yes	Yes	—	Yes	Phthiocerol dimycocerosate esters

[a] 'Yes' indicates that an ortholog is present. [b] See text for cluster abbreviations. [c] Dimodular NRPS, the Ml ortholog has only one module.

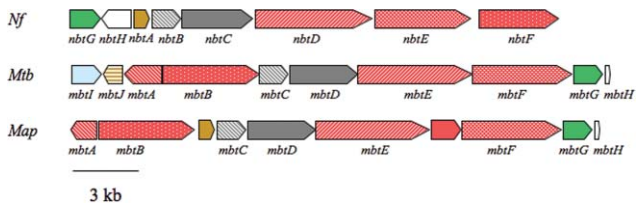


Fig. 13 The mycobactin-like clusters from *N. farcinica* (*Nf*), *M. tuberculosis* H37Rv (*Mtb*) (substantially identical clusters in *M. tuberculosis* CDC1551 and *M. bovis*) and *M. avium* subsp. *paratuberculosis* (*Map*). Color codes are: red, NRPS; grey, PKS; green, lysine-*N*-oxygenase; brown, (thio)esterase; light blue, salycilate biosynthesis; white, other functions. Ortholog genes have identical patterns.

the synthesis of a mycobactin-related siderophore,[221] although no siderophore has yet been described from *Nocardia* spp.

7.2 Genus *Corynebacterium*

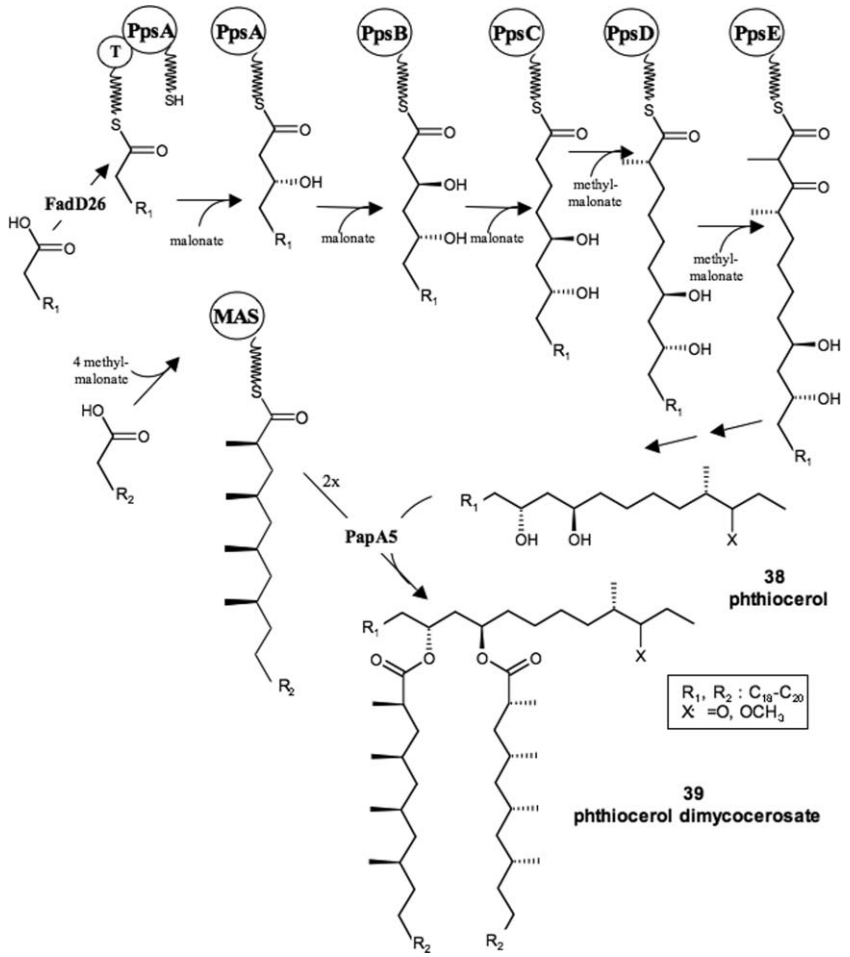
Corynebacteria present irregular cell morphology and a broad G + C content (51–70%). Representatives characterized through genome sequencing include: *C. diphtheriae*, the etiological agent of diphtheria; *C. glutamicum*, widely used for the industrial production of L-glutamate and L-lysine; *C. efficiens*, a close relative of *C. glutamicum*; and *C. jeikeium*, a human skin flora inhabitant that can become a serious, multi-resistant nosocomial pathogen. In addition to the *pks13* cluster (Fig. 12), all corynebacteria share a gene of unknown function encoding a 130 kDa protein containing A and T domains. Additional TMS genes are described below.

C. diphtheriae[211] contains a cluster encoding a hybrid PKS–NRPS (one module each), a separate monomodular NRPS and an ABC transporter. This cluster may direct siderophore synthesis, since the NRPS is similar to *P. aeruginosa* PchF and the ABC transporter to *Y. pestis* YbtP. *C. jeikeium* contains two TMS clusters, one encoding a monomodular NRPS and an isolated A domain, and the other a trimodular NRPS.

7.3 Genus *Mycobacterium*

The genus *Mycobacterium* includes severe pathogens, such as *M. tuberculosis* (*Mtb*), the causative agent of tuberculosis, *M. leprae* (*Ml*), the causative agent of leprosy, and *M. avium*, an important opportunistic pathogen in immunocompromised patients. In addition to mycolic acids, the mycobacterial cell envelope contains a vast repertoire of multimethyl-branched long-chain acids and alcohols, including mycocerosic and mycopenic acids, as well as variously substituted phthiocerols **38**. Five mycobacterial genomes were available (Table 10): the *Mtb* strains H37Rv and CDC1551, and one strain each for *Ml*, *M. bovis* (*Mb*) and *M. avium* subsp. *paratubercolosis* (*Map*).

Mtb strain H37Rv harbors the full set of TMS genes found in the other mycobacteria organized as 16 distinct clusters (Table 11). A function has been assigned to some of them.[232] The *mas* and *ppsABCDE* genes are clustered on the *Mtb* chromosome in a region that also encodes acyl-CoA synthases and transporters important for biosynthesis and translocation of dimycocerosate esters.[233,234] The enzymatic activities of PpsA, PpsB and PpsE have been experimentally determined.[232] The five Pps enzymes elongate long-chain fatty acids with malonate or methylmalonate units (Scheme 8). The acyl-AMP ligase FadD26 activates an acyl chain and transfers it to the first T domain of PpsA. PpsA, -B and -C elongate the acyl chain with one malonate unit each, followed by β-ketoreduction (PpsA and -B) or by full processing to methylene (PpsC). Next, PpsD and PpsE add one methylmalonate, followed by reduction of the β-keto group to methylene. PpsE, which does not contain any processing domains, adds the final methylmalonate contain and uses its C domain for chain release.[234] In a process which has yet to be completely defined, the released chain undergoes decarboxylation and other modifications (*e.g.* reduction and methylation) to yield the phthiocerols **38** (Scheme 8). Mycocerosic acids are synthesized by MAS, another PKS consisting of KS-AT-DH-ER-KR-ACP domains (Scheme 8), and selective for methylmalonyl-CoA, producing multimethyl-branched acids from *n*-acyl primers.[235] The PapA5 protein, also encoded by the *pps-mas* cluster, consists of a C domain that mediates mycocerosic acid transfer from MAS to phthiocerol[235] and is responsible for the last step in the assembly of the (phenol)phthiocerol dimycocerosate esters **39**. In a variation of the theme, the Pks15/1 protein (containing KS, AT, KR, DH, ER, ACP domain) is required for adding a *p*-hydroxybenzoic acid moiety to the growing chain, ultimately yielding the phenolphthiocerol derivatives.[236]



Scheme 8

Mtb H37Rv contains seven loci,[234] designated *msl* and consisting of single genes or adjacent gene pairs, each encoding full set(s) of the domains KS-AT-DH-ER-KR-ACP, which would be sufficient for the synthesis of a branched fatty acid as exemplified by MAS. These loci can be subdivided into two families on the basis of sequence identity: one includes *msl1* (*pks5*), *msl2* (*pks2*), and *msl3* (*pks3/4*), as well as *mas* itself; and *msl4* (*pks7*), *msl5* (*pks8/17*), *msl6* (*pks12*), and *msl7* (*Pks15/1*). Different *msl* loci probably direct the synthesis of the different multiple methyl-branched fatty acids found in *Mtb*.

Orthologs of *msl1* (role not yet known) are present in *Ml* and *Mb* but not in the *Mtb* strain CDC1551 (Table 11). Orthologs of *msl2* are present in *Map* and *Mb*, in addition to strain CDC1551 (Table 11). This locus is responsible for the formation of the hydroxyphthioceranates, the major acyl constituents of sulfolipids, which are present uniquely in virulent *Mtb* strains where they play a significant role in interaction with the host.[237]

The *msl3* locus consists of two separate genes (*pks3* and *pks4*) in strain H37Rv, fused as a single gene in strain CDC1551 and in *Mb*. The *msl3* locus also encodes the proteins PapA1, PapA2 and PapA3, which may catalyze *O*-esterification of trehalose with the methyl-branched polyketides produced by Pks3 and Pks4. Since methyl-branched polyketides are not found as free acids in *Mtb*—and these PKSs do not contain a TE domain—the Pap proteins may directly transfer ACP-bound polyketide to threalose, a reaction reminiscent of the acyl transferases involved in lipid A synthesis. *Mtb* contains an unusual cluster consisting of *msl4*, *msl5* and *pks9* flanked by two genes encoding type III PKSs. This cluster has been implicated in the synthesis of dimycocerosate esters.[237]

Pks12 (*msl6* locus) was recently shown to be responsible for the synthesis of an acylated derivative of mannosyl-β-1-phosphate in human macrophages.[232] Pks12 consists of two head-to-tail sets of MAS domains, with the unique feature that the first set is selective for malonyl-CoA and the second for methylmalonyl-CoA. Orthologs are present in *Mb* and *Map* (Table 11). The *msl7* locus has been proposed to catalyze the elongation of *p*-hydroxybenzoic acid with malonyl-CoA units to form *p*-hydroxyphenylalkanoic acid derivatives, which in turn are used by the Pps proteins to yield phenolphthiocerol **38** and its relatives, as confirmed by the finding that inactivation of *msl7* abolishes phenolphthiocerol **38** production.[236] The *msl7* locus is present in all mycobacteria except *Map* (Table 11).

In addition to the *msl* loci, mycobacteria encode other TMSs of unknown function: Pks6 is a monomodular enzyme with ACP-KS-AT-TE domains, with orthologs present in strain CDC1551 and, as two separate proteins, in *Mb*. Pks16 consists of an A domain only and is present in all mycobacteria (Table 11). It should be noted that all *Mtb* PKSs from strain H37Rv have an ortholog in *Mb*, one more than in *Mtb* strain CDC1551. However, only ten and five PKS orthologs are present in *Map* and *Ml*, respectively.

The lower number of PKSs in *Map* might be due to the lack of dextrorotatory multimethyl-branched acids in this strain.

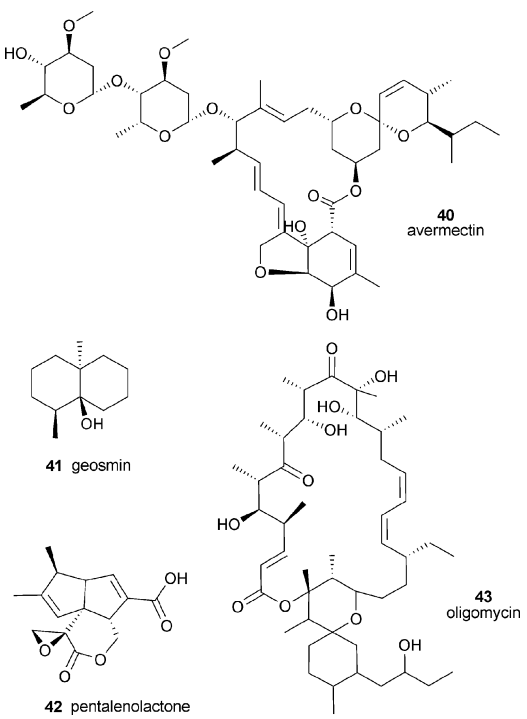
Finally, a bimodular NRPS is also encoded by the *Mtb* genome. Bimodular and monomodular orthologs are present in *Mb* and in *Ml*, respectively. While this NRPS is needed for formation of dimycocerosate esters, its specific role is unknown.[238]

7.4 Genus *Streptomyces*

Members of this genus are soil inhabitants that thrive on complex organic polymers. Furthermore, *Streptomyces* spp. represent the most prolific and versatile genus of natural product producers. Two genome sequences have been published, from the model organism *S. coelicolor* and from the avermectin **40** producer *Streptomyces avermitilis*, and each carries more than 20 clusters for secondary metabolism. Except for the geosmin **41** cluster, all clusters are unique to each of the *Streptomyces* strains.

7.4.1 *Streptomyces avermitilis*

Avermectin **40**, a complex of eight related pentacyclic lactones carrying an oleandrose disaccharide, is an important compound in human and veterinary medicine. The *S. avermitilis* genome—the largest in the set analyzed here—consists of a 9.0 Mb linear chromosome and a small plasmid. Its chromosome possesses at least 30 clusters for secondary metabolism (271 genes, or 6.6% of the genome), involved in the synthesis of melanin, carotenoid, siderophore, polyketide and peptide compounds.[239] Over half of the clusters are located in the subtelomeric regions, while metabolites commonly produced by several *Streptomyces* spp. are directed by clusters located in the 6.5 Mb internal core region of the chromosome (*e.g.* the clusters for geosmin **41**, pentalenolactone **42** and oligomycin **43** formation).



Omura *et al.*[239] performed a detailed analysis of the *S. avermitilis* TMS genes. Only one region, designated *nrps5* and encoding A and T domains, with the A domain predicted to be specific for proline (Table 12),[239] has an ortholog, situated in a similar genetic context, in *N. farcinica*. Seven additional NRPS clusters are unique to *S. avermitilis*,[239] although none of the corresponding metabolites have been identified yet. The domain composition of each module and the predicted amino acids recognized by the A domains are summarized in Table 12. The *nrps1*, *nrps2* and *nrps3* clusters encode three NRPSs each. On the basis of module number, these clusters are predicted to make tetra-, hexa- and tripeptides, respectively.[239] *Nrps4* encodes a monomodular NRPS. The *nrps6* cluster encodes C-A-T domains (with proline specificity) and a long chain fatty acyl-CoA ligase, suggesting that it directs the synthesis of an acylated amino acid. The *nrps7* cluster presents a complex architecture: it encodes 14 distinct polypeptides, many of which carry a single NRPS domain, for a total of 8 A, 11 T, 7 C and 1 TE domains, as well as a PKS module with KS-AT-ACP domains. If active, *nrps7* may direct the synthesis of an octapeptide containing two pipecolate residues and a polyketide moiety. *Nrps8* encodes a single A domain.

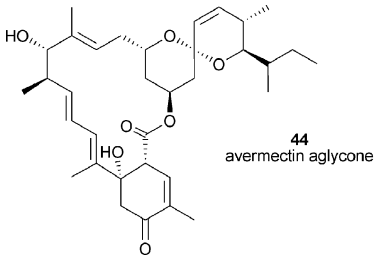
Table 12 Predicted substrates and domain organization of *S. avermitilis* NRPSs

Cluster	Polypeptide–module[a]	Substrate	Domains
nrps1	1	Threonine	C-A-T
	2	Leucine	C-A-T-E
	3–1	Serine	A-T
	3–2	Tyrosine	C-A-T
nrps2	1	Glutamine	A-T
	2–1	Leucine	C-A-T-E
	2–2	Asparagine	C-A-T
	3–1	Serine	C-A-T-E
	3–2	Asparagine	C-A-T-E
	3–3	Asparagine	C-A-T-TE
nrps3	1	Ornithine	C-A-T
	2		E-TE
	3	Serine	A-T
nrps4		Asparagine	A-T
nrps5		Proline	A-T
nrps6		Proline	C-A-T
nrps7	1	Valine	A-T
	2	Glutamate	A
	3	Pipecolate	A-T
	4	Valine	A-T-C
	5	Glutamate	A-T
	6	Pipecolate	C-A-T
	7	Serine	A-T
	8		T-C
	9		C
	10		C
	11	Cysteine	A-T
	12		T-C-T-Te
	13		T-C
nrps8		Alanine	A
nrps9		Cysteine	A

[a] The first and second numbers indicate polypeptides and modules, respectively, if more than one.

