NPR

Natural Product Reports

Current developments in natural products chemistry

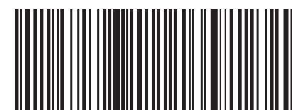
www.rsc.org/npr

Volume 30 | Number 3 | March 2013 | Pages 377–476

ISSN 0265-0568

RSC Publishing

REVIEW ARTICLE
Christian Hertweck *et al.*
A genomic approach to the cryptic secondary metabolome of the anaerobic world



NPR

RSCPublishing

REVIEW

View Article Online
View Journal | View Issue

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Cite this: *Nat. Prod. Rep.*, 2013, **30**, 392

A genomic approach to the cryptic secondary metabolome of the anaerobic world

Anne-Catrin Letzel,†[a] Sacha J. Pidot†[a] and Christian Hertweck*[ab]

Received 18th October 2012

DOI: 10.1039/c2np20103h

www.rsc.org/npr

Covering: up to September 2012.

A total of 211 complete and published genomes from anaerobic bacteria are analysed for the presence of secondary metabolite biosynthesis gene clusters, in particular those tentatively coding for polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS). We investigate the distribution of these gene clusters according to bacterial phylogeny and, if known, correlate these to the type of metabolic pathways they encode. The potential of anaerobes as secondary metabolite producers is highlighted.

1 Introduction
2 Organisation of PKS and NRPS systems
3 Bioinformatics and genome mining
4 Overview of anaerobe genomes
5 Phylum Firmicutes
5.1 Genus *Clostridium*
5.1.1 Solventogenic clostridia
5.1.2 Cellulolytic clostridia
5.1.3 Human-associated clostridia
5.1.4 Other Clostridia
5.2 Other members of phylum Firmicutes
5.2.1 Other members of the order Clostridiales
5.2.1.1 Family Heliobacteriacae
5.2.1.2 Clostridiales Family XI *incertae sedis*
5.2.1.3 Family Eubacteriaceae
5.2.1.4 Family Syntrophomonadaceae
5.2.1.5 Family Lachnospiraceae
5.2.1.6 Family Ruminococcaceae
5.2.1.7 Family Oscillospiraceae
5.2.1.8 Family Peptococcaceae
5.2.2 Order Thermoanaerobacterales
5.2.3 Order Natranaerobiales
5.2.4 Order Halanaerobiales
5.2.5 Other Firmicutes
6 Phylum Bacteroidetes
7 Phylum Proteobacteria
7.1 Conserved PUFA biosynthesis gene clusters amongst anaerobic Proteobacteria
7.2 Order Desulfuromonadales
7.2.1 Genus *Geobacter*
7.2.2 Genus Pelobacter
7.3 Order Syntrophobacterales
7.4 Order Desulfovibrionales
7.5 Order Desulfobacterales
8 Phylum Actinobacteria
8.1 Family Bifidobacteriaceae
8.2 Family Propionibacteriaceae
8.3 Family Coriobacteriaceae
9 Phylum Thermotogae
10 Phylum Spirochaetes
11 Other phyla
11.1 Phylum Verrucomicrobia
11.2 Phylum Elusimicrobia
11.3 Phylum Fusobacteria
11.4 Phylum Fibrobacteres
11.5 Unclassified bacteria
12 Conclusions
13 References

[a] *Leibniz Institute for Natural Product Research and Infection Biology HKI, Beutenbergstr. 11a, Jena, 07745, Germany. E-mail: christian.hertweck@hki-jena.de*
[b] *Chair of Natural Product Chemistry, Friedrich Schiller University, Jena, Germany*
† These authors contributed equally to this work.

1 Introduction

Secondary metabolites represent a pharmaceutically important class of compounds with activities ranging from antibiotic and antiparasitic to immunosuppressant and anti-cancer and even lowering cholesterol.[1] Despite their medical relevance to humans, many of these compounds also play roles within their host organisms (usually bacteria or fungi) as pigments, defence molecules or virulence factors. Many strains of bacteria have been screened over the years for their ability to produce

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

secondary metabolites, however, arguably the most prolific producers are members of the genus *Streptomyces*.[2] Other genera are also able to produce secondary metabolites and analysis of increasing numbers of bacterial genome sequences has shown that bacteria previously considered "non-producers" possess genes for secondary metabolite production.[3] Many of these "non-producers" fall into the category of neglected organisms, which have been overlooked in terms of the time and effort spent on investigating the diversity of their secondary metabolome.

With respect to natural products, one group of highly neglected bacteria are those that grow without oxygen: the anaerobes. Anaerobic bacteria are ubiquitous in the environment and play major roles in the natural cycling of both carbon and sulfur,[4] and in the degradation of organic matter,[5] are important in biofuel manufacture,[6] and certain species are also well known pathogens.[7] A major obstacle to the investigation of natural products in anaerobes comes from the long held view amongst secondary metabolite researchers that these bacteria are incapable of producing secondary metabolites.[8] However, the isolation of the first antibiotic from an anaerobe, closthiamide from *Clostridium cellulolyticum*, has turned this supposition on its head.[9] Once thought of as difficult to manipulate, advances in recent years have helped to overcome problems of genetic intractability for some anaerobic genera.[10–13] This has paved the way for further studies on anaerobes and has opened the door to secondary metabolite researchers interested in exploring the anaerobic world.