S. avermitilis harbors eight PKS clusters,[239] including the previously characterized *ave* cluster for avermectin **40** biosynthesis.[240] The *ave* PKSs consist of twelve modules contributing the 55 catalytic functions required for the formation of the 6,8*a*-seco-6,8a-deoxy-5-oxoavermectin **44**, using isobutyryl- or 2-methylbutyryl-CoA as starter units. The *olm* cluster, involved in the formation of the macrocyclic lactone oligomycin **43**, encodes seven PKSs for a total of 17 modules and 79 catalytic domains. Direct confirmation

of its role came from disruption of selected *olm* genes, which abolished oligomycin production.[239] The *pte* cluster, encoding five PKSs for a total of 13 modules and 57 domains, is predicted to synthesize a 26-membered pentaene macrolide that is yet to be identified.

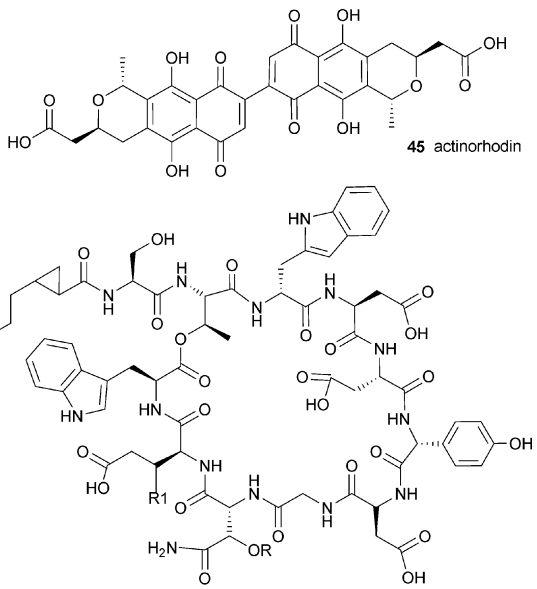


44
avermectin aglycone

S. avermitilis contains five other PKS clusters of unknown function.[239] *Pks1* encodes two unusual modules, consisting of KS-AT-ACP-KR domains (with atypical order and divergent sequence for the ACP-KR pair) and of KS-AT domains, as well as a monofunctional ACP. *Pks2* encodes PKS modules and an A domain, and has been suggested to direct the synthesis of an unknown macrolactam. Clusters *pks3*, *pks4*, *pks5* and *pks7* complete the set of genes possibly participating in the synthesis of structurally undefined polyketides.

7.4.2 *Streptomyces coelicolor*

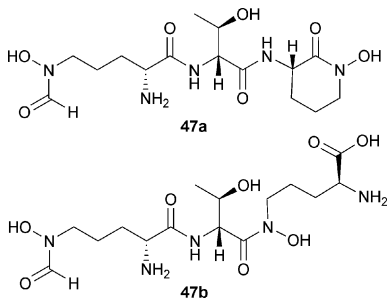
Streptomyces coelicolor A3(2) has an 8.6 Mb linear chromosome with more than 20 secondary metabolism clusters.[28] In addition to the previously characterized clusters for the aromatic polyketides actinorhodin **45** and the structurally undefined spore pigment, *S. coelicolor* harbors 5 clusters containing TMS genes, including the PUFA-like cluster (Fig. 6). Of these, the *cda* cluster, encoding three NRPSs for a total of 11 modules and 37 domains required for the synthesis of the cyclic lipoundecapeptide named CDA **46** (calcium-dependent antibiotic), and the *red* cluster, for the synthesis of the oligopyrrole prodiginines **8**, were previously characterized.

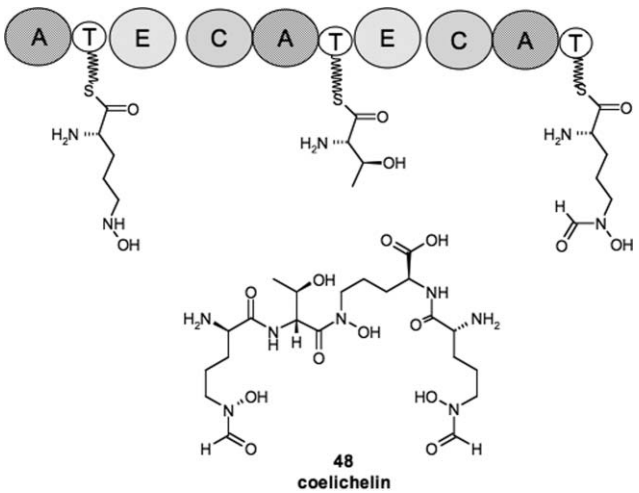


45 actinorhodin

46 CDA CDA1 CDA2 CDA3b CDA4b

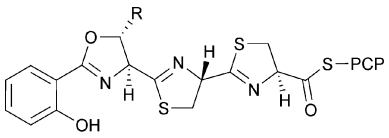
The *cch* cluster encodes the CchH NRPS with an A-T-E-C-A-T-E-C-A-T domain architecture. The substrates for the three A domains were predicted to be L-δ-*N*-formyl-δ-*N*-hydroxyornithine, L-threonine and L-δ-*N*-hydroxyornithine, respectively.[241] This prediction led to two alternate structures (**47a** and **47b**) for the hypothesized product, which was expected to chelate metal ions and therefore dubbed coelichelin.[241] In subsequent work, an iron chelating compound was detected from *S. coelicolor* wild type but not from a *cch* mutant.[242] Using a combination of high resolution and tandem MS and high field NMR,[242] coelichelin was demonstrated to be the tetrapeptide **48**, in which a second D-δ-*N*-formyl-δ-*N*-hydroxyornithine residue was linked to the α-amino group of the L-δ-*N*-hydroxyornithine residue. Consequently, the third Cch module is predicted to act twice. Gene knockouts established the essential role of CchJ, leading to a proposed route for coelichelin formation (Scheme 9).





Scheme 9

Another NRPS cluster, encoding two polypeptides with T-C-A-T-C-A-T and C-A-T-TE domains, is proposed to catalyze the biosynthesis of a novel siderophore named coelibactin. On the basis of the predicted amino acids incorporated by the A domains, the structural core **49**[28] was predicted for this as yet unidentified compound. *S. coelicolor* harbors two additional TMS loci. One is represented by the *cpk* cluster, which encodes a PKS loading module, five extension modules and an additional reductase domain, for which a putative intermediate has been proposed.[243,244] Another TMS cluster encodes two NRPSs with A-T-C-T and C-A-T-TE domains.



49 coelibactin R = H/CH_3

7.5 Genus *Nocardia*

This genus of filamentous soil inhabitants includes the causative agents of nocardiosis in human and animal lungs. The genome of the *N. farcinica* clinical isolate IFM 10152 consists of a single circular 6.0 Mb chromosome and two smaller plasmids.[221] The chromosome contains 22 TMS genes, organized into 13 clusters. To our knowledge, *N. farcinica* is not known to produce secondary metabolites.

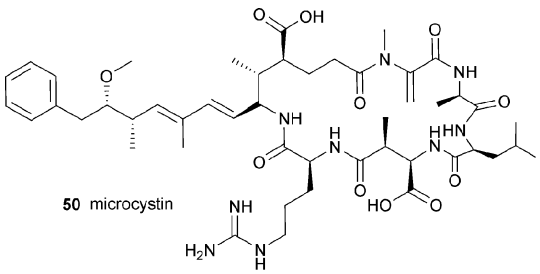
N. farcinica contains an ortholog of the *pks13* locus (Fig. 12), one of the mycobactin cluster (Fig. 13) and one of the *S. avermitilis nrps5* region. In addition, this strain harbors seven TMS clusters of unknown function, orthologs of which are not present in the other bacterial genomes. One cluster encodes a large, 14474-aa NRPS consisting of twelve elongation C-A-T modules and a final C-A module, a second cluster specifies four NRPSs, for a total of 13 A, 12 T, 9 C and 3 TE domains, the third two NRPSs for a total of 13 modules with a terminal TE domain; three separate NRPS genes account for proteins consisting of C-A-T-C-$(C\text{-}A\text{-}T)_3$-C-TE, C-A-T-C-$(C\text{-}A\text{-}T)_4$-TE, and $(C\text{-}A\text{-}T)_5$-TE domains, and three monomodular PKSs are encoded by three unlinked genes. Finally, there are two linked genes encoding individual A and C domains.

7.6 *Propionibacterium acnes*

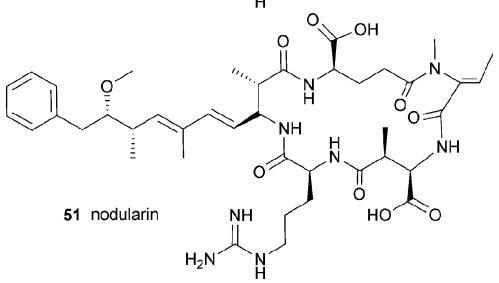
P. acnes is an anaerobic bacterium commonly found on human skin, and the causative agent of acne. Its genome contains only one TMS cluster of unknown function, encoding two NRPSs (for a total of two modules), a separate TE and a PPTase.

8 Phylum Cyanobacteria

The photosynthetic Cyanobacteria commonly proliferate in marine and freshwater habitats, often resulting in bloom formation. Terrestrial genera (*e.g. Nostoc*) can be found too. Secondary metabolism is widespread in Cyanobacteria and often leads to the production of toxic substances associated with cyanobacterial blooms. In particular microcystins **50** and nodularins **51** are cyclic peptides which cause acute hepatotoxicity.[245] These toxins are produced by a hybrid NRPS–PKS, and it has been postulated that a microcystin cluster was present in the last common ancestor of a large portion of modern Cyanobacteria.[245] Several other substances are produced by TMS genes, which are present in most members of the Cyanobacteria.[246] The current set of published genome sequences is biased towards *Synechococcus* and *Prochlorococcus* spp., free living organisms responsible for a large part of the carbon fixation that occurs in marine environments and which possess small genomes (1.7–2.7 Mb) devoid of TMS genes. In contrast, a significant fraction of TMS genes is present in the two strains with larger genomes (Table 13).



50 microcystin



51 nodularin

8.1 Order Nostocales, Genus *Anabaena*

These cyanobacteria are capable of fixing carbon and nitrogen, and can be found worldwide in aquatic and, occasionally, terrestrial environments. They form long filaments presenting a variety of cell types, including heterocysts where nitrogen fixation takes place. *Anabaena* spp. produce toxic blooms in aquatic environments that are harmful or fatal to animals and humans.

8.1.1 *Anabaena* sp. PCC7120

This strain is also referred to as "*Nostoc* sp. PCC7120". A cluster encodes three PKSs consisting of KS-AT-ACP-ACP, AT-KS and KS domains. A nearly identical cluster (>94% identity), located in the same genetic context, is found also in the unpublished genome sequence of *A. variabilis*.[11] The first PKS is highly related (81%

Table 13 Analyzed genomes from the *Cyanobacteria*

Strain	Order	Genome[a]	TMS[b]	Reference
Anabaena sp. PCC 7120	Nostocales	7.2	104.8	247
Gloeobacter violaceus PCC 7421	Gloeobacterales	4.7	50.6	248
Prochlorococcus marinus str. MIT 9313	Prochlorales	2.4	0	249
Prochlorococcus marinus subsp. *marinus*	Prochlorales	1.8	0	247
Prochlorococcus marinus subsp. *pastoris*	Prochlorales	1.7	0	250
Synechococcus sp. WH 8102	Chroococcales	2.4	0	251
Synechocystis sp PCC6803	Chroococcales	3.9	0	252
Thermosynechococcus elongatus BP-1	Chroococcales	2.6	0	253

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb.

Table 14 Analyzed genomes from the class *Bacteroidetes*

Strain	Order	Genome[a]	TMS[b]	Reference
Bacteroides fragilis NCTC 9343	Bacteroidales	5.2	3.9	256
Bacteroides fragilis YCH46	Bacteroidales	5.3	3.9	257
Bacteroides thetaiotaomicron VPI-5482	Bacteroidales	6.3	0	258
Porphyromonas gingivalis W83	Bacteroidales	2.3	0	259
Salinibacter ruber DSM 13855	Sphingobacteria	3.6	0	260

[a] Cumulative size of chromosomes and plasmids, in Mb. [b] Cumulative size of TMS genes, in kb.

identity) to HglE from *Nostoc punctiforme* ATCC 29133, which has been demonstrated to be required for glycolipid synthesis in heterocysts.[254] Glycolipids are needed for protecting the oxygen-sensitive nitrogenase from inactivation, and *hlgE* mutants are unable to fix nitrogen under atmospheric conditions.[254] *Anabaena* sp. PCC7120 shares with *A. variabilis* another TMS cluster, encoding four PKSs and one NRPS, which display over 93% identity for all proteins. However, the first PKS gene in *A. variabilis* is disrupted by an IS, separating KS-AT from KR-ACP domains in the resulting polypeptides. No evidence is available regarding its role.

Additional TMS clusters of unknown function are present. One large cluster consists of nine genes encoding a total of ten NRPS and two PKS modules. Another cluster specifies one monomodular NRPS and two monomodular PKSs. A final cluster encodes a monomodular NRPS and enzymes involved in cell wall biogenesis.

8.2 Order Gloeobacterales: *Gloeobacter violaceus*

This organism was one of the earliest to diverge from the cyanobacterial line. It is an obligate photoautotroph that lacks thylakoid membranes and probably has its photosynthetic machinery in the cytoplasmic membrane with various components exposed to the periplasm. *G. violaceus* contains a cluster resembling the *hgl* locus and encoding orthologs of HglC, HglD and HglE. It should be noted that this cluster encodes an additional monomodular PKS and that *G. violaceus* is a unicellular organism that does not form heterocysts. Thus, this cluster might be involved in the synthesis of a lipid different from those made by *Anabaena* and *Nostoc*. *G. violaceus* encodes other PKSs of unknown function: two monomodular proteins and a cluster comprising two PKS modules and an isolated KS domain.

9 Other phyla

9.1 Phylum Planctomycetes

These microorganisms represent one of the deepest branching bacterial phyla, with several peculiar characteristics, such as lack of peptidoglycan in the cell wall and unique cellular compartimentalization. The 7.1 Mb genome from the marine budding bacterium *Rhodopirellula baltica* SH1[255] encodes two small NRPSs, two monomodular PKSs and a bimodular NRPS-PKS. These five genes are located in different regions of the genome, suggesting that they are involved in the synthesis of five different, unknown products.

9.2 Other phyla

Four genomes (from three species) have been completed for the class Bacteroidetes and a single genome for the class Sphingobacteria (Table 14). TMS genes are present only in two different strains of *Bacteroides fragilis*, an opportunistic pathogen that is the most common anaerobe isolated from clinical specimens. Both strains contain two NRPS genes of unknown function, encoding an isolated A domain and an A-T didomain polypeptide. The genes are 98% identical and in both cases they are approximately 400 kb apart.

No TMS genes are present in the published genomic sequences of strains from the phyla Aquificae, Chlamydiae, Chlorobi, Deinococci, Fusobacteria, Spirochaetes and Thermotogae (Table 15).

10 Conclusions

The current set of available bacterial genomes is providing important insights into the occurrence of genes for natural product biosynthesis in different bacterial lineages. *In primis*, it has confirmed the long-held view that the ability to produce secondary metabolites is not uniformly distributed within the bacterial world.[14] In fact, bacterial taxa known for natural products (*i.e.* the genera *Streptomyces*, *Bacillus* and *Pseudomonas* among those covered in this report) stand out for the high percentage of TMS genes in their genomes. In addition, the available data suggest that natural product formation is a "luxury" that only relatively large genomes can afford, but not all large-genome bacteria possess this feature.

Bacterial genomics has also confirmed the occurrence of the same or very similar natural products in divergent bacterial lineages. Notably, most of these common metabolites are involved in the uptake of essential nutrients. Two observations are noteworthy. The first is that these clusters, which are not uniformly present in all representatives of a given taxon, are relatively prone to inactivation. The second observation is their high conservation which makes them easy to recognize. Thus, horizontal gene transfer has likely played a strong role in spreading these biosynthetic pathways. Until metabolites have been characterized from sequenced bacterial strains, we cannot formally exclude the possibility that highly divergent clusters also direct the synthesis of some of these common metabolites. In any case, we still have a limited understanding about the interplay between vertical and horizontal transmission for natural product biosynthesis.[281]

Despite the fact that the current set of sequenced genomes is biased against secondary metabolite producers, it has highlighted a remarkable number of TMS clusters that do not match those

Table 15 Analyzed genomes from phyla devoid of TMS genes

Phylum	Strain	Genome[a]	Reference
Aquificae	*Aquifex aeolicus* VF5	1.6	261
Chlamydiae	*Chlamydia muridarum* Nigg	1.1	262
Chlamydiae	*Chlamidya trachomatis* A/HAR-13	1.1	263
Chlamydiae	*Chlamidya trachomatis* D/UW-3/CX	1.0	264
Chlamydiae	*Chlamydophila abortus* S26/3	1.1	265
Chlamydiae	*Chlamydophila caviae* GPIC	1.2	266
Chlamydiae	*Chlamydophila pneumoniae* AR39	1.2	266
Chlamydiae	*Chlamydophila pneumoniae* CWL029	1.2	267
Chlamydiae	*Chlamydophila pneumoniae* J138	1.2	268
Chlamydiae	*Candidatus Protochlamydia amoebophila* UWE25	2.4	269
Chlorobi	*Chlorobium* tepidum TLS	2.2	270
Deinococci	*Deinococcus radiodurans* R1	3.3	271
Deinococci	*Thermus thermophilus* HB27	2.1	272
Fusobacteria	*Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586	2.2	273
Spirochaetes	*Borrelia burgdorferi* B31	1.5	274
Spirochaetes	*Borrelia garinii* Pbi	1.0	275
Spirochaetes	*Leptospira interrogans* sv. Copenhageni str. Fiocruz L1-130	4.6	276
Spirochaetes	*Leptospira interrogans* sv. Lai str. 56601	4.7	277
Spirochaetes	*Treponema denticola* ATCC 35405	2.8	278
Spirochaetes	*Treponema pallidum* subsp. *pallidum* str. Nichols	1.1	279
Thermotogae	*Thermotoga maritima* MSB8	1.9	280

[a] Cumulative size of chromosomes and plasmids, in Mb.

directing the synthesis of known compounds. In order to lead to a natural product, a cluster must consist of active, coordinately expressed genes, and the resulting products must undergo the necessary posttranslational modifications and find the appropriate precursors to produce a compound in a concentration high enough for detection. Our understanding of these processes is still limited and, coupled with the inevitable sequencing errors of large-scale sequencing projects, prevents drawing firm conclusions about the number of active clusters discovered from bacterial genomes. From the available data, it appears that similar compounds are made by similar clusters even in divergent lineages. If the reverse is also true, *i.e.* that different clusters direct the synthesis of different compounds, then bacterial genomics has uncovered a large number of potentially new chemical classes. In the late 1980s, when DNA sequencing was first applied to the elucidation of whole biosynthetic pathways, the classes of characterized natural products far outnumbered the known gene clusters. Many clusters for known compounds have been identified since. In a period when new classes of natural products are discovered by chemistry- or bioassay-based approaches at a declining rate, bacterial genomes are revealing a plethora of new gene combinations at an increasing pace. As chemical information provided the impetus for gene sequencing at the end of the last century, so we can expect that genomics will inspire renewed chemical efforts, hopefully providing the sorely-needed novel chemical classes from microbial sources.

11 References

1 A. E. Allsop, *Curr. Opin. Biotechnol.*, 1998, **9**, 637–642; D. McDevitt and M. Rosenberg, *Trends Microbiol.*, 2001, **9**, 611–617.
2 A. Wack and R. Rappuoli, *Curr. Opin. Immunol.*, 2005, **17**, 411–418; D. Serruto and R. Rappuoli, *FEBS Lett.*, 2006, **580**, 2985–2992.
3 R. McDaniel, M. Welch and C. R. Hutchinson, *Chem. Rev.*, 2005, **105**, 543–558; K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925–936; A. M. Hill, *Nat. Prod. Rep.*, 2006, **23**, 256–320.
4 J. Grunewald and M. A. Marahiel, *Microbiol. Mol. Biol. Rev.*, 2006, **70**, 121–146.
5 C. T. Walsh, A. M. Gehring, P. H. Weinreb, L. E. Quadri and R. S. Flugel, *Curr. Opin. Chem. Biol.*, 1997, **1**, 309–315; R. Finking and M. A. Marahiel, *Annu. Rev. Microbiol.*, 2004, **58**, 453–488.
6 G. Yadav, R. S. Gokhale and D. Mohanty, *J. Mol. Biol.*, 2003, **328**, 335–363.
7 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493–505; G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211–224; C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben and D. H. Huson, *Nucleic Acids Res.*, 2005, **33**, 5799–5808.
8 C. D. Reeves, S. Murli, G. W. Ashley, M. Piagentini, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2001, **40**, 15464–15470.
9 P. Caffrey, *ChemBioChem*, 2003, **4**, 654–657; P. Caffrey, *Chem. Biol.*, 2005, **12**, 1060–1062; R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72–79.
10 S. J. Moss, C. J. Martin and B. Wilkinson, *Nat. Prod. Rep.*, 2004, **21**, 575–593.
11 http://img.jgi.doe.gov/cgi-bin/pub/main.cgi.
12 http://www.ncbi.nlm.nih.gov/BLAST/.
13 H. Reichenbach and G. Höfle, *Biotechnol. Adv.*, 1993, **11**, 219–277; K. Gerth, S. Pradella, O. Perlova, S. Beyer and R. Müller, *J. Biotechnol.*, 2003, **106**, 233–253.
14 A. T. Bull, A. C. Ward and M. Goodfellow, *Microbiol. Mol. Biol. Rev.*, 2000, **64**, 573–606.
15 C. K. Stover, X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. L. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K.-S. Wong, Z. Wu, I. T. Paulsenk, J. Reizer, M. H. Saier, R. E. W. Hancock, S. Lory and M. V. Olson, *Nature*, 2000, **406**, 959–964.
16 J. H. Crosa and C. T. Walsh, *Microbiol. Mol. Biol. Rev.*, 2002, **66**, 223–249.
17 J. J. May, T. M. Wendrich and M. A. Marahiel, *J. Biol. Chem.*, 2001, **276**, 7209–7217.
18 D. Rauscher, D. Expert, B. F. Matzanke and A. X. Trautwein, *J. Biol. Chem.*, 2002, **277**, 2385–2395.
19 J. R. Butterton, M. H. Choi, P. I. Watnick, P. A. Carroll and S. B. Calderwood, *J. Bacteriol.*, 2000, **182**, 1731–1738.
20 A. M. Cerdeno, M. J. Bibb and G. L. Challis, *Chem. Biol.*, 2001, **8**, 817; M. G. Thomas, M. D. Burkart and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 171–184.

21 C. T. Walsh, S. Garneau-Tsodikova and A. R. Howard-Jones, *Nat. Prod. Rep.*, 2006, **23**, 517–531.
22 A. K. P. Harris, N. R. Williamson, H. Slater, A. Cox, S. Abbasi, I. Foulds, H. T. Simonsen, F. J. Leeper and G. P. C. Salmond, *Microbiology*, 2004, **150**, 3547–3560.
23 H. Jeong, J. H. Yim, C. Lee, S.-H. Choi, Y. K. Park, S. H. Yoon, C.-G. Hur, H.-Y. Kang, D. Kim, H. H. Lee, K. H. Park, S.-H. Park, H.-S. Park, H. K. Lee, T. K. Oh and J. F. Kim, *Nucleic Acids Res.*, 2005, **33**, 7066–7073.
24 U. Kaulmann and C. Hertweck, *Angew. Chem., Int. Ed.*, 2002, **41**, 1866–1869.
25 B. A. Methé, K. E. Nelson, J. W. Deming, B. Momen, E. Melamud, X. Zhang, J. Moult, R. Madupu, W. C. Nelson, R. J. Dodson, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, R. T. DeBoy, J. F. Kolonay, S. A. Sullivan, L. Zhou, T. M. Davidsen, M. Wu, A. L. Huston, M. Lewis, B. Weaver, J. F. Weidman, H. Khouri, T. R. Utterback, T. V. Feldblyum and C. M. Fraser, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10913–10918.
26 A. Vezzi, S. Campanaro, M. D'Angelo, F. Simonato, N. Vitulo, F. M. Lauro, A. Cestaro, G. Malacrida, B. Simionati, N. Cannata, C. Romualdi, D. H. Bartlett and G. Valle, *Science*, 2005, **307**, 1459–1461.
27 S. D. Bentley, K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill and D. A. Hopwood, *Nature*, 2002, **417**, 141–147.
28 S. L. Chen, C.-S. Hung, J. Xu, C. S. Reigstad, V. Magrini, A. Sabo, D. Blasiar, T. Bieri, R. R. Meyer, P. Ozersky, J. R. Armstrong, R. S. Fulton, J. P. Latreille, J. Spieth, T. M. Hooton, E. R. Mardis, S. J. Hultgren and J. I. Gordon, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 5977–5982.
29 Brazilian National Genome Project Consortium, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 11660–11665.
30 H. H. Barnes and C. A. Ishimaru, *BioMetals*, 1999, **12**, 83–87.
31 K. S. Bell, M. Sebaihia, L. Pritchard, M. T. G. Holden, L. J. Hyman, M. C. Holeva, N. R. Thomson, S. D. Bentley, L. J. C. Churcher, K. Mungall, R. Atkin, N. Bason, K. Brooks, T. Chillingworth, K. Clark, J. Doggett, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, H. Norbertczak, D. Ormond, C. Price, M. A. Quail, M. Sanders, D. Walker, S. Whitehead, G. P. C. Salmond, P. R. J. Birch, J. Parkhill and I. K. Toth, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11105–11110.
32 B. H. Bellaire, P. H. Elzer, S. Hagius, J. Walker, C. L. Baldwin and R. M. Roop, II, *Infect. Immun.*, 2003, **71**, 1794–1803.
33 I. T. Paulsen, C. M. Press, J. Ravel, D. Y. Kobayashi, G. S. Myers, D. V. Mavrodi, R. T. DeBoy, R. Seshadri, Q. Ren, R. Madupu, R. J. Dodson, A. S. Durkin, L. M. Brinkac, S. C. Daugherty, S. A. Sullivan, M. J. Rosovitz, M. L. Gwinn, L. Zhou, D. J. Schneider, S. W. Cartinhour, W. C. Nelson, J. Weidman, K. Watkins, K. Tran, H. Khouri, E. A. Pierson, L. S. Pierson, 3rd, L. S. Thomashow and J. E. Loper, *Nat. Biotechnol.*, 2005, **23**, 873–878.
34 P. A. Sokol, *J. Clin. Microbiol.*, 1986, **23**, 560–562; L. Farmer and M. S. Thomas, *J. Bacteriol.*, 2004, **186**, 270–277.
35 M. T. G. Holden, R. W. Titball, S. J. Peacock, A. M. Cerdeño-Tárraga, T. Atkins, L. C. Crossman, T. Pitt, C. Churcher, K. Mungall, S. D. Bentley, M. Sebaihia, N. R. Thomson, N. Bason, I. R. Beacham, K. Brooks, K. A. Brown, N. F. Brown, G. L. Challis, I. Cherevach, T. Chillingworth, A. Cronin, B. Crossett, P. Davis, D. DeShazer, T. Feltwell, A. Fraser, Z. Hance, H. Hauser, S. Holroyd, K. Jagels, K. E. Keith, M. Maddison, S. Moule, C. Price, M. A. Quail, E. Rabbinowitsch, K. Rutherford, M. Sanders, M. Simmonds, S. Songsivilai, K. Stevens, S. Tumapa, M. Vesaratchavest, S. Whitehead, C. Yeats, B. G. Barrell, P. C. F. Oyston and J. Parkhill, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14240–14245.
36 H. S. Kim, M. A. Schell, Y. Yu, R. L. Ulrich, S. H. Sarria, W. C. Nierman and D. DeShazer, *BMC Genomics*, 2005, **6**, 174.
37 E. Kiss, B. L. Reuhs, J. S. Kim, A. Kereszt, G. Petrovics, P. Putnoky, I. S. Dusha, R. W. Carlson and A. Kondorosi, *J. Bacteriol.*, 1997, **179**, 2132–2140; A. Kereszt, E. Kiss, B. L. Reuhs, R. W. Carlson, A. Kondorosi and P. Putnoky, *J. Bacteriol.*, 1998, **180**, 5426–5431.
38 P. Chain, J. Lamerdin, F. Larimer, W. Regala, V. Lao, M. Land, L. Hauser, A. Hooper, M. Klotz, J. Norton, L. Sayavedra-Soto, D. Arciero, N. Hommes, M. Whittaker and D. Arp, *J. Bacteriol.*, 2003, **185**, 2759–2773.
39 J. Parkhill, M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, A. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell and D. J. Maskell, *Nat. Genet.*, 2003, **35**, 32–40.
40 D. W. Wood, J. C. Setubal, R. Kaul, D. E. Monks, J. P. Kitajima, V. K. Okura, Y. Zhou, L. Chen, G. E. Wood, N. F. Almeida, Jr., L. Woo, Y. Chen, I. T. Paulsen, J. A. Eisen, P. D. Karp, D. Bovee, Sr., P. Chapman, J. Clendenning, G. Deatherage, W. Gillet, C. Grant, T. Kutyavin, R. Levy, M. J. Li, E. McClelland, A. Palmieri, C. Raymond, G. Rouse, C. Saenphimmachak, Z. Wu, P. Romero, D. Gordon, S. Zhang, H. Yoo, Y. Tao, P. Biddle, M. Jung, W. Krespan, M. Perry, B. Gordon-Kamm, L. Liao, S. Kim, C. Hendrick, Z. Y. Zhao, M. Dolan, F. Chumley, S. V. Tingey, J. F. Tomb, M. P. Gordon, M. V. Olson and E. W. Nester, *Science*, 2001, **294**, 2317–2323; B. Goodner, G. Hinkle, S. Gattung, N. Miller, M. Blanchard, B. Qurollo, B. S. Goldman, Y. Cao, M. Askenazi, C. Halling, L. Mullin, K. Houmiel, J. Gordon, M. Vaudin, O. Iartchouk, A. Epp, F. Liu, C. Wollam, M. Allinger, D. Doughty, C. Scott, C. Lappas, B. Markelz, C. Flanagan, C. Crowell, J. Gurson, C. Lomo, C. Sear, G. Strub, C. Cielo and S. Slater, *Science.*, 2001, **294**, 2323–2328.
41 K. A. Brayton, L. S. Kappmeyer, D. R. Herndon, M. J. Dark, D. L. Tibbals, G. H. Palmer, T. C. McGuire and D. P. Knowles, Jr., *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 844–849.
42 C. M. Alsmark, A. C. Frank, E. O. Karlberg, B. A. Legault, D. H. Ardell, B. Canback, A. S. Eriksson, A. K. Naslund, S. A. Handley, M. Huvet, B. La Scola, M. Holmberg and S. G. Andersson, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9716–9721.
43 T. Kaneko, Y. Nakamura, S. Sato, K. Minamisawa, Y. Uchiumi, S. Sasamoto, A. Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima, M. Kohara, M. Matsumoto, S. Shimpo, H. Tsuruoka, T. Wada, M. Yamada and S. Tabata, *DNA Res.*, 2002, **9**, 189–197.
44 S. M. Halling, B. D. Peterson-Burch, B. J. Bricker, R. L. Zuerner, Z. Qing, L.-L. Li, V. Kapur, D. P. Alt and S. C. Olsen, *J. Bacteriol.*, 2005, **187**, 2715–2726.
45 V. G. DelVecchio, V. Kapatral, R. J. Redkar, G. Patra, C. Mujer, T. Los, N. Ivanova, I. Anderson, A. Bhattacharyya, A. Lykidis, G. Reznik, L. Jablonski, N. Larsen, M. D'Souza, A. Bernal, M. Mazur, E. Goltsman, E. Selkov, P. H. Elzer, S. Hagius, D. O'Callaghan, J. J. Letesson, R. Haselkorn, N. Kyrpides and R. Overbeek, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 443–448.
46 P. S. Chain, D. J. Comerci, M. E. Tolmasky, F. W. Larimer, S. A. Malfatti, L. M. Vergez, F. Aguero, M. L. Land, R. A. Ugalde and E. Garcia, *Infect. Immun.*, 2005, **73**, 8353–8361.
47 I. T. Paulsen, R. Seshadri, K. E. Nelson, J. A. Eisen, J. F. Heidelberg, T. D. Read, R. J. Dodson, L. Umayam, L. M. Brinkac, M. J. Beanan, S. C. Daugherty, R. T. Deboy, A. S. Durkin, J. F. Kolonay, R. Madupu, W. C. Nelson, B. Ayodeji, M. Kraul, J. Shetty, J. Malek, S. E. Van Aken, S. Riedmuller, H. Tettelin, S. R. Gill, O. White, S. L. Salzberg, D. L. Hoover, L. E. Lindler, S. M. Halling, S. M. Boyle and C. M. Fraser, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 13148–13153.
48 W. C. Nierman, T. V. Feldblyum, M. T. Laub, I. T. Paulsen, K. E. Nelson, J. A. Eisen, J. F. Heidelberg, M. R. Alley, N. Ohta, J. R. Maddock, I. Potocka, W. C. Nelson, A. Newton, C. Stephens, N. D. Phadke, B. Ely, R. T. DeBoy, R. J. Dodson, A. S. Durkin, M. L. Gwinn, D. H. Haft, J. F. Kolonay, J. Smit, M. B. Craven, H. Khouri, J. Shetty, K. Berry, T. Utterback, K. Tran, A. Wolf, J. Vamathevan, M. Ermolaeva, O. White, S. L. Salzberg, J. C. Venter, L. Shapiro and C. M. Fraser, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4136–4141.
49 R. Frutos, A. Viari, C. Ferraz, A. Morgat, S. Eychenie, Y. Kandassamy, I. Chantal, A. Bensaid, E. Coissac, N. Vachiery, J. Demaille and D. Martinez, *J. Bacteriol.*, 2006, **188**, 2533–2542.
50 N. E. Collins, J. Liebenberg, E. P. de Villiers, K. A. Brayton, E. Louw, A. Pretorius, F. E. Faber, H. van Heerden, A. Josemans, M. van Kleef, H. C. Steyn, M. F. van Strijp, E. Zweygarth, F. Jongejan, J. C. Maillard,