In recent years, the field of natural product research has witnessed a convergence of genomics, microbiology and chemistry. In practical terms, this has meant that preliminary evaluations of microorganisms for their secondary metabolic potential can now be performed without inoculating a single culture. The use of bioinformatic tools to identify genes for natural product formation has become widespread, and several novel compounds have been identified as a direct result of these efforts. Although this strategy has been applied to whole genera of sequenced organisms,[14] it has not yet been applied to organisms from multiple genera with similar metabolic features. Our aim here is to use bioinformatics to probe the genomes of a cohort of anaerobic bacteria and, as a result, unveil their secondary metabolic potential, which has previously been discounted by the natural products community. We have focused our efforts on identifying polyketide synthase (PKS) and non-ribosomal peptide synthase (NRPS) genes of obligately anaerobic eubacteria whose genomes have been completely sequenced and published, with reference to incomplete or draft genome sequences where appropriate from a comparative genomics perspective.

2 Organisation of PKS and NRPS systems

Both PKS and NRPS systems represent large biosynthetic machines that work in concert to produce molecules with a wide range of complexities. For the purpose of this review, we will briefly outline the enzymatic content of these molecular assembly lines and how one can use the genetic make up of a PKS or NRPS to predict the structure of a product. While both of these types of systems produce structurally unrelated compounds, they have comparable logic in their mechanisms of action.

Anne-Catrin Letzel earned her diploma degree (2009) in Applied Natural Science from the Technical University of Freiberg, Germany. Since 2010 she has been a Ph.D. student at the Department of Biomolecular Chemistry, Leibniz Institue for Natural Product Research and Infection Biology, Jena under the supervision of Prof. Christian Hertweck. Her research is on secondary metabolites from Clostridium spp.

Sacha Pidot was born and raised in Melbourne, Australia. Following undergraduate studies at Monash University, where he received a BSc (Hons) in Microbiology, he joined the lab of Associate Professor Tim Stinear. His Ph. D. studies centered on molecular mechanisms of pathogenesis in the mycolactone producing mycobacterium M. ulcerans *and he was awarded his Ph.D. in 2011. Later that year, he joined the laboratory of Prof. Christian Hertweck as a Humboldt Postdoctoral Fellow, where he is currently investigating secondary metabolites from anaerobic bacteria.*

Christian Hertweck gained a Ph.D. in Organic Chemistry at the University of Bonn and at the MPI for Chemical Ecology (with Prof. Boland). From 1999–2000 he was Feodor Lynen postdoctoral fellow of Profs Floss and Moore at the University of Washington, Seattle, before starting his own research group at the HKI Jena. Since 2006 he has been Full Professor at the Friedrich Schiller University, Jena, and Head of the Department of Biomolecular Chemistry at the Leibniz Institute for Natural Product Research and Infection Biology (HKI). His research interests encompass various aspects of bacterial and fungal natural products, with focus on biosynthesis, genome mining and microbial interactions.

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

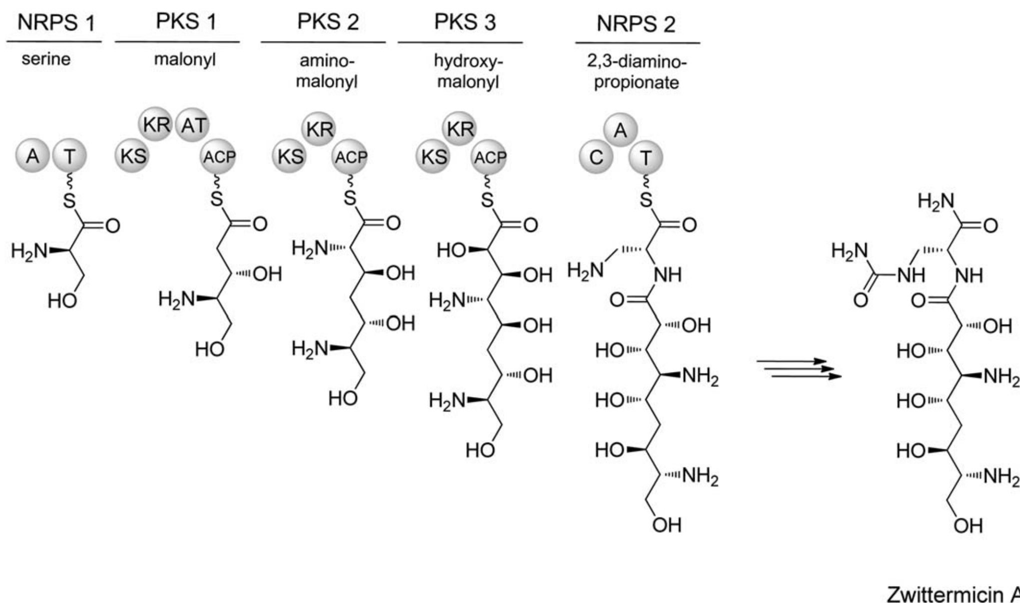


Fig. 1 A simplified mechanism of action for NRPS and PKS modules within the hybrid synthetase involved in the biosynthesis of zwittermicin A.[15]