D. Berthier, M. Botha, F. Joubert, C. H. Corton, N. R. Thomson, M. T. Allsopp and B. A. Allsopp, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 838–843.
51 C. Prust, M. Hoffmeister, H. Liesegang, A. Wiezer, W. F. Fricke, A. Ehrenreich, G. Gottschalk and U. Deppenmeier, *Nat. Biotechnol.*, 2005, **23**, 195–200.
52 T. Kaneko, Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada and S. Tabata, *DNA Res.*, 2000, **7**, 331–338.
53 S. J. Giovannoni, H. J. Tripp, S. Givan, M. Podar, K. L. Vergin, D. Baptista, L. Bibbs, J. Eads, T. H. Richardson, M. Noordewier, M. S. Rappe, J. M. Short, J. C. Carrington and E. J. Mathur, *Science.*, 2005, **309**, 1242–1245.
54 F. W. Larimer, P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M. L. Land, D. A. Pelletier, J. T. Beatty, A. S. Lang, F. R. Tabita, J. L. Gibson, T. E. Hanson, C. Bobst, J. L. Torres, C. Peres, F. H. Harrison, J. Gibson and C. S. Harwood, *Nat. Biotechnol.*, 2004, **22**, 55–61.
55 H. Ogata, S. Audic, P. Renesto-Audiffren, P. E. Fournier, V. Barbe, D. Samson, V. Roux, P. Cossart, J. Weissenbach, J. M. Claverie and D. Raoult, *Science*, 2001, **293**, 2093–2098.
56 H. Ogata, P. Renesto, S. Audic, C. Robert, G. Blanc, P. E. Fournier, H. Parinello, J. M. Claverie and D. Raoult, *PLoS Biol.*, 2005, **3**, e248.
57 S. G. Andersson, A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. Alsmark, R. M. Podowski, A. K. Naslund, A. S. Eriksson, H. H. Winkler and C. G. Kurland, *Nature*, 1998, **396**, 133–140.
58 M. P. McLeod, X. Qin, S. E. Karpathy, J. Gioia, S. K. Highlander, G. E. Fox, T. Z. McNeill, H. Jiang, D. Muzny, L. S. Jacob, A. C. Hawes, E. Sodergren, R. Gill, J. Hume, M. Morgan, G. Fan, A. G. Amin, R. A. Gibbs, C. Hong, X. J. Yu, D. H. Walker and G. M. Weinstock, *J. Bacteriol.*, 2004, **186**, 5842–5855.
59 M. A. Moran, A. Buchan, J. M. Gonzalez, J. F. Heidelberg, W. B. Whitman, R. P. Kiene, J. R. Henriksen, G. M. King, R. Belas, C. Fuqua, L. Brinkac, M. Lewis, S. Johri, B. Weaver, G. Pai, J. A. Eisen, E. Rahe, W. M. Sheldon, W. Ye, T. R. Miller, J. Carlton, D. A. Rasko, I. T. Paulsen, Q. Ren, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, M. J. Rosovitz, D. H. Haft, J. Selengut and N. Ward, *Nature*, 2004, **432**, 910–912.
60 F. Galibert, T. M. Finan, S. R. Long, A. Puhler, P. Abola, F. Ampe, F. Barloy-Hubler, M. J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R. W. Davis, S. Dreano, N. A. Federspiel, R. F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R. W. Hyman, T. Jones, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M. C. Peck, T. M. Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thebault, M. Vandenbol, F. J. Vorholter, S. Weidner, D. H. Wells, K. Wong, K. C. Yeh and J. Batut, *Science*, 2001, **293**, 668–672.
61 M. Wu, L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M. B. Young, T. Utterback, J. Weidman, W. C. Nierman, I. T. Paulsen, K. E. Nelson, H. Tettelin, S. L. O'Neill and J. A. Eisen, *PloS Biol.*, 2004, **2**, E69.
62 J. Foster, M. Ganatra, I. Kamal, J. Ware, K. Makarova, N. Ivanova, A. Bhattacharyya, V. Kapatral, S. Kumar, J. Posfai, T. Vincze, J. Ingram, L. Moran, A. Lapidus, M. Omelchenko, N. Kyrpides, E. Ghedin, S. Wang, E. Goltsman, V. Joukov, O. Ostrovskaya, K. Tsukerman, M. Mazur, D. Comb, E. Koonin and B. Slatko, *PloS Biol.*, 2005, **3**, E121.
63 J. S. Seo, H. Chong, H. S. Park, K. O. Yoon, C. Jung, J. J. Kim, J. H. Hong, H. Kim, J. H. Kim, J. I. Kil, C. J. Park, H. M. Oh, J. S. Lee, S. J. Jin, H. W. Um, H. J. Lee, S. J. Oh, J. Y. Kim, H. L. Kang, S. Y. Lee, K. J. Leeand and H. S. Kang, *Nat. Biotechnol.*, 2005, **23**, 63–68.
64 I. González Carreró, F. J. Sangari, J. Agüero and J. M. García Lobo, *Microbiology*, 2002, **148**, 353–360.
65 R. Rabus, M. Kube, J. Heider, A. Beck, K. Heitmann, F. Widdel and R. Reinhardt, *Arch. Microbiol.*, 2005, **183**, 27–36.
66 W. C. Nierman, D. DeShazer, H. S. Kim, H. Tettelin, K. E. Nelson, T. Feldblyum, R. L. Ulrich, C. M. Ronning, L. M. Brinkac, S. C. Daugherty, T. D. Davidsen, R. T. Deboy, G. Dimitrov, R. J. Dodson, A. S. Durkin, M. L. Gwinn, D. H. Haft, H. Khouri, J. F. Kolonay, R. Madupu, Y. Mohammoud, W. C. Nelson, D. Radune, C. M. Romero, S. Sarria, J. Selengut, C. Shamblin, S. A. Sullivan, O. White, Y. Yu, N. Zafar, L. Zhou and C. M. Fraser, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14246–14251.
67 H. Tettelin, N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, W. C. Nelson, M. L. Gwinn, R. DeBoy, J. D. Peterson, E. K. Hickey, D. H. Haft, S. L. Salzberg, O. White, R. D. Fleischmann, B. A. Dougherty, T. Mason, A. Ciecko, D. S. Parksey, E. Blair, H. Cittone, E. B. Clark, M. D. Cotton, T. R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H. O. Smith, C. M. Fraser, E. R. Moxon, R. Rappuoli and J. C. Venter, *Science*, 2000, **287**, 1809–1815.
68 J. Parkhill, M. Achtman, K. D. James, S. D. Bentley, C. Churcher, S. R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R. M. Davies, P. Davis, K. Devlin, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Leather, S. Moule, K. Mungall, M. A. Quail, M. A. Rajandream, K. M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B. G. Spratt and B. G. Barrell, *Nature*, 2000, **404**, 502–506.
69 M. Salanoubat, S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A. Billault, P. Brottier, J. C. Camus, L. Cattolico, M. Chandler, N. Choisne, C. Claudel-Renard, S. Cunnac, N. Demange, C. Gaspin, M. Lavie, A. Moisan, C. Robert, W. Saurin, T. Schiex, P. Siguier, P. Thébault, M. Whalen, P. Wincker, M. Levy, J. Weissenbach and C. A. Boucher, *Nature*, 2002, **415**, 497–502.
70 H. R. Beller, P. S. Chain, T. E. Letain, A. Chakicherla, F. W. Larimer, P. M. Richardson, M. A. Coleman, A. P. Wood and D. P. Kelly, *J. Bacteriol.*, 2006, **188**, 1473–1488.
71 V. Barbe, D. Vallenet, N. Fonknechten, A. Kreimeyer, S. Oztas, L. Labarre, S. Cruveiller, C. Robert, S. Duprat, P. Wincker, L. N. Ornston, J. Weissenbach, P. Marlière, G. N. Cohen and C. Médigue, *Nucleic Acids Res.*, 2004, **32**, 5766–5779.
72 R. Gil, F. J. Silva, E. Zientz, F. Delmotte, F. Gonzalez-Candelas, A. Latorre, C. Rausell, J. Kramerbeek, J. Gadau, B. Hoelldobler, R. C. H. J. van Ham, R. Gross and A. Moya, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 9388–9393.
73 P. H. Degnan, A. B. Lazarus and J. J. Wernegreen, *Genome Res.*, 2005, **15**, 1023–1033.
74 S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki and H. Ishikawa, *Nature*, 2000, **407**, 81–86.
75 R. C. Van Ham, J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernandez, L. Jimenez, M. Postigo, F. J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Moran and A. Moya, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 581–586.
76 I. Tamas, L. Klasson, B. Canback, A. K. Naslund, A. S. Eriksson, J. J. Wernegreen, J. P. Sandstrom, N. A. Moran and S. G. Andersson, *Science*, 2002, **296**, 2376–2379.
77 R. Seshadri, I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser and J. F. Heidelberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5455–5460.
78 F. R. Blattner, G. Plunkett, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao, *Science*, 1997, **277**, 1453–1462.
79 N. T. Perna, G. Plunkett, III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Pósfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch and F. R. Blattner, *Nature*, 2001, **409**, 529–535.
80 R. A. Welch, V. Burland, G. Plunkett, III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Donnenberg and F. R. Blattner, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 17020–17024.
81 T. Hayashi, K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C.-G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori and H. Shinagawa, *DNA Res.*, 2001, **8**, 11–22.
82 P. Larsson, P. C. Oyston, P. Chain, M. C. Chu, M. Duffield, H. H. Fuxelius, E. Garcia, G. Halltorp, D. Johansson, K. E. Isherwood, P. D.

Karp, E. Larsson, Y. Liu, S. Michell, J. Prior, R. Prior, S. Malfatti, A. Sjostedt, K. Svensson, N. Thompson, L. Vergez, J. K. Wagg, B. W. Wren, L. E. Lindler, S. G. Andersson, M. Forsman and R. W. Titball, *Nat. Genet.*, 2005, **37**, 153–159.
83 A. Harrison, D. W. Dyer, A. Gillaspy, W. C. Ray, R. Mungur, M. B. Carson, H. Zhong, J. Gipson, M. Gipson, L. S. Johnson, L. Lewis, L. O. Bakaletz and R. S. Munson, Jr., *J. Bacteriol.*, 2005, **187**, 4627–4636.
84 R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. G. Sutton, W. FitzHugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L. I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith and J. C. Venter, *Science*, 1995, **269**, 496–512.
85 S. Hou, J. H. Saw, K. S. Lee, T. A. Freitas, C. Belisle, Y. Kawarabayasi, S. P. Donachie, A. Pikina, M. Y. Galperin, E. V. Koonin, K. S. Makarova, M. V. Omelchenko, A. Sorokin, Y. I. Wolf, Q. X. Li, Y. S. Keum, S. Campbell, J. Denery, S. Aizawa, S. Shibata, A. Malahoff and M. Alam, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 18036–18041.
86 C. Cazalet, C. Rusniok, H. Bruggemann, N. Zidane, A. Magnier, L. Ma, M. Tichit, S. Jarraud, C. Bouchier, F. Vandenesch, F. Kunst, J. Etienne, P. Glaser and C. Buchrieser, *Nat. Genet.*, 2004, **36**, 1165–1173.
87 M. Chien, I. Morozova, S. Shi, H. Sheng, J. Chen, S. M. Gomez, G. Asamani, K. Hill, J. Nuara, M. Feder, J. Rineer, J. J. Greenberg, V. Steshenko, S. H. Park, B. Zhao, E. Teplitskaya, J. R. Edwards, S. Pampou, A. Georghiou, I.-C. Chou, W. Iannuccilli, M. E. Ulz, D. H. Kim, A. Geringer-Sameth, C. Goldsberry, P. Morozov, S. G. Fischer, G. Segal, X. Qu, A. Rzhetsky, P. Zhang, E. Cayanis, P. J. De Jong, J. Ju, S. Kalachikov, H. A. Shuman and J. J. Russo, *Science*, 2004, **305**, 1966–1968.
88 S. H. Hong, J. S. Kim, S. Y. Lee, Y. H. In, S. S. Choi, J. K. Rih, C. H. Kim, H. Jeong, C. G. Hur and J. J. Kim, *Nat. Biotechnol.*, 2004, **22**, 1275–1281.
89 N. Ward, Ø. Larsen, J. Sakwa, L. Bruseth, H. Khouri, A. S. Durkin, G. Dimitrov, L. Jiang, D. Scanlan, K. H. Kang, M. Lewis, K. E. Nelson, B. Methé, M. Wu, J. F. Heidelberg, I. T. Paulsen, D. Fouts, J. Ravel, H. Tettelin, Q. Ren, T. Read, R. T. DeBoy, R. Seshadri, S. L. Salzberg, H. B. Jensen, N. K. Birkeland, W. C. Nelson, R. J. Dodson, S. H. Grindhaug, I. Holt, I. Eidhammer, I. Jonasen, S. Vanaken, T. Utterback, T. V. Feldblyum, C. M. Fraser, I. R. Lillehaug and J. A. Eisen, *PLoS Biol.*, 2004, **2**, 1616–1628.
90 B. J. May, Q. Zhang, L. L. Li, M. L. Paustian, T. S. Whittam and V. Kapur, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 3460–3465.
91 E. Duchaud, C. Rusniok, L. Frangeul, C. Buchrieser, A. Givaudan, S. Taourit, S. Bocs, C. Boursaux-Eude, M. Chandler, J. F. Charles, E. Dassa, R. Derose, S. Derzelle, G. Freyssinet, S. Gaudriault, C. Medigue, A. Lanois, K. Powell, P. Siguier, R. Vincent, V. Wingate, M. Zouine, P. Glaser, N. Boemare, A. Danchin and F. Kunst, *Nat. Biotechnol.*, 2003, **21**, 1307–1313.
92 K. E. Nelson, C. Weinel, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A. P. Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanana, R. T. DeBoy, S. Daugherty, J. Kolonay, R. Maduou, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. Chris Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzez, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Welder, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. Eisen, K. N. Timmis, A. Düsterhöft, B. Tümmler and C. M. Fraser, *Environ. Microbiol.*, 2002, **4**, 799–808.
93 H. Feil, W. S. Feil, P. Chain, F. Larimer, G. DiBartolo, A. Copeland, A. Lykidis, S. Trong, M. Nolan, E. Goltsman, J. Thiel, S. Malfatti, J. E. Loper, A. Lapidus, J. C. Detter, M. Land, P. M. Richardson, N. C. Kyrpides, N. Ivanova and W. E. Lindow, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 11064–11069.
94 V. Joardar, M. Lindeberg, R. W. Jackson, J. Selengut, R. Dodson, L. M. Brinkac, S. C. Daugherty, R. DeBoy, A. S. Durkin, M. G. Giglio, R. Madupu, W. C. Nelson, M. J. Rosovitz, S. Sullivan, J. Crabtree, T. Creasy, T. Davidsen, D. H. Haft, N. Zafar, L. Zhou, R. Halpin, T. Holley, H. Khouri, T. Feldblyum, O. White, C. M. Fraser, A. K. Chatterjee, S. Cartinhour, D. J. Schneider, J. Mansfield, A. Collmer and C. R. Buell, *J. Bacteriol.*, 2005, **187**, 6488–6498.
95 C. R. Buell, V. Joardar, M. Lindeberg, J. Selengut, I. T. Paulsen, M. L. Gwinn, R. J. Dodson, R. T. Deboy, A. S. Durkin, J. F. Kolonay, R. Madupu, S. Daugherty, L. Brinkac, M. J. Beanan, D. H. Haft, W. C. Nelson, T. Davidsen, N. Zafar, L. Zhou, J. Liu, Q. Yuan, H. Khouri, N. Fedorova, B. Tran, D. Russell, K. Berry, T. Utterback, S. E. Van Aken, T. V. Feldblyum, M. D'Ascenzo, W. L. Deng, A. R. Ramos, J. R. Alfano, S. Cartinhour, A. K. Chatterjee, T. P. Delaney, S. G. Lazarowitz, G. B. Martin, D. J. Schneider, X. Tang, C. L. Bender, O. White, C. M. Fraser and A. Collmer, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 10181–10186.
96 C.-H. Chiu, P. Tang, C. Chu, S. Hu, Q. Bao, J. Yu, Y.-Y. Chou, H.-S. Wang and Y.-S. Lee, *Nucleic Acids Res.*, 2005, **33**, 1690–1698.
97 M. McClelland, K. E. Sanderson, S. W. Clifton, P. Latreille, S. Porwollik, A. Sabo, R. Meyer, T. Bieri, P. Ozersky, M. McLellan, C. R. Harkins, C. Wang, C. Nguyen, A. Berghoff, G. Elliott, S. Kohlberg, C. Strong, F. Du, J. Carter, C. Kremizki, D. Layman, S. Leonard, H. Sun, L. Fulton, W. Nash, T. Miner, P. Minx, K. Delehaunty, C. Fronick, V. Magrini, M. Nhan, W. Warren, L. Florea, J. Spieth and R. K. Wilson, *Nat. Genet.*, 2004, **36**, 1268–1274.
98 J. Parkhill, G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. G. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead and B. G. Barrell, *Nature*, 2001, **413**, 848–852.
99 W. Deng, S.-R. Liou, G. Plunkett, III, G. F. Mayhew, D. J. Rose, V. Burland, V. Kodoyianni, D. C. Schwartz and F. R. Blattner, *J. Bacteriol.*, 2003, **185**, 2330–2337.
100 M. McClelland, K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston and R. K. Wilson, *Nature*, 2001, **413**, 852–856.
101 J. F. Heidelberg, I. T. Paulsen, K. E. Nelson, E. J. Gaidos, W. C. Nelson, T. D. Read, J. A. Eisen, R. Seshadri, N. Ward, B. Methe, R. A. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. T. DeBoy, R. J. Dodson, A. S. Durkin, D. H. Haft, J. F. Kolonay, R. Madupu, J. D. Peterson, L. A. Umayam, O. White, A. M. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. R. Utterback, L. A. McDonald, T. V. Feldblyum, H. O. Smith, J. C. Venter, K. H. Nealson and C. M. Fraser, *Nat. Biotechnol.*, 2002, **20**, 1118–1123.
102 F. Yang, J. Yang, X. Zhang, L. Chen, Y. Jiang, Y. Yan, X. Tang, J. Wang, Z. Xiong, J. Dong, Y. Xue, Y. Zhu, X. Xu, L. Sun, S. Chen, H. Nie, J. Peng, J. Xu, Y. Wang, Z. Yuan, Y. Wen, Z. Yao, Y. Shen, B. Qiang, Y. Hou, J. Yu and Q. Jin, *Nucleic Acids Res.*, 2005, **33**, 6445–6458.
103 J. Wei, M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett, III, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz and F. R. Blattner, *Infect. Immun.*, 2003, **71**, 2775–2786.
104 Q. Jin, Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, L. Sun, Y. Xue, A. Zhao, Y. Gao, J. Zhu, B. Kan, K. Ding, S. Chen, H. Cheng, Z. Yao, B. He, R. Chen, D. Ma, B. Qiang, Y. Wen, Y. Hou and J. Yu, *Nucleic Acids Res.*, 2002, **30**, 4432–4441.
105 J. F. Heidelberg, J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter and C. M. Fraser, *Nature*, 2000, **406**, 477–483.
106 E. G. Ruby, M. Urbanowski, J. Campbell, A. Dunn, M. Faini, R. Gunsalus, P. Lostroh, C. Lupp, J. McCann, D. Millikan, A. Schaefer, E. Stabb, A. Stevens, K. Visick, C. Whistler and E. P. Greenberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 3004–3009.
107 K. Makino, K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, Y. Kubota, S. Kimura, T. Yasunaga, T. Honda, H. Shinagawa, M. Hattori and T. Iida, *Lancet*, 2003, **361**, 743–749.
108 C.-Y. Chen, K.-M. Wu, Y.-C. Chang, C.-H. Chang, H.-C. Tsai, T.-L. Liao, Y.-M. Liu, H.-J. Chen, A. B.-T. Shen, J.-C. Li, T.-L. Su,

C.-P. Shao, C.-T. Lee, L. I. Hor and S.-F. Tsai, *Genome Res.*, 2003, **13**, 2577–2587.
109 L. Akman, A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori and S. Aksoy, *Nat. Genet.*, 2002, **32**, 402–407.
110 A. C. R. da Silva, J. A. Ferro, F. C. Reinach, C. S. Farah, L. R. Furlan, R. B. Quaggio, C. B. Monteiro-Vitorello, M. A. Van Sluys, N. F. Almeida, L. M. C. Alves, A. M. do Amaral, M. C. Bertolini, L. E. A. Camargo, G. Camarotte, F. Cannavan, J. Cardozo, F. Chambergo, L. P. Ciapina, R. M. B. Cicarelli, L. L. Coutinho, J. R. Cursino-Santos, H. El-Dorry, J. B. Faria, A. J. S. Ferreira, R. C. C. Ferreira, M. I. T. Ferro, E. F. Formighieri, M. C. Franco, C. C. Greggio, A. Gruber, A. M. Katsuyama, L. T. Kishi, R. P. Leite, E. G. M. Lemos, M. V. F. Lemos, E. C. Locali, M. A. Machado, A. M. B. N. Madeira, N. M. Martinez-Rossi, E. C. Martins, J. Meidanis, C. F. M. Menck, C. Y. Miyaki, D. H. Moon, L. M. Moreira, M. T. M. Novo, V. K. Okura, M. C. Oliveira, V. R. Oliveira, H. A. Pereira, A. Rossi, J. A. D. Sena, C. Silva, R. F. de Souza, L. A. F. Spinola, M. A. Takita, R. E. Tamura, E. C. Teixeira, R. I. D. Tezza, M. Trindade dos Santos, D. Truffi, S. M. Tsai, F. F. White, J. C. Setubal and J. P. Kitajima, *Nature*, 2002, **417**, 459–463.
111 W. Qian, Y. Jia, S.-X. Ren, Y.-Q. He, J.-X. Feng, L.-F. Lu, Q. Sun, G. Ying, D.-J. Tang, H. Tang, W. Wu, P. Hao, L. Wang, B.-L. Jiang, S. Zeng, W.-Y. Gu, G. Lu, L. Rong, Y. Tian, Z. Yao, G. Fu, B. Chen, R. Fang, B. Qiang, Z. Chen, G.-P. Zhao, J.-L. Tang and C. He, *Genome Res.*, 2005, **15**, 757–767.
112 F. Thieme, R. Koebnik, T. Bekel, C. Berger, J. Boch, D. Büttner, C. Caldana, L. Gaigalat, A. Goesmann, S. Kay, O. Kirchner, C. Lanz, B. Linke, A. C. McHardy, F. Meyer, G. Mittenhuber, D. H. Nies, U. Niesbach-Klösgen, T. Patschkowski, C. Rückert, O. Rupp, S. Schneiker, S. C. Schuster, F.-J. Vorhölter, E. Weber, A. Pühler, U. Bonas, D. Bartels and O. Kaiser, *J. Bacteriol.*, 2005, **197**, 7254–7266.
113 B.-M. Lee, Y.-J. Park, D.-S. Park, H.-W. Kang, J.-G. Kim, E.-S. Song, I.-C. Park, U.-H. Yoon, J.-H. Hahn, B.-S. Koo, G.-B. Lee, H. Kim, H.-S. Park, K.-O. Yoon, J.-H. Kim, C.-H. Jung, N.-H. Koh, J.-S. Seo and S.-J. Go, *Nucleic Acids Res.*, 2005, **33**, 577–586.
114 A. J. Simpson, F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bove, M. R. Briones, M. R. Bueno, A. A. Camargo, L. E. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El-Dorry, A. P. Facincani, A. J. Ferreira, V. C. Ferreira, J. A. Ferro, J. S. Fraga, S. C. Franca, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C. Leite, E. G. Lemos, M. V. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M. Madeira, H. M. Madeira, C. L. Marino, M. V. Marques, E. A. Martins, E. M. Martins, A. Y. Matsukuma, C. F. Menck, E. C. Miracca, C. Y. Miyaki, C. B. Monteriro-Vitorello, D. H. Moon, M. A. Nagai, A. L. Nascimento, L. E. Netto, A. Nhani, Jr., F. G. Nobrega, L. R. Nunes, M. A. Oliveira, M. C. de Oliveira, R. C. de Oliveira, D. A. Palmieri, A. Paris, B. R. Peixoto, G. A. Pereira, H. A. Pereira, Jr., J. B. Pesquero, R. B. Quaggio, P. G. Roberto, V. Rodrigues, A. J. de M. Rosa, V. E. de Rosa, Jr., R. G. de Sa, R. V. Santelli, H. E. Sawasaki, A. C. da Silva, A. M. da Silva, F. R. da Silva, W. A. da Silva, Jr., J. F. da Silveira, M. L. Silvestri, W. J. Siqueira, A. A. de Souza, A. P. de Souza, M. F. Terenzi, D. Truffi, S. M. Tsai, M. H. Tsuhako, H. Vallada, M. A. Van Sluys, S. Verjovski-Almeida, A. L. Vettore, M. A. Zago, M. Zatz, J. Meidanis and J. C. Setubal, *Nature*, 2000, **406**, 151–157.
115 M. A. Van Sluys, M. C. de Oliveira, C. B. Monteiro-Vitorello, C. Y. Miyaki, L. R. Furlan, L. E. Camargo, A. C. da Silva, D. H. Moon, M. A. Takita, E. G. Lemos, M. A. Machado, M. I. Ferro, F. R. da Silva, M. H. Goldman, G. H. Goldman, M. V. Lemos, H. El-Dorry, S. M. Tsai, H. Carrer, D. M. Carraro, R. C. de Oliveira, L. R. Nunes, W. J. Siqueira, L. L. Coutinho, E. T. Kimura, E. S. Ferro, R. Harakava, E. E. Kuramae, C. L. Marino, E. Giglioti, I. L. Abreu, L. M. Alves, A. M. do Amaral, G. S. Baia, S. R. Blanco, M. S. Brito, F. S. Cannavan, A. V. Celestino, A. F. da Cunha, R. C. Fenille, J. A. Ferro, E. F. Formighieri, L. T. Kishi, S. G. Leoni, A. R. Oliveira, V. E. Rosa, Jr., F. T. Sassaki, J. A. Sena, A. A. de Souza, D. Truffi, F. Tsukumo, G. M. Yanai, L. G. Zaros, E. L. Civerolo, A. J. Simpson, N. F. Almeida, Jr., J. C. Setubal and J. P. Kitajima, *J. Bacteriol.*, 2003, **185**, 1018–1026.
116 J. Parkhill, B. W. Wren, N. R. Thomson, R. W. Titbal, M. T. G. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeño-Tárraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. F. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead and B. G. Barrell, *Nature*, 2003, **413**, 523–527.
117 W. Deng, V. Burland, G. Plunkett, III, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna, D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L. Nilles, J. S. Matson, F. R. Blattner and R. D. Perry, *J. Bacteriol.*, 2002, **184**, 4601–4611; Y. Song, Z. Tong, J. Wang, L. Wang, Z. Guo, Y. Han, J. Zhang, D. Pei, D. Zhou, H. Qin, X. Pang, Y. Han, J. Zhai, M. Li, B. Cui, Z. Qi, L. Jin, R. Dai, F. Chen, S. Li, C. Ye, X. Du, W. Lin, J. Wang, J. Yu, H. Yang, J. Wang, P. Huang and R. Yang, *DNA Res.*, 2004, **11**, 179–197.
118 P. S. G. Chain, E. Carniel, F. W. Larimer, J. Lamerdin, P. O. Stoutland, W. M. Regala, A. M. Georgescu, L. M. Vergez, M. L. Land, V. L. Motin, R. R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V. Chenal-Francisque, B. Souza, D. Dacheux, J. M. Elliott, A. Derbise, L. J. Hauser and E. Garcia, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 13826–13831.
119 T. A. Ciche, M. Blackburn, J. R. Carney and J. C. Ensign, *Appl. Environ. Microbiol.*, 2003, **69**, 4706–4713.
120 J. M. Meyer, *Arch. Microbiol.*, 2000, **174**, 135–142.
121 J. Ravel and P. Cornelis, *Trends Microbiol.*, 2003, **11**, 195–200.
122 C. R. Howell and R. D. Stipanovic, *Phytopathology*, 1980, **70**, 712–715.
123 (*a*) B. Nowak-Thompson, N. Chaney, J. S. Wing, S. J. Gould and J. E. Loper, *J. Bacteriol.*, 1999, **181**, 2166–2174; (*b*) H. Gross, V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper and W. H. Gerwick, *Chem. Biol.*, 2007, **14**, 1–11.
124 C. L. Bender, F. Alarcon-Chaidez and D. C. Gross, *Microbiol. Mol. Biol. Rev.*, 1999, **63**, 266–292.
125 I. Grgurina, F. Mariotti, V. Fogliano, M. Gallo, A. Scaloni, N. S. Iacobellis, P. Lo Cantore, L. Mannina, V. van Axel Castelli, M. L. Greco and A. Graniti, *Biochim. Biophys. Acta*, 2002, **1597**, 81–89.
126 A. Bultreys, I. Gheysen and E. de Hoffmann, *Appl. Environ. Microbiol.*, 2006, **72**, 3814–3825.
127 L. Gardan, C. Gouy, R. Christen and R. Samson, *Int. J. Syst. Evol. Microbiol.*, 2003, **53**, 381–391.
128 www.npbiogene.com.
129 R. H. Ffrench-Constant, N. Waterfield, V. Burland, N. T. Perna, P. J. Daborn, D. Bowen and F. R. Blattner, *Appl. Environ. Microbiol.*, 2000, **66**, 3310–3329.
130 M. Achtman, K. Zurth, G. Morelli, G. Torrea, A. Guiyoule and E. Carniel, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14043–14048.
131 A. A. DiSpirito, J. A. Zahn, D. W. Graham, H. J. Kim, C. K. Larive, T. S. Derrick, C. D. Cox and A. Taylor, *J. Bacteriol.*, 1998, **180**, 3606–3613; C. M. Tellez, K. P. Gaus, D. W. Graham, R. G. Arnold and R. Z. Guzman, *Appl. Environ. Microbiol.*, 1998, **64**, 1115–1122.
132 C. M. Litwin, T. W. Rayback and J. Skinner, *Infect. Immun.*, 1996, **64**, 2834–2838.
133 S. Rendulic, P. Jagtap, A. Rosinus, M. Eppinger, C. Baar, C. Lanz, H. Keller, C. Lambert, K. J. Evans, A. Goesmann, F. Meyer, R. E. Sockett and S. C. Schuster, *Science*, 2004, **303**, 689–692.
134 R. Rabus, A. Ruepp, T. Frickey, T. Rattei, B. Fartmann, M. Stark, M. Bauer, A. Zibat, T. Lombardot, I. Becker, J. Amann, K. Gellner, H. Teeling, W. D. Leuschner, F. O. Glockner, A. N. Lupas, R. Amann and H. P. Klenk, *Environ. Microbiol.*, 2004, **6**, 887–902.
135 J. F. Heidelberg, R. Seshadri, S. A. Haveman, C. L. Hemme, I. T. Paulsen, J. F. Kolonay, J. A. Eisen, N. Ward, B. Methe, L. M. Brinkac, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, D. Fouts, D. H. Haft, J. Selengut, J. D. Peterson, T. M. Davidsen, N. Zafar, L. Zhou, D. Radune, G. Dimitrov, M. Hance, K. Tran, H. Khouri, J. Gill, T. R. Utterback, T. V. Feldblyum, J. D. Wall, G. Voordouw and C. M. Fraser, *Nat. Biotechnol.*, 2004, **22**, 554–559.
136 B. A. Methe, K. E. Nelson, J. A. Eisen, I. T. Paulsen, W. Nelson, J. F. Heidelberg, D. Wu, M. Wu, N. Ward, M. J. Beanan, R. J. Dodson, R. Madupu, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, M. Gwinn, J. F. Kolonay, S. A. Sullivan, D. H. Haft, J. Selengut, T. M. Davidsen, N. Zafar, O. White, B. Tran, C. Romero, H. A. Forberger, J. Weidman, H. Khouri, T. V. Feldblyum, T. R. Utterback, S. E. Van