PKS systems come in three different varieties: type I, type II and type III PKSs. The mechanism of action of each class differs slightly and the reader is referred to a number of specialized reviews on this subject for detailed enzymatic and mechanistic information.[16–18] The three broad classes of PKSs share the ability to utilise simple building blocks, such as malonyl-CoA or methylmalonyl-CoA, to produce a myriad of different molecular structures. The type I modular PKS are the most well studied and consist of large proteins, which are divided into modules, with each module containing a set of enzymatic activities. These enzymatic domains are the workhorses of the PKS that act cooperatively to extend the growing polyketide chain. Examples of PKS domains include acyltransferases (AT), acyl-carrier proteins (ACP), ketosynthases (KS), dehydratases (DH), enoyl-reductases (ER), ketoreductases (KR) and thioesterases (TE), although not every type of domain is present in each system. The majority of type I PKSs follow the co-linearity rule,[19] meaning that each domain performs a dedicated step in the biosynthetic process. Variations in PKS structure also exist, with some PKSs using an AT domain encoded by a distinct gene (a *trans*-AT), which binds to an intramodular *trans*-AT binding site to perform its acyltransferase function.[20]

NRPS systems use proteogenic and/or non-proteogenic amino acids as substrates for the generation of a range of different molecules.[21–23] In a similar fashion to type I PKSs, NRPS are also divided into modules that possess a discrete set of enzymatic domains, including adenylation (A), peptidyl carrier protein (PCP, also called thiolation (T) domain), condensation (C), epimerisation (E), methylation (M), reduction (R) and cyclisation (Cy) domains. Accessory genes are often part of PKS and NRPS systems, usually being present within the biosynthetic locus. One of the most important of these accessory genes is that encoding a 4′-phosphopantetheineyl transferase (PPTase), which is responsible for adding a 4′-phosphopantetheine group to either ACPs or PCPs, so that PKS or NRPS processing can proceed. Often assembly lines are found that comprise both PKS and NRPS functionalities. Such PKS-NRPS hybrids produce compounds composed of amino acids and malonyl-derived units, examples of which include bleomycin, epothilone and zwittermicin A (Fig. 1).[15,24]

3 Bioinformatics and genome mining

The fact that production of a polyketide or non-ribosomal peptide occurs in an orderly fashion has been seized upon by bioinformaticians in recent years to develop tools for the prediction of a product given the DNA or amino acid sequence of a PKS or NRPS assembly line. The possibility of an NRPS "code", similar in nature to the genetic code, was first raised in the 1970s,[25] and much work has been done since this time to decipher this code. Using advances in structural biology and phylogenetics, several groups were able to identify amino acid residues in both NRPS adenylation and PKS acyltransferase domains responsible for substrate selection.[26–28] These discoveries have been further built upon and have been turned into software tools (many of which are freely available online) to predict substrate specificity for A and AT domains.[28–35] Although prediction algorithms were initially developed for *cis*-AT domains, starter unit selection in *trans*-ATs can now be predicted,[36] as can domain specificities for a number of other polyketide and non-ribosomal peptide domains and modifying enzymes.[37–39]

Many of these tools have been put together in the recent release of antiSMASH, which is capable of searching whole genomes for natural product biosynthesis gene clusters (including those other than coding for NRPS and PKS), predicting both substrate selection and the associated product structure.[41] Although the advances in this area have been immense, multiple NRPS and PKS systems that violate the colinearity principle have been identified, and the current crop

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

of software tools cannot predict the products associated with such gene clusters.[42–45]

Many of the bioinformatic tools described above can be used to identify secondary metabolite genes in both targeted and metagenomic DNA sequencing projects. In the past five years there has been a seven-fold increase in the number of sequenced bacterial genomes (over 1800 bacterial genomes are available at the time of writing).[46] Although the currently available bacterial genomes are biased towards pathogens, efforts are underway to sequence relatively unknown branches of the tree of life, such as the Genomic Encyclopaedia of Archaea and Bacteria (GEBA) project.[47] The sequencing of these unusual lineages has been shown to increase the probability of discovering novel genes[48] and it is possible, therefore, that exploration of natural product biosynthesis in obscure bacteria will reveal novel secondary metabolite genes and pathways.