Aken, D. R. Lovley and C. M. Fraser, *Science*, 2003, **302**, 1967–1969.
137 J. Parkhill, B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M. A. Rajandream, K. M. Rutherford, A. H. van Vliet, S. Whitehead and B. G. Barrell, *Nature*, 2000, **403**, 665–668.
138 D. E. Fouts, E. F. Mongodin, R. E. Mandrell, W. G. Miller, D. A. Rasko, J. Ravel, L. M. Brinkac, R. T. DeBoy, C. T. Parker, S. C. Daugherty, R. J. Dodson, A. S. Durkin, R. Madupu, S. A. Sullivan, J. U. Shetty, M. A. Ayodeji, A. Shvartsbeyn, M. C. Schatz, J. H. Badger, C. M. Fraser and K. E. Nelson, *PloS Biol.*, 2005, **3**, E15.
139 S. Suerbaum, C. Josenhans, T. Sterzenbach, B. Drescher, P. Brandt, M. Bell, M. Droge, B. Fartmann, H. P. Fischer, Z. Ge, A. Horster, R. Holland, K. Klein, J. Konig, L. Macko, G. L. Mendz, G. Nyakatura, D. B. Schauer, Z. Shen, J. Weber, M. Frosch and J. G. Fox, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 7901–7906.
140 J. F. Tomb, O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser and J. C. Venter, *Nature*, 1997, **388**, 539–547.
141 R. A. Alm, L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. de Jonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis and T. J. Trust, *Nature*, 1999, **397**, 176–180.
142 C. Baar, M. Eppinger, G. Raddatz, J. Simon, C. Lanz, O. Klimmek, R. Nandakumar, R. Gross, A. Rosinus, H. Keller, P. Jagtap, B. Linke, F. Meyer, H. Lederer and S. C. Schuster, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 11690–11695.
143 T. D. Read, S. N. Peterson, N. Tourasse, L. W. Baillie, I. T. Paulsen, K. E. Nelson, H. Tettelin, D. E. Fouts, J. A. Eisen, S. R. Gill, E. K. Holtzapple, O. A. Okstad, E. Helgason, J. Rilstone, M. Wu, J. F. Kolonay, M. J. Beanan, R. J. Dodson, L. M. Brinkac, M. Gwinn, R. T. DeBoy, R. Madpu, S. C. Daugherty, A. S. Durkin, D. H. Haft, W. C. Nelson, J. D. Peterson, M. Pop, H. M. Khouri, D. Radune, J. L. Benton, Y. Mahamoud, L. Jiang, I. R. Hance, J. F. Weidman, K. J. Berry, R. D. Plaut, A. M. Wolf, K. L. Watkins, W. C. Nierman, A. Hazen, R. Cline, C. Redmond, J. E. Thwaite, O. White, S. L. Salzberg, B. Thomason, A. M. Friedlander, T. M. Koehler, P. C. Hanna, A. B. Kolsto and C. M. Fraser, *Nature*, 2003, **423**, 81–86.
144 D. A. Rasko, M. R. Altherr, C. S. Han and J. Ravel, *FEMS Microbiol. Rev.*, 2005, **29**, 303–329.
145 D. A. Rasko, J. Ravel, O. A. Økstad, E. Helgason, R. Z. Cer, L. Jiang, K. A. Shores, D. E. Fouts, N. J. Tourasse, S. V. Angiuoli, J. Kolonay, W. C. Nelson, A.-B. Kolstø, C. M. Fraser and T. D. Read, *Nucleic Acids Res.*, 2004, **32**, 977–988.
146 N. Ivanova, A. Sorokin, I. Anderson, N. Galleron, B. Candelon, V. Kapatral, A. Bhattacharyya, G. Reznik, N. Mikhailova, A. Lapidus, L. Chu, M. Mazur, E. Goltsman, N. Larsen, M. D'Souza, T. Walunas, Y. Grechkin, G. Pusch, R. Haselkorn, M. Fonstein, S. D. Ehrlich, R. Overbeek and N. Kyrpides, *Nature*, 2003, **423**, 87–91.
147 H. Takami, K. Nakasone, Y. Takaki, G. Maeno, R. Sasaki, N. Masui, F. Fuji, C. Hirama, Y. Nakamura, N. Ogasawara, S. Kuhara and K. Horikoshi, *Nucleic Acids Res.*, 2000, **28**, 4317–4331.
148 M. W. Rey, P. Ramaiya, B. A. Nelson, S. D. Brody-Karpin, E. J. Zaretsky, M. Tang, A. Lopez de Leon, H. Xiang, V. Gusti, I. G. Clausen, P. B. Olsen, M. D. Rasmussen, J. T. Andersen, P. L. Jorgensen, T. S. Larsen, A. Sorokin, A. Bolotin, A. Lapidus, N. Galleron, S. D. Ehrlich and R. M. Berka, *Genome Biol.*, 2004, **5**, R77; B. Veith, C. Herzberg, S. Steckel, J. Feesche, K. H. Maurer, P. Ehrenreich, S. Bäumer, A. Henne, H. Liesegang, R. Merkl, A. Ehrenreich and G. Gottschalk, *J. Mol. Microbiol. Biotechnol.*, 2004, **7**, 204–211.
149 F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton and A. Danchin, *et al.*, *Nature*, 1997, **390**, 249–256.
150 N. Shankar, A. S. Baghdayan and M. S. Gilmore, *Nature*, 2002, **417**, 746–750.
151 H. Takami, Y. Takaki, G.-J. Chee, S. Nishi, S. Shimamura, H. Suzuki, S. Matsui and I. Uchiyama, *Nucleic Acids Res.*, 2004, **32**, 6292–6303.
152 E. Altermann, W. M. Russell, M. A. Azcarate-Peril, R. Barrangou, B. L. Buck, O. McAuliffe, N. Souther, A. Dobson, T. Duong, M. Callanan, S. Lick, A. Hamrick, R. Cano and T. R. Klaenhammer, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 3906–3912.
153 R. D. Pridmore, B. Berger, F. Desiere, D. Vilanova, C. Barretto, A. C. Pittet, M. C. Zwahlen, M. Rouvet, E. Altermann, R. Barrangou, B. Mollet, A. Mercenier, T. Klaenhammer, F. Arigoni and M. A. Schell, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 2512–2517.
154 M. Kleerebezem, J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos and R. J. Siezen, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1990–1995.
155 S. Chaillou, M. C. Champomier-Verges, M. Cornet, A. M. Crutz-Le Coq, A. M. Dudez, V. Martin, S. Beaufils, E. Darbon-Rongere, R. Bossy, V. Loux and M. Zagorec, *Nat. Biotechnol.*, 2005, **23**, 1527–1533.
156 A. Bolotin, P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Erlich and A. Sorokin, *Genome Res.*, 2001, **11**, 731–753.
157 P. Glaser, L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couve, A. de Daruvar, P. Dehoux, E. Domann, G. Dominguez-Bernal, E. Duchaud, L. Durant, O. Dussurget, K. D. Entian, H. Fsihi, F. Garcia-del Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gomez-Lopez, T. Hain, J. Hauf, D. Jackson, L. M. Jones, U. Kaerst, J. Kreft, M. Kuhn, F. Kunst, G. Kurapkat, E. Madueno, A. Maitournam, J. M. Vicente, E. Ng, H. Nedjari, G. Nordsiek, S. Novella, B. de Pablos, J. C. Perez-Diaz, R. Purcell, B. Remmel, M. Rose, T. Schlueter, N. Simoes, A. Tierrez, J. A. Vazquez-Boland, H. Voss, J. Wehland and P. Cossart, *Science*, 2001, **294**, 849–852.
158 E. Nelson, D. E. Fouts, E. F. Mongodin, J. Ravel, R. T. DeBoy, J. F. Kolonay, D. A. Rasko, S. V. Angiuoli, S. R. Gill, I. T. Paulsen, J. Peterson, O. White, W. C. Nelson, W. Nierman, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. J. Dodson, A. S. Durkin, R. Madupu, D. H. Haft, J. Selengut, S. Van Aken, H. Khouri, N. Fedorova, H. Forberger, B. Tran, S. Kathariou, L. D. Wonderling, G. A. Uhlich, D. O. Bayles, J. B. Luchansky and C. M. Fraser, *Nucleic Acids Res.*, 2004, **32**, 2386–2395.
159 H. Takami, Y. Takaki and I. Uchiyama, *Nucleic Acids Res.*, 2002, **30**, 3927–3935.
160 S. R. Gill, D. E. Fouts, G. L. Archer, E. F. Mongodin, R. T. Deboy, J. Ravel, I. T. Paulsen, J. F. Kolonay, L. Brinkac, M. Beanan, R. J. Dodson, S. C. Daugherty, R. Madupu, S. V. Angiuoli, A. S. Durkin, D. H. Haft, J. Vamathevan, H. Khouri, T. Utterback, C. Lee, G. Dimitrov, L. Jiang, H. Qin, J. Weidman, K. Tran, K. Kang, I. R. Hance, K. E. Nelson and C. M. Fraser, *J. Bacteriol.*, 2005, **187**, 2426–2438.
161 M. T. Holden, E. J. Feil, J. A. Lindsay, S. J. Peacock, N. P. Day, M. C. Enright, T. J. Foster, C. E. Moore, L. Hurst, R. Atkin, A. Barron, N. Bason, S. D. Bentley, C. Chillingworth, T. Chillingworth, C. Churcher, L. Clark, C. Corton, A. Cronin, J. Doggett, L. Dowd, T. Feltwell, Z. Hance, B. Harris, H. Hauser, S. Holroyd, K. Jagels, K. D. James, N. Lennard, A. Line, R. Mayes, S. Moule, K. Mungall, D. Ormond, M. A. Quail, E. Rabbinowitsch, K. Rutherford, M. Sanders, S. Sharp, M. Simmonds, K. Stevens, S. Whitehead, B. G. Barrell, B. G. Spratt and J. Parkhill, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9786–9791.
162 T. Baba, F. Takeuchi, M. Kuroda, H. Yuzawa, K. Aoki, A. Oguchi, Y. Nagai, N. Iwama, K. Asano, T. Naimi, H. Kuroda, L. Cui, K. Yamamoto and K. Hiramatsu, *Lancet*, 2002, **359**, 1819–1827.
163 M. Kuroda, T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi and K. Hiramatsu, *Lancet*, 2001, **357**, 1225–1240.

164 Y. Q. Zhang, S. X. Ren, H. L. Li, Y. X. Wang, G. Fu, J. Yang, Z. Q. Qin, Y. G. Miao, W. Y. Wang, R. S. Chen, Y. Shen, Z. Chen, Z. H. Yuan, G. P. Zhao, D. Qu, A. Danchin and Y. M. Wen, *Mol. Microbiol.*, 2003, **49**, 1577–1593.

165 F. Takeuchi, S. Watanabe, T. Baba, H. Yuzawa, T. Ito, Y. Morimoto, M. Kuroda, L. Cui, M. Takahashi, A. Ankai, S. Baba, S. Fukui, J. C. Lee and K. Hiramatsu, *J. Bacteriol.*, 2005, **187**, 7292–7308.

166 M. Kuroda, A. Yamashita, H. Hirakawa, M. Kumano, K. Morikawa, M. Higashide, A. Maruyama, Y. Inose, K. Matoba, H. Toh, S. Kuhara, M. Hattori and T. Ohta, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13272–13277.

167 H. Tettelin, V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi and C. M. Fraser, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12391–12396.

168 H. Tettelin, V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. Deboy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarity Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli and C. M. Fraser, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13950–13955.

169 P. Glaser, C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couve, L. Lalioui, C. Poyart, P. Trieu-Cuot and F. Kunst, *Mol. Microbiol.*, 2002, **45**, 1499–1513.

170 D. Ajdic, W. M. McShan, R. E. McLaughlin, G. Savic, J. Chang, M. B. Carson, C. Primeaux, R. Tian, S. Kenton, H. Jia, S. Lin, Y. Qian, S. Li, H. Zhu, F. Najar, H. Lai, J. White, B. A. Roe and J. J. Ferretti, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14434–14439.

171 J. Hoskins, W. E. Alborn, Jr., J. Arnold, L. C. Blaszczak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D. J. Fu, W. Fuller, C. Geringer, R. Gilmour, J. S. Glass, H. Khoja, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenney, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P. M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. Zhao, C. A. Zook, R. H. Baltz, S. R. Jaskunas, P. R. Rosteck, Jr., P. L. Skatrud and J. I. Glass, *J. Bacteriol.*, 2001, **183**, 5709–5717.

172 H. Tettelin, K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead and C. M. Fraser, *Science*, 2001, **293**, 498–506.

173 J. J. Ferretti, W. M. McShan, D. Ajdic, D. J. Savic, G. Savic, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najar, Q. Ren, H. Zhu, L. Song, J. White, X. Yuan, S. W. Clifton, B. A. Roe and R. McLaughlin, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4658–4663.

174 P. Sumby, S. F. Porcella, A. G. Madrigal, K. D. Barbian, K. Virtaneva, S. M. Ricklefs, D. E. Sturdevant, M. R. Graham, J. Vuopio-Varkila, N. P. Hoe and J. M. Musser, *J. Infect. Dis.*, 2005, **192**, 771–782.

175 M. Green, S. Zhang, S. F. Porcella, M. J. Nagiec, K. D. Barbian, S. B. Beres, R. B. LeFebvre and J. M. Musser, *J. Infect. Dis.*, 2005, **192**, 760–770.

176 D. J. Banks, S. F. Porcella, K. D. Barbian, S. B. Beres, L. E. Philips, J. M. Voyich, F. R. DeLeo, J. M. Martin, G. A. Somerville and J. M. Musser, *J. Infect. Dis.*, 2004, **190**, 727–738.

177 S. B. Beres, G. L. Sylva, K. D. Barbian, B. Lei, J. S. Hoff, N. D. Mammarella, M. Y. Liu, J. C. Smoot, S. F. Porcella, L. D. Parkins, D. S. Campbell, T. M. Smith, J. K. McCormick, D. Y. Leung, P. M. Schlievert and J. M. Musser, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 10078–10083.

178 J. C. Smoot, K. D. Barbian, J. J. Van Gompel, L. M. Smoot, M. S. Chaussee, G. L. Sylva, D. E. Sturdevant, S. M. Ricklefs, S. F. Porcella, L. D. Parkins, S. B. Beres, D. S. Campbell, T. M. Smith, Q. Zhang, V. Kapur, J. A. Daly, L. G. Veasy and J. M. Musser, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4668–4673.

179 I. Nakagawa, K. Kurokawa, A. Yamashita, M. Nakata, Y. Tomiyasu, N. Okahashi, S. Kawabata, K. Yamazaki, T. Shiba, T. Yasunaga, H. Hayashi, M. Hattori and S. Hamada, *Genome Res.*, 2003, **13**, 1042–1055.

180 A. Bolotin, B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau and P. Hols, *Nat. Biotechnol.*, 2004, **22**, 1554–1558.

181 K. Tsuge, T. Ano, M. Hirai, Y. Nakamura and M. Shoda, *Antimicrob. Agents Chemother.*, 1999, **43**, 2183–2192.

182 D. Mootz, R. Finking and M. A. Marahiel, *J. Biol. Chem.*, 2001, **276**, 37289–37298.

183 F. Peypoux, J. M. Bonmatin and J. Wallach, *Appl. Microbiol. Biotechnol.*, 1999, **51**, 553–563.

184 C. Scotti, M. Piatti, A. Cuzzoni, P. Perani, A. Tognoni, G. Grandi, A. Galizzi and A. M. Albertini, *Gene*, 1993, **130**, 65–71.

185 T. Stein, *Mol. Microbiol.*, 2005, **56**, 845–857.

186 J. Hofemeister, B. Conrad, B. Adler, B. Hofemeister, J. Feesche, N. Kucheryava, G. Steinborn, P. Franke, N. Grammel, A. Zwintscher, F. Leenders, G. Hitzeroth and J. Vater, *Mol. Genet. Genomics.*, 2004, **272**, 363–378.

187 D. Konz, S. Doekel and M. A. Marahiel, *J. Bacteriol.*, 1999, **181**, 133–140.

188 D. Konz, A. Klens, K. Schorgendorfer and M. A. Marahiel, *Chem. Biol.*, 1997, **4**, 927–937.

189 S. Cendrowski, W. MacArthur and P. Hanna, *Mol. Microbiol.*, 2004, **51**, 407–417.

190 R.-Y. Park, M.-H. Choi, H.-Y. Sun and S.-H. Shin, *Biol. Pharm. Bull.*, 2005, **28**, 1132–1135.

191 E. H. Duitman, L. W. Hamoen, M. Rembold, G. Venema, H. Seitz, W. Saenger, F. Bernhard, R. Reinhardt, M. Schmidt, C. Ullrich, T. Stein, F. Leenders and J. Vater, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 13294–13299.

192 M. Ehling-Schulz, N. Vukov, A. Schulz, R. Shaheen, M. Andersson, E. Märtlbauer and S. Scherer, *Appl. Environ. Microbiol.*, 2005, **71**, 105–113.

193 E. A. B. Emmert, A. K. Klimowicz, M. G. Thomas and J. Hadelsman, *Appl. Environ. Microbiol.*, 2004, **70**, 104–113.

194 M. Wu, Q. Ren, A. S. Durkin, S. C. Daugherty, L. M. Brinkac, R. J. Dodson, R. Madupu, S. A. Sullivan, J. F. Kolonay, D. H. Haft, W. C. Nelson, L. J. Tallon, K. M. Jones, L. E. Ulrich, J. M. Gonzalez, I. B. Zhulin, F. T. Robb and J. A. Eisen, *PLoS Genet.*, 2005, **1**, E65.

195 J. Nolling, G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin and D. R. Smith, *J. Bacteriol.*, 2001, **183**, 4823–4838.

196 T. Shimizu, K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori, S. Kuhara and H. Hayashi, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 996–1001.

197 H. Bruggemann, S. Baumer, W. F. Fricke, A. Wiezer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne and G. Gottschalk, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1316–1321.

198 Q. Bao, Y. Tian, W. Li, Z. Xu, Z. Xuan, S. Hu, W. Dong, J. Yang, Y. Chen, Y. Xue, Y. Xu, X. Lai, L. Huang, X. Dong, Y. Ma, L. Ling, H. Tan, R. Chen, J. Wang, J. Yu and H. Yang, *Genome Res.*, 2002, **12**, 689–700.

199 K. Oshima, S. Kakizawa, H. Nishigawa, H. Y. Jung, W. Wei, S. Suzuki, R. Arashida, D. Nakata, S. Miyata, M. Ugaki and S. Namba, *Nat. Genet.*, 2004, **36**, 27–29.

200 L. Papazisi, T. S. Gorton, G. Kutish, P. F. Markham, G. F. Browning, D. Kim Nguyen, S. Swartzell, A. Madan, G. Mahairas and S. J. Geary, *Microbiology*, 2003, **149**, 2307–2316.

201 C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, R. D. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. F. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S.

Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, 3rd and J. V. Venter, *Science*, 1995, **270**, 397–403.
202 F. C. Minion, E. J. Lefkowitz, M. L. Madsen, B. J. Cleary, S. M. Swartzell and G. G. Mahairas, *J. Bacteriol.*, 2004, **186**, 7123–7133.
203 A. T. Vasconcelos, H. B. Ferreira, C. V. Bizarro, S. L. Bonatto, M. O. Carvalho, P. M. Pinto, D. F. Almeida, L. G. Almeida, R. Almeida, L. Alves-Filho, E. N. Assuncao, V. A. Azevedo, M. R. Bogo, M. M. Brigido, M. Brocchi, H. A. Burity, A. A. Camargo, S. S. Camargo, M. S. Carepo, D. M. Carraro, J. C. de Mattos Cascardo, L. A. Castro, G. Cavalcanti, G. Chemale, R. G. Collevatti, C. W. Cunha, B. Dallagiovanna, B. P. Dambros, O. A. Dellagostin, C. Falcao, F. Fantinatti-Garboggini, M. S. Felipe, L. Fiorentin, G. R. Franco, N. S. Freitas, D. Frias, T. B. Grangeiro, E. C. Grisard, C. T. Guimaraes, M. Hungria, S. N. Jardim, M. A. Krieger, J. P. Laurino, L. F. Lima, M. I. Lopes, E. L. Loreto, H. M. Madeira, G. P. Manfio, A. Q. Maranhao, C. T. Martinkovics, S. R. Medeiros, M. A. Moreira, M. Neiva, C. E. Ramalho-Neto, M. F. Nicolas, S. C. Oliveira, R. F. Paixao, F. O. Pedrosa, S. D. Pena, M. Pereira, L. Pereira-Ferrari, I. Piffer, L. S. Pinto, D. P. Potrich, A. C. Salim, F. R. Santos, R. Schmitt, M. P. Schneider, A. Schrank, I. S. Schrank, A. F. Schuck, H. N. Seuanez, D. W. Silva, R. Silva, S. C. Silva, C. M. Soares, K. R. Souza, R. C. Souza, C. C. Staats, M. B. Steffens, S. M. Teixeira, T. P. Urmenyi, M. H. Vainstein, L. W. Zuccherato, A. J. Simpson and A. Zaha, *J. Bacteriol.*, 2005, **187**, 5568–5577.
204 J. D. Jaffe, N. Stange-Thomann, C. Smith, D. DeCaprio, S. Fisher, J. Butler, S. Calvo, T. Elkins, M. G. FitzGerald, N. Hafez, C. D. Kodira, J. Major, S. Wang, J. Wilkinson, R. Nicol, C. Nusbaum, B. Birren, H. C. Berg and G. M. Church, *Genome Res.*, 2004, **14**, 1447–1461.
205 J. Westberg, A. Persson, A. Holmberg, A. Goesmann, J. Lundeberg, K. E. Johansson, B. Pettersson and M. Uhlen, *Genome Res.*, 2004, **14**, 221–227.
206 Y. Sasaki, J. Ishikawa, A. Yamashita, K. Oshima, T. Kenri, K. Furuya, C. Yoshino, A. Horino, T. Shiba, T. Sasaki and M. Hattori, *Nucleic Acids Res.*, 2002, **30**, 5293–5300.
207 T. Dandekar, M. Huynen, J. T. Regula, B. Ueberle, C. U. Zimmermann, M. A. Andrade, T. Doerks, L. Sanchez-Pulido, B. Snel, M. Suyama, Y. P. Yuan, R. Herrmann and P. Bork, *Nucleic Acids Res.*, 2000, **28**, 3278–3288; R. Himmelreich, H. Hilbert, H. Plagens, E. Pirkl, B. C. Li and R. Herrmann, *Nucleic Acids Res.*, 1996, **24**, 4420–4449.
208 I. Chambaud, R. Heilig, S. Ferris, V. Barbe, D. Samson, F. Galisson, I. Moszer, K. Dybvig, H. Wroblewski, A. Viari, E. P. Rocha and A. Blanchard, *Nucleic Acids Res.*, 2001, **29**, 2145–2153.
209 J. I. Glass, E. J. Lefkowitz, J. S. Glass, C. R. Heiner, E. Y. Chen and G. H. Cassell, *Nature*, 2000, **407**, 757–762.
210 M. A. Schell, M. Karmirantzou, B. Snel, D. Vilanova, B. Berger, G. Pessi, M.-C. Zwahlen, F. Desiere, P. Bork, M. Delley, R. D. Pridmore and F. Arigoni, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14422–14427.
211 A. M. Cerdeno-Tarraga, A. Efstratiou, L. G. Dover, M. T. Holden, M. Pallen, S. D. Bentley, G. S. Besra, C. Churcher, K. D. James, A. De Zoysa, T. Chillingworth, A. Cronin, L. Dowd, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Moule, M. A. Quail, E. Rabbinowitsch, K. M. Rutherford, N. R. Thomson, L. Unwin, S. Whitehead, B. G. Barrell and J. Parkhill, *Nucleic Acids Res.*, 2003, **31**, 6516–6523.
212 Y. Nishio, Y. Nakamura, Y. Kawarabayasi, Y. Usuda, E. Kimura, S. Sugimoto, K. Matsui, A. Yamagishi, H. Kikuchi, K. Ikeo and T. Gojobori, *Genome Res.*, 2003, **13**, 1572–1579.
213 J. Kalinowski, B. Bathe, D. Bartels, N. Bischoff, M. Bott, A. Burkovski, N. Dusch, L. Eggeling, B. J. Eikmanns, L. Gaigalat, A. Goesmann, M. Hartmann, K. Huthmacher, R. Krämer, B. Linke, A. C. McHardy, F. Meyer, B. Möckel, W. Pfefferle, A. Pühler, D. A. Rey, C. Rückert, O. Rupp, H. Sahm, V. F. Wendisch, I. Wiegräbe and A. Tauch, *J. Biotechnol.*, 2003, **104**, 5–25.
214 A. Tauch, O. Kaiser, T. Hain, A. Goesmann, B. Weisshaar, A. Albersmeier, T. Bekel, N. Bischoff, I. Brune, T. Chakraborty, J. Kalinowski, F. Meyer, O. Rupp, S. Schneiker, P. Viehoever and A. Pühler, *J. Bacteriol.*, 2005, **187**, 4671–4682.
215 C. B. Monteiro-Vitorello, L. E. Camargo, M. A. Van Sluys, J. P. Kitajima, D. Truffi, A. M. do Amaral, R. Harakava, J. C. de Oliveira, D. Wood, M. C. de Oliveira, C. Miyaki, M. A. Takita, A. C. da Silva, L. R. Furlan, D. M. Carraro, G. Camarotte, N. F. Almeida, Jr., H. Carrer, L. L. Coutinho, H. A. El-Dorry, M. I. Ferro, P. R. Gagliardi, E. Giglioti, M. H. Goldman, G. H. Goldman, E. T. Kimura, E. S. Ferro, E. E. Kuramae, E. G. Lemos, M. V. Lemos, S. M. Mauro, M. A. Machado, C. L. Marino, C. F. Menck, L. R. Nunes, R. C. Oliveira, G. G. Pereira, W. Siqueira, A. A. de Souza, S. M. Tsai, A. S. Zanca, A. J. Simpson, S. M. Brumbley and J. C. Setubal, *Mol. Plant-Microbe Interact.*, 2004, **17**, 827–836.
216 L. Li, J. P. Bannantine, Q. Zhang, A. Amonsin, B. J. May, D. Alt, N. Banerji, S. Kanjilal and V. Kapur, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 12344–12349.
217 A. Tauch, O. Kaiser, T. Hain, A. Goesmann, B. Weisshaar, A. Albersmeier, T. Bekel, N. Bischoff, I. Brune, T. Chakraborty, J. Kalinowski, F. Meyer, O. Rupp, S. Schneiker, P. Viehoever and A. Puhler, *J. Bacteriol.*, 2005, **187**, 4671–4682.
218 S. T. Cole, K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward and B. G. Barrell, *Nature*, 2001, **409**, 1007–1011.
219 R. D. Fleischmann, D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter and C. M. Fraser, *J. Bacteriol.*, 2002, **184**, 5479–5490.
220 S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, W. Whitehead and B. G. Barrell, *Nature*, 1998, **11**(393), 537–544.
221 J. Ishikawa, A. Yamashita, Y. Mikami, Y. Hoshino, H. Kurita, K. Hotta, T. Shiba and M. Hattori, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14925–14930.
222 H. Bruggemann, A. Henne, F. Hoster, H. Liesegang, A. Wiezer, A. Strittmatter, S. Hujer, P. Durre and G. Gottschalk, *Science*, 2004, **305**, 671–673.
223 H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori and S. Omura, *Nat. Biotechnol.*, **21**, 526–531.
224 K. Ueda, A. Yamashita, J. Ishikawa, M. Shimada, T. Watsuji, K. Morimura, H. Ikeda, M. Hattori and T. Beppu, *Nucleic Acids Res.*, 2004, **32**, 4937–4944.
225 D. Raoult, H. Ogata, S. Audic, C. Robert, K. Suhre, M. Drancourt and J. M. Claverie, *Genome Res.*, 2003, **13**, 1800–1809.
226 S. D. Bentley, M. Maiwald, L. D. Murphy, M. J. Pallen, C. A. Yeats, L. G. Dover, H. T. Norbertczak, G. S. Besra, M. A. Quail, D. E. Harris, A. von Herbay, A. Goble, S. Rutter, R. Squares, S. Squares, B. G. Barrell, J. Parkhill and D. A. Relman, *Lancet*, 2003, **361**, 637–644.
227 O. H. Shiratori, M.-J. Park, Y. Saito, Y. Kumon, N. Yamashita, A. Hirata, H. Nishida, K. Ueda and T. Beppu, *Int. J. Syst. Evol. Microbiol.*, 2000, **50**, 1829–1832; R.-Y. Park, M.-H. Choi, H.-Y. Sun and S.-H. Shin, *Biol. Pharm. Bull.*, 2005, **28**, 1132–1135.
228 D. Portevin, C. de Sousa-D'Auria, H. Montrozier, C. Houssin, A. Stella, M.-A. Lanéelle, F. Bardou, C. Guilhot and M. Daffé, *J. Biol. Chem.*, 2005, **280**, 8862–8874.
229 K. Takayama, C. Wang and G. S. Besra, *Clin. Microbiol. Rev.*, 2005, **18**, 81–101.
230 L. E. N. Quadri, J. Sello, T. A. Keating, P. H. Weinreb and C. Walsh, *Chem. Biol.*, 1998, **5**, 631–645.
231 R. Krithika, U. Marathe, P. Saxena, M. Z. Ansari, D. Mohanty and R. S. Gokhale, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 2069–2074.
232 I. Matsunaga, A. Bhatt, D. C. Young, T. Y. Cheng, S. J. Eyles, G. S. Besra, V. Briken, S. A. Porcelli, C. E. Costello, W. R. Jacobs, Jr. and D. B. Moody, *J. Exp. Med.*, 2004, **200**, 1559–1569.
233 D. E. Minnikin, L. Kremer, L. G. Dover and G. S. Besra, *Chem. Biol.*, 2002, **9**, 545–553.
234 O. A. Trivedi, P. Arora, A. Vats, M. Z. Ansari, R. Tickoo, V. Sridharan, D. Mohanty and R. S. Gokhale, *Mol. Cell*, 2005, **17**, 631–643.
235 M. Mathur and P. E. Kolattukudy, *J. Biol. Chem.*, 1992, **267**, 19388–19395.