The incredible increase in the amount of sequencing data, mainly driven by the falling cost of DNA sequencing, has led to a paradigm shift in the approach to secondary metabolite discovery. As more genomes have been sequenced, it has become apparent that the plethora of secondary metabolite gene clusters far outweighs the number of described secondary metabolites.[49] The sequencing of the *Streptomyces avermitilis* genome is a classic example of this, with 25 secondary metabolite clusters being identified, many of which did not have an assigned product.[50] This means genome sequence data can be used to reveal previously unknown secondary metabolite clusters with no *a priori* knowledge of a strain's ability to produce natural products. This approach to secondary metabolite discovery is known as genome mining and has been successfully used to identify a number of novel compounds.[51–57] To date, however, genome mining has predominantly been used to identify the products of silent gene clusters (*i.e.* those clusters inactive under standard culture conditions) from well-established secondary metabolite producers (such as actinomycetes and myxobacteria).[58–60] There are a limited number of examples of genome mining of microorganisms that have long been neglected with respect to their biosynthetic potential, including representatives of the genera *Pseudomonas* and *Burkholderia*[36,61–67]

4 Overview of anaerobe genomes

With the genome mining strategy in mind, we have analysed the genomes of the 211 currently completed and published anaerobe genomes, which cover a variety of different branches of the eubacterial phylogenetic tree (Fig. 2, Table 1). Firmicute genomes make up almost 40% of completed and published anaerobe genomes, due to the presence of the clostridia, whose several pathogenic species have been the focus of a number of comparative genomic studies.[68–70] As relatively few published genome reports highlight features related to secondary metabolite production, especially for anaerobe genomes, we have used a combination of publicly available web-based bioinformatic tools to assist in the hunt for natural products in anaerobes. An overview of the secondary metabolic capacity of each genome was given by using antiSMASH, a recently developed tool for the genome-wide prediction of secondary metabolite biosynthesis gene clusters.[41] Although exactly which genes constitute a "cluster" is a matter of debate for uncharacterised gene loci, we have taken a logical approach and included genes that may play roles in product tailoring and transport. Once identified, individual genes were analysed using both Napdos and BLAST searches to try to gain further information about the potential product of each cluster and orthologous genes in other genomes.[31,71] All gene clusters containing a gene with putative PKS or NRPS functionality were included in the analysis. As such, 69 (33%) of the analysed genomes contain PKS or NRPS genes, for a total of just over 1 Mb of secondary metabolite biosynthesis related DNA from a total of 640 Mb of genome sequence (0.16% coverage) (Table 1). This total is approximately 70% lower than that found for bacteria in general,[3] indicating that secondary metabolites in anaerobes are not as common as those in facultative or aerobic bacteria.

As the reader will appreciate, there is virtually no chemical data on secondary metabolites from the anaerobic world. As such, almost all gene clusters identified here do not have an associated product, however, where clusters are similar to those linked to a known product, we have mentioned this in the text. A unifying feature of all the identified anaerobe PKS and NRPS genes is the low level of sequence homology to functionally characterised genes, suggesting the importance of conserved residues in large multi-domain enzymes with diverse amino acid sequences. For almost all of the identified gene clusters, we invite the reader to assume that encoded protein functions are putative, unless stated otherwise, to avoid frequent repetition of this word.

An analysis of the amount of secondary metabolite DNA relative to total genomic DNA shows that, similar to other bacteria, the larger the genome, the more likely an anaerobe is to possess secondary metabolite genes (Fig. 3).[3] Below a genome size of 3 Mb there is a very low chance of having secondary metabolite genes, however, even as genome size increases, large secondary metabolite biosynthesis gene clusters (>40 kb) are relatively rare (Fig. 3). Support for this notion is shown in Fig. 4, where it can be seen that, cumulatively, the largest genomes do not contribute the most secondary metabolite genes (Fig. 4). The non-uniform distribution of secondary metabolite biosynthesis gene clusters across phyla can be seen in both Fig. 2 and Table 1, a feature previously identified for other bacterial genome data sets.[3] The Firmicutes are overrepresented in terms of the amount of secondary metabolite DNA, while spirochaetes and anaerobic actinobacteria have little to no ability to produce secondary metabolites. Both the δ-proteobacteria and Firmicutes contain similar coding densities for secondary metabolite biosynthesis gene clusters, however, more genomes from other phyla are needed to give a clearer picture of their metabolic potential.

It seems that when it comes to anaerobes, the type of habitat or environment from which the organism was isolated is the best correlate of secondary metabolic potential (Fig. 5a). Anaerobes isolated from soil samples or similar habitats are more likely to contain secondary metabolite biosynthesis gene clusters (29 of 58 isolates), than those from other habitats. Isolates from habitats related to soil have three times more secondary metabolite DNA than anaerobes from all other habitat groups combined (Fig. 5a). On the opposite end of the