236 P. Constant, E. Perez, W. Malaga, M. A. Laneelle, O. Saurel, M. Daffe and C. Guilhot, *J. Biol. Chem.*, 2002, **277**, 38148–38158.
237 T. D. Sirakova, A. K. Thirumala, V. S. Dubey, H. Sprecher and P. E. Kolattukudy, *J. Biol. Chem.*, 2001, **276**, 16833–16839.
238 G. S. Hotter, B. J. Wards, P. Mouat, G. S. Besra, J. Gomes, M. Singh, S. Bassett, P. Kawakami, P. R. Wheeler, G. W. de Lisle and D. M. Collins, *J. Bacteriol.*, 2005, **187**, 2267–2277.
239 S. Omura, H. Ikeda, J. Ishikawa, A. Hanamoto, C. Takahashi, M. Shinose, Y. Takahashi, H. Horikawa, H. Nakazawa, T. Osonoe, H. Kikuchi, T. Shiba, Y. Sakaki and M. Hattori, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 12215–12220.
240 H. Ikeda, T. Nonomiya, M. Usami, T. Ohta and S. Omura, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9509–9514.
241 G. L. Challis and J. Ravel, *FEMS Microbiol. Lett.*, 2000, **187**, 111–114.
242 S. Lautru, R. Deeth, L. M. Bailey and G. L. Challis, *Nat. Chem. Biol.*, 2005, **1**, 265–269.
243 E. Takano, H. Kinoshita, V. Mersinias, G. Bucca, G. Hotchkiss, T. Nihira, C. P. Smith, M. Bibb, W. Wohlleben and K. Chater, *Mol. Microbiol.*, 2005, **56**, 465–479.
244 K. Pawlik, M. Kotowska, K. F. Chater, K. Kuczek and E. Takano, *Arch. Microbiol.*, 2007, **187**, 87–99.
245 A. Rantala, D. P. Fewer, M. Hisbergues, L. Rouhiainen, J. Vaitomaa, T. Börner and K. Sivonen, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 568–573.
246 B. A. Neilan, E. Dittmann, L. Rouhiainen, R. A. Bass, V. Schaub, K. Sivonen and T. Börner, *J. Bacteriol.*, 1999, **181**, 4089–4097; I. M. Ehrenreich, J. B. Waterbury and E. A. Webb, *Appl. Environ. Microbiol.*, 2005, **71**, 7401–7413.
247 T. Kaneko, Y. Nakamura, C. P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, M. Kohara, M. Matsumoto, A. Matsuno, A. Muraki, N. Nakazaki, S. Shimpo, M. Sugimoto, M. Takazawa, M. Yamada, M. Yasuda and S. Tabata, *DNA Res.*, 2001, **8**, 205–213.
248 Y. Nakamura, T. Kaneko, S. Sato, M. Mimuro, H. Miyashita, T. Tsuchiya, S. Sasamoto, A. Watanabe, K. Kawashima, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, N. Nakazaki, S. Shimpo, C. Takeuchi, M. Yamada and S. Tabata, *DNA Res.*, 2003, **10**, 137–145.
249 G. Rocap, F. W. LarimeR, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser and S. W. Chisholm, *Nature*, 2003, **424**, 1042–1047.
250 A. Dufresne, M. Salanoubat, F. Partensky, F. Artiguenave, I. M. Axmann, V. Barbe, S. Duprat, M. Y. Galperin, E. V. Koonin, F. Le Gall, K. S. Makarova, M. Ostrowski, S. Oztas, C. Robert, I. B. Rogozin, D. J. Scanlan, N. Tandeau de Marsac, J. Weissenbach, P. Wincker, Y. I. Wolf and W. R. Hess, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 10020–10025.
251 B. Palenik, B. Brahamsha, F. W. Larimer, M. Land, L. Hauser, P. Chain, J. Lamerdin, W. Regala, E. E. Allen, J. McCarren, I. Paulsen, A. Dufresne, F. Partensky, E. A. Webb and J. Waterbury, *Nature*, 2003, **424**, 1037–1042.
252 T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda and S. Tabata, *DNA Res.*, 1996, **3**, 109–136.
253 Y. Nakamura, T. Kaneko, S. Sato, M. Ikeuchi, H. Katoh, S. Sasamoto, A. Watanabe, M. Iriguchi, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada and S. Tabata, *DNA Res.*, 2002, **9**, 123–130.
254 E. L. Campbell, M. F. Cohen and J. C. Meeks, *Arch. Microbiol.*, 1997, **167**, 251–258.
255 F. O. Glockner, M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann and R. Reinhardt, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 8298–8303.
256 A. M. Cerdeno-Tarraga, S. Patrick, L. C. Crossman, G. Blakely, V. Abratt, N. Lennard, I. Poxton, B. Duerden, B. Harris, M. A. Quail, A. Barron, L. Clark, C. Corton, J. Doggett, M. T. Holden, N. Larke, A. Line, A. Lord, H. Norbertczak, D. Ormond, C. Price, E. Rabbinowitsch, J. Woodward, B. Barrell and J. Parkhill, *Science*, 2005, **307**, 1463–1465.
257 T. Kuwahara, A. Yamashita, H. Hirakawa, H. Nakayama, H. Toh, N. Okada, S. Kuhara, M. Hattori, T. Hayashi and Y. Ohnishi, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14919–14924.
258 J. Xu, M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper and J. I. Gordon, *Science*, 2003, **299**, 2074–2076.
259 K. E. Nelson, R. D. Fleischmann, R. T. DeBoy, I. T. Paulsen, D. E. Fouts, J. A. Eisen, S. C. Daugherty, R. J. Dodson, A. S. Durkin, M. Gwinn, D. H. Haft, J. F. Kolonay, W. C. Nelson, T. Mason, L. Tallon, J. Gray, D. Granger, H. Tettelin, H. Dong, J. L. Galvin, M. J. Duncan, F. E. Dewhirst and C. M. Fraser, *J. Bacteriol.*, 2003, **185**, 5591–5601.
260 E. F. Mongodin, K. E. Nelson, S. Daugherty, R. T. Deboy, J. Wister, H. Khouri, J. Weidman, D. A. Walsh, R. T. Papke, G. Sanchez Perez, A. K. Sharma, C. L. Nesbo, D. MacLeod, E. Bapteste, W. F. Doolittle, R. L. Charlebois, B. Legault and F. Rodriguez-Valera, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 18147–18152.
261 G. Deckert, P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen and R. V. Swanson, *Nature*, 1998, **392**, 353–358.
262 T. D. Read, R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen and C. M. Fraser, *Nucleic Acids Res.*, 2000, **28**, 1397–1406.
263 J. H. Carlson, S. F. Porcella, G. McClarty and H. D. Caldwell, *Infect. Immun.*, 2005, **73**, 6407–6418.
264 R. S. Stephens, S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin and R. W. Davis, *Science*, 1998, **282**, 754–759.
265 N. R. Thomson, C. Yeats, K. Bell, M. T. Holden, S. D. Bentley, M. Livingstone, A. M. Cerdeno-Tarraga, B. Harris, J. Doggett, D. Ormond, K. Mungall, K. Clarke, T. Feltwell, Z. Hance, M. Sanders, M. A. Quail, C. Price, B. G. Barrell, J. Parkhill and D. Longbottom, *Genome Res.*, 2005, **15**, 629–640.
266 T. D. Read, G. S. Myers, R. C. Brunham, W. C. Nelson, I. T. Paulsen, J. Heidelberg, E. Holtzapple, H. Khouri, N. B. Federova, H. A. Carty, L. A. Umayam, D. H. Haft, J. Peterson, M. J. Beanan, O. White, S. L. Salzberg, R. C. Hsia, G. McClarty, R. G. Rank, P. M. Bavoil and C. M. Fraser, *Nucleic Acids Res.*, 2003, **31**, 2134–2147.
267 S. Kalman, W. Mitchell, R. Marathe, C. Lammel, J. Fan, R. W. Hyman, L. Olinger, J. Grimwood, R. W. Davis and R. S. Stephens, *Nat. Genet.*, 1999, **21**, 385–389.
268 M. Shirai, H. Hirakawa, M. Kimoto, M. Tabuchi, F. Kishi, K. Ouchi, T. Shiba, K. Ishii, M. Hattori, S. Kuhara and T. Nakazawa, *Nucleic Acids Res.*, 2000, **28**, 2311–2314.
269 M. Horn, A. Collingro, S. Schmitz-Esser, C. L. Beier, U. Purkhold, B. Fartmann, P. Brandt, G. J. Nyakatura, M. Droege, D. Frishman, T. Rattei, H. W. Mewes and M. Wagner, *Science*, 2004, **304**, 728–730.
270 J. A. Eisen, K. E. Nelson, I. T. Paulsen, J. F. Heidelberg, M. Wu, R. J. Dodson, R. Deboy, M. L. Gwinn, W. C. Nelson, D. H. Haft, E. K. Hickey, J. D. Peterson, A. S. Durkin, J. L. Kolonay, F. Yang, I. Holt, L. A. Umayam, T. Mason, M. Brenner, T. P. Shea, D. Parksey, W. C. Nierman, T. V. Feldblyum, C. L. Hansen, M. B. Craven, D. Radune, J. Vamathevan, H. Khouri, O. White, T. M. Gruber, K. A. Ketchum, J. C. Venter, H. Tettelin, D. A. Bryant and C. M. Fraser, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9509–9514.
271 O. White, J. A. Eisen, J. F. Heidelberg, E. K. Hickey, J. D. Peterson, R. J. Dodson, D. H. Haft, M. L. Gwinn, W. C. Nelson, D. L. Richardson, K. S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J. J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K. S. Makarova, L. Aravind, M. J. Daly, K. W. Minton, R. D. Fleischmann, K. A. Ketchum, K. E. Nelson, S. Salzberg, H. O. Smith, J. C. Venter and C. M. Fraser, *Science*, 1999, **286**, 1571–1577.
272 A. Henne, H. Bruggemann, C. Raasch, A. Wiezer, T. Hartsch, H. Liesegang, A. Johann, T. Lienard, O. Gohl, R. Martinez-Arias, C. Jacobi, V. Starkuviene, S. Schlenczeck, S. Dencker, R. Huber, H. P. Klenk, W. Kramer, R. Merkl, G. Gottschalk and H. J. Fritz, *Nat. Biotechnol.*, 2004, **22**, 547–553.
273 V. Kapatral, I. Anderson, N. Ivanova, G. Reznik, T. Los, A. Lykidis, A. Bhattacharyya, A. Bartman, W. Gardner, G. Grechkin, L. Zhu, O. Vasieva, L. Chu, Y. Kogan, O. Chaga, E. Goltsman, A. Bernal, N. Larsen, M. D'Souza, T. Walunas, G. Pusch, R. Haselkorn, M. Fonstein, N. Kyrpides and R. Overbeek, *J. Bacteriol.*, 2002, **184**, 2005–2018.

274 C. M. Fraser, S. Casjens, W. M. Huang, G. G. Sutton, R. Clayton, R. Lathigra, O. White, K. A. Ketchum, R. Dodson, E. K. Hickey, M. Gwinn, B. Dougherty, J. F. Tomb, R. D. Fleischmann, D. Richardson, J. Peterson, A. R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. van Vugt, N. Palmer, M. D. Adams, J. Gocayne, J. Weidman, T. Utterback, L. Watthey, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fuji, M. D. Cotton, K. Horst, K. Roberts, B. Hatch, H. O. Smith and J. C. Venter, *Nature*, 1997, **390**, 580–586.
275 G. Glockner, R. Lehmann, A. Romualdi, S. Pradella, U. Schulte-Spechtel, M. Schilhabel, B. Wilske, J. Suhnel and M. Platzer, *Nucleic Acids Res.*, 2004, **32**, 6038–6046.
276 A. L. Nascimento, A. I. Ko, E. A. Martins, C. B. Monteiro-Vitorello, P. L. Ho, D. A. Haake, S. Verjovski-Almeida, R. A. Hartskeerl, M. V. Marques, M. C. Oliveira, C. F. Menck, L. C. Leite, H. Carrer, L. L. Coutinho, W. M. Degrave, O. A. Dellagostin, H. El-Dorry, E. S. Ferro, M. I. Ferro, L. R. Furlan, M. Gamberini, E. A. Giglioti, A. Goes-Neto, G. H. Goldman, M. H. Goldman, R. Harakava, S. M. Jeronimo, I. L. Junqueira-de-Azevedo, E. T. Kimura, E. E. Kuramae, E. G. Lemos, M. V. Lemos, C. L. Marino, L. R. Nunes, R. C. de Oliveira, G. G. Pereira, M. S. Reis, A. Schriefer, W. J. Siqueira, P. Sommer, S. M. Tsai, A. J. Simpson, J. A. Ferro, L. E. Camargo, J. P. Kitajima, J. C. Setubal and M. A. Van Sluys, *J. Bacteriol.*, 2004, **186**, 2164–2172.
277 S. X. Ren, G. Fu, X. G. Jiang, R. Zeng, Y. G. Miao, H. Xu, Y. X. Zhang, H. Xiong, G. Lu, L. F. Lu, H. Q. Jiang, J. Jia, Y. F. Tu, J. X. Jiang, W. Y. Gu, Y. Q. Zhang, Z. Cai, H. H. Sheng, H. F. Yin, Y. Zhang, G. F. Zhu, M. Wan, H. L. Huang, Z. Qian, S. Y. Wang, W. Ma, Z. J. Yao, Y. Shen, B. Q. Qiang, Q. C. Xia, X. K. Guo, A. Danchin, I. Saint Girons, R. L. Somerville, Y. M. Wen, M. H. Shi, Z. Chen, J. G. Xu and G. P. Zhao, *Nature*, 2003, **422**, 888–893.
278 R. Seshadri, G. S. Myers, H. Tettelin, J. A. Eisen, J. F. Heidelberg, R. J. Dodson, T. M. Davidsen, R. T. DeBoy, D. E. Fouts, D. H. Haft, J. Selengut, Q. Ren, L. M. Brinkac, R. Madupu, J. Kolonay, S. A. Durkin, S. C. Daugherty, J. Shetty, A. Shvartsbeyn, E. Gebregeorgis, K. Geer, G. Tsegaye, J. Malek, B. Ayodeji, S. Shatsman, M. P. McLeod, D. Smajs, J. K. Howell, S. Pal, A. Amin, P. Vashisth, T. Z. McNeill, Q. Xiang, E. Sodergren, E. Baca, G. M. Weinstock, S. J. Norris, C. M. Fraser and I. T. Paulsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 5646–5651.
279 C. M. Fraser, S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, Y. Utterback, L. McDonald, P. Artiach, C. Bowman, M. D. Cotton, C. Fujii, S. Garland, B. Hatch, K. Horst, K. Roberts, M. Sandusky, J. Weidman, H. O. Smith and J. C. Venter, *Science*, 1998, **281**, 375–388.
280 K. E. Nelson, R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter and C. M. Fraser, *Nature*, 1999, **399**, 323–329.
281 G. L. Challis and D. A. Hopwood, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14555–14561.