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

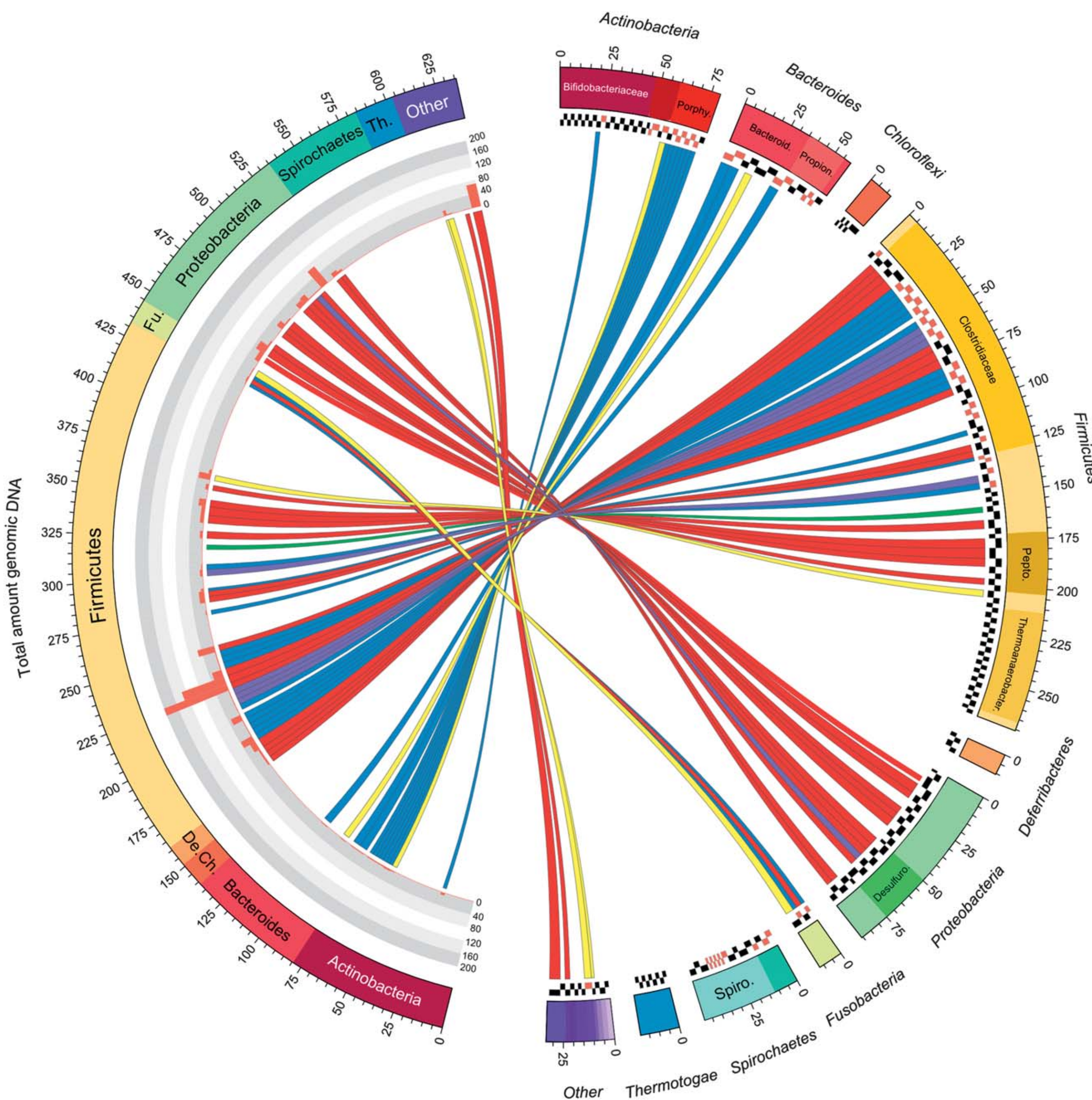


Fig. 2 Natural product biosynthesis gene clusters amongst all anaerobe genomes. The outermost circle represents the total amount of genomic DNA from all anaerobe genomes analysed in this study, divided into individual phyla (left side), or the total amount of genomic DNA in each phylum, with colour divisions signifying largest families or orders within each phylum (right side). The next innermost circle (right half) labels each genome as either pathogen (orange blocks) or non-pathogen (black blocks). The graph (third circle, left side) shows the total amount of NRPS and PKS encoding DNA in each genome (in Kb). Links connect the position of an organism within a phylum to its corresponding amount of NRPS or PKS encoding DNA and are colour coded according to habitat (red, soil; blue, human; yellow, animal; green, extreme environment; purple, other). Figure compiled using Circos.[40] Further information is given in Table 1 and the text.

scale, anaerobes from extreme environments, including those isolated from hot springs, as well as thermohalophilic or thermoalkaliphilic organisms, contain next to no PKS or NRPS genes. Although natural products have been isolated from organisms from these unusual environments,[72,73] none of these have been from anaerobes, suggesting that these organisms' niche may be so restricted as to not require these molecules. Furthermore, pathogenic anaerobes appear to contain much fewer PKS and NRPS gene clusters than non-pathogens (Fig. 5b). Although only 25% of the investigated anaerobe genomes were from pathogens, when NRPS and PKS DNA is evaluated as a function of total genomic DNA, non-pathogens have 2.5 times more secondary metabolite DNA than pathogenic anaerobes.

The 211 strains analysed here contain a total of 92 NRPS, PKS and hybrid gene clusters. Although some of the strains lacking NRPS or PKS gene clusters appear to have the potential to produce other secondary metabolites, such as bacteriocins, lantibiotics, and terpenes, our focus is on NRPS and PKS gene clusters, and we have not analysed these other types of natural products in detail here.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Table 1 Distribution of secondary metabolite genes amongst anaerobic genomes

Phylum	Number of strains	Number of strains with NRPS or PKS genes	Amount of PKS and NRPS DNA (kb)	Total genome size (Mb)	Coding density[a]
Actinobacteria	34	9	34.38	79.43	0.04
Bacteroidetes	14	4	9.77	59.08	0.02
Chlorobi	1	0	0	2.15	0.00
Chloroflexi	6	0	0	12.55	0.00
Chrysiogenetes	1	0	0	2.93	0.00
Proteobacteria	24	14	200.68	90.67	0.23
Deferribacteres	4	0	0	10.50	0.00
Deinococcus-Thermus	1	0	0	2.36	0.00
Elusimicrobia	1	1	16.28	1.64	0.99
Fibrobacteres	1	1	2.97	3.84	0.08
Firmicutes	83	34	690.21	275.12	0.25
Fusobacteria	4	4	12.38	12.25	0.10
Spirochaetes	21	0	0	51.18	0.00
Synergistetes	3	0	0	5.82	0.00
Thermodesulfobacteria	1	0	0	2.32	0.00
Thermotogae	10	0	0	19.30	0.00
Unclassified bacteria	1	1	1.03	2.75	0.04
Verrucomicrobia	2	1	68.88	8.62	0.80
Total	**211**	**69**	**1036.59**	**640.48**	**0.16**

[a] The amount of PKS and NRPS encoding DNA as a percentage of total genome size.

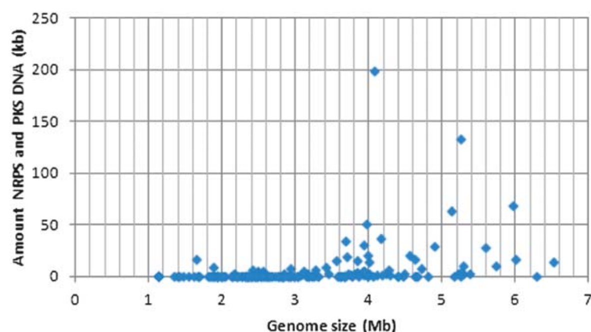


Fig. 3 Correlation between genome size and amount of secondary metabolite biosynthesis encoding DNA in completed and published anaerobe genomes.

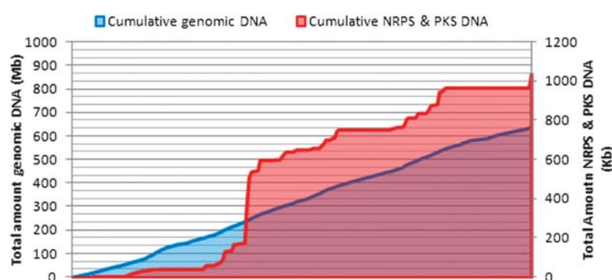
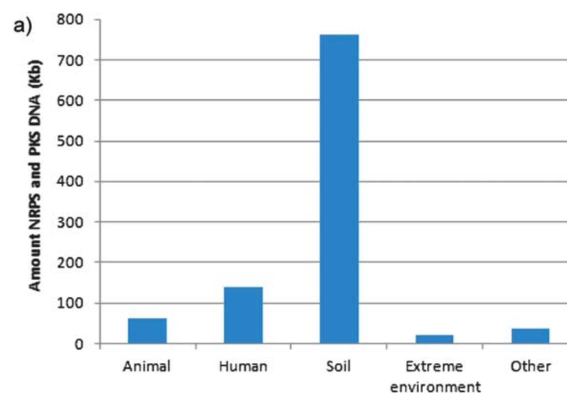


Fig. 4 Comparison of cumulative genome size and cumulative amount of secondary metabolite biosynthesis encoding DNA across anaerobe genomes analysed in this study.



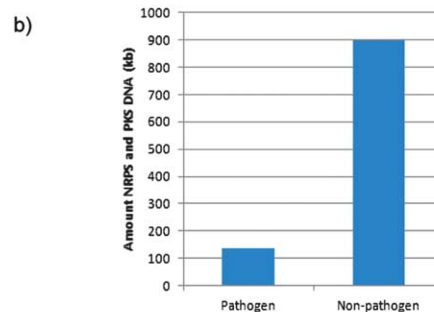


Fig. 5 a) Comparison of isolation site and amount of PKS and NRPS DNA for anaerobe genomes. Animal, isolates collected from both animals and insects; Human, pathogenic and non-pathogenic isolates collected from humans or used in food production; Soil, isolates collected from mud, soil, sludge or from reactors inoculated with soil or sludge; Extreme environment, organisms isolated from deep sea vents, hot springs, alkaline and saline lakes and oil wells; Other, samples whose origin does not fit into other categories or is unknown. b) Comparison of NRPS and PKS DNA harboured by pathogenic and non-pathogenic anaerobes.

Within the analysed set of genomes, NRPS gene clusters appear to be more common than either PKS or PKS-NRPS hybrid gene clusters, however, the data set is not large enough to determine if this is a general trend amongst all anaerobes. Both type I and type III PKS gene clusters were identified within the

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

target genome set, however, type II PKS genes were not identified in any of the genomes analysed. This is not unusual, given that type II PKSs are almost exclusively found in actinomycetes.[16] Due to the limited number of anaerobes that contain secondary metabolite biosynthesis gene clusters, we have not attempted to further subdivide PKS categories into *trans*- or *cis*-AT systems,[20] although both types are present within the analysed genomes.

As many of the gene clusters mentioned below do not have an associated product, the genes that form part of these loci do not have gene names. For ease of identification, we have used the locus tag assigned to individual genes as part of each genome sequencing project. To avoid confusion between the gene and protein level, we have italicised all locus tags when referring to genes, however, when referring to proteins the locus tag is in normal type.

5 Phylum Firmicutes

The phylum Firmicutes contains the highest number of sequenced anaerobe genomes, owing mainly to the presence of the clostridia, whose pathogen genomes are overrepresented amongst those of sequenced anaerobes. Of the 83 analysed genomes, 34 contain secondary metabolite biosynthesis gene clusters. However, these gene loci are not distributed evenly amongst genera and the presence of these particular gene loci does not seem to be related to phylogenetic position.

5.1 Genus *Clostridium*

The genus *Clostridium* has a measure of Jekyll and Hyde about it. On the one hand, this diverse group of Gram-positive, obligately anaerobic, spore forming rods includes some of the most fearsome human pathogens that have scourged humanity for centuries, including *Clostridium tetani* (tetanus), *Clostridium botulinum* (botulism) and *Clostridium perfringens* (gas gangrene), as well as pathogens that cause serious nosocomial infections, such as *Clostridium difficile* (diarrhoea).[74] These species are (in)famous for their range of virulence factors, including highly potent neuro- and enterotoxins. On the other hand, several clostridial species may be potentially beneficial as treatments for certain carcinomas,[75,76] and non-pathogenic clostridia, such as *Clostridium acetobutylicum*, can be used for the industrial scale production of organic solvents like acetone, 1-butanol, ethanol and acetic and butyric acids through fermentation of different carbohydrate sources.[6]

To date 32 clostridial genomes have been analysed. The presence of secondary metabolite genes in this genus is quite widespread. Environmental isolates possess a larger number of secondary metabolite biosynthesis gene clusters and a more variable spectrum of cluster compositions than their human associated relatives, whilst cellulolytic clostridia possess the largest proportion of secondary metabolite DNA in their genomes. Only those clostridia that contain secondary metabolite biosynthesis gene clusters will be discussed here. For a full list of analysed genomes, see Table 2.

5.1.1 Solventogenic clostridia. The solventogenic clostridia are a group of organisms with biotechnological relevance for their ability to convert complex carbohydrates into economically useful end products, such as acetone, butanol and ethanol. Surprisingly, each of these organisms also contains secondary metabolite biosynthesis gene clusters.

Clostridium acetobutylicum – All analysed *C. acetobutylicum* strains possess one PKS gene (*ca_c3355* in *C. acetobutylicum* ATCC 834) with a KS-AT-DH-KR-PPTase encoded domain structure. The closest characterised homologues are MxaD, MxaE and StiC from the myxobacterium *Stigmatella aurantiaca*, with a range of 32–40% amino acid identity. These three proteins play roles in the biosynthesis of the potent electron transport chain inhibitors myxalamid and stigmatellin, with MxaD and MxaE being single module PKSs involved in myxalamid biosynthesis, while StiC is a single module PKS involved in stigmatellin biosynthesis.[77,78] All three *Stigmatella* proteins have the same domain structure as CA_C3355, however, in each case the genes for these proteins are part of much larger biosynthesis gene clusters, whilst *ca_c3355* appears to be an orphan. Standalone single module PKS genes, which encode an identical structure to *ca_c3355*, exist in various *Paenibacillus mucilaginosus* strains, which also have not been further characterized. However, in each case, there are other PKS related genes in the vicinity, including those coding for *trans*-ATs, β-lactamases and thioesterases. None of these types of genes are present either upstream or downstream of *ca_c3355*, although there are several different regulators and efflux pumps in the vicinity. It remains unclear at present whether this is an active PKS or a remnant of what was once a much larger secondary metabolite biosynthesis gene cluster.

Clostridium beijerinckii NCIMB 8052 – *Clostridium beijerinckii* NCIMB 8052, originally thought to be another strain of *C. acetobutylicum*, was isolated from garden soil in the 1920s and is a well-known solvent and organic acid producer.[79] It offers the advantage that organic waste products can be used as a carbon source, compared with *C. acetobutylicum*, which requires starch.[80] Within the genome of *C. beijerinckii* one secondary metabolite biosynthesis gene cluster was detected. Homology searches showed only weak amino acid identity (35–46% similarity) to the *nrsABCDEF* gene cluster of *Bacillus amyloliquefaciens* FZB42, a plant growth promoting rhizobacterium (Fig. 6).[81] The *nrs* gene cluster has much lower GC content than the *B. amyloliquefaciens* FZB42 genome (26–34% GC *vs.* 46% in the genome) and appears to be the only NRPS or PKS gene cluster acquired by horizontal gene transfer in this strain.[82] Interestingly, the GC content of the *C. beijerinckii* gene cluster also varies from 26–34%, suggesting that these two loci may be derived from the same ancestral cluster. The *B. amyloliquefaciens* FZB42 genome contains gene clusters coding for up to nine different compounds of PKS or NRPS origin, however, the *nrs* gene cluster is the only one without an associated compound.[81]

The two gene clusters share a number of similarities: a) both clusters encode a three module NRPS (NrsC and Cbei_0250) with three cysteine-specific A domains; b) both clusters encode a CAL-domain (specific for the incorporation of an fatty acid moiety), as part of *cbei_0250* in *C. beijerinckii* and as a standalone gene (*nrsF*) in *B. amyloliquefaciens* FZB42; c) both clusters contain accessory genes, encoding a thioesterase (*nrsA* and

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Table 2 Analysed genomes from the class Clostridia (genus *Clostridium*)

Strain[a]	Genome size (Mb)	PKS	NRPS[b]	Hybrid	Size (kb) (%) of genome	Reference
Clostridium acetobutylicum DSM 1731	3.94	1			5.39 (0.14)	83
Clostridium acetobutylicum ATCC 824	3.94	1			5.39 (0.14)	84
Clostridium acetobutylicum EA 2018	3.94	1			5.39 (0.14)	85
Clostridium beijerinckii NCIMB 8052	6.00			1	17.24 (0.29)	80
Clostridium botulinum A2 BoNT/A2 Kyoto-F	4.16		2*		36.95 (0.89)	86
Clostridium botulinum A BoNT/A1, ATCC 19397	3.86		1*		1.83 (0.05)	86
Clostridium botulinum A BoNT/A1 Hall	3.76		1*		1.83 (0.05)	86
Clostridium botulinum BoNT/B1 Okra	3.96		1*		1.83 (0.05)	86
Clostridium botulinum BoNT/A3 Loch Maree	3.99		1*		1.83 (0.05)	86
Clostridium botulinum Ba4 657	3.98		1*		1.83 (0.05)	86
Clostridium botulinum B925KT015+	2.77				—	70
Clostridium botulinum F 230613	3.99		1*		1.83 (0.05)	87
Clostridium botulinum H04402 065	3.93		2*		30.74 (0.78)	88
Clostridium botulinum A str. ATCC 3502	3.76				1.83 (0.05)	89
Clostridium cellulolyticum H10	4.07		1	4	199.58 (4.90)	90
Clostridium cellulovorans ATCC 35296	5.26		5	2	132.99 (2.53)	91
Clostridium clariflavum EBR 45, DSM 19732	4.90	1	1		29.13 (0.59)	92
Clostridium difficile 630 X	4.29		1**		2.19 (0.05)	93
Clostridium difficile CD196	4.11		1**		2.19 (0.05)	94
Clostridium difficile R20291	4.19		1**		2.19 (0.05)	94
Clostridium kluyveri DSM 555	3.96			3	51.36 (1.30)	8
Clostridium ljungdahlii DSM 13528+	4.63				—	95
Clostridium novyi NT+	2.55				—	96
Clostridium perfringens ATCC 13124+	3.26				—	69
Clostridium perfringens SM101+	2.90				—	69
Clostridium perfringens 13+	3.03				—	97
Clostridium tetani Massachusetts E88+	2.80				—	98
Clostridium thermocellum DSM 1313	3.56	1	1		15.65 (0.44)	99
Clostridium thermocellum ATCC 27405	3.84	1	1		15.65 (0.41)	99
Clostridium saccharolyticum WM1, DSM 2544+	4.66				—	90
Clostridium sticklandii DSM 519+	2.72				—	100
Clostridium sp. SY8519	2.84		1*		3.22 (0.11)	101

[a] Strains marked with + do not contain NRPS or PKS encoding DNA [b] Conserved domain structure: *A-T, **A.

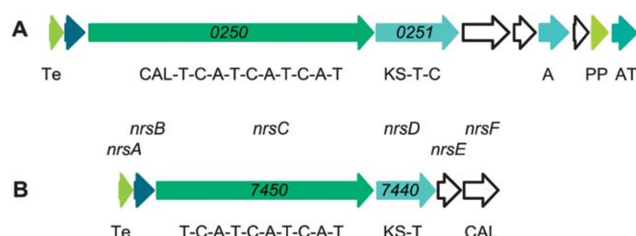


Fig. 6 Comparison of a homologous secondary metabolite biosynthesis gene cluster of *C. beijerinckii* NCIMB 8052 (*cbei_xxxx*) **A** and *B. amyloliquefaciens* (*rbam_02xxxx*) **B**. Identical colours indicate similar gene functions, white arrows indicate no similarity. Numbers represent the locus tag for each gene within the genome sequence of each organism.

cbei_0248), an McbC-like-oxidoreductase (*nrsB* and *cbei_0249*) and an acyl-CoA synthetase (*nrsD* and *cbei_0253*); d) a hybrid PKS-NRPS component encoded by both gene clusters (NrsD and Cbei_0251) contain KS and T domains. Additionally the *C. beijerinckii* PKS gene (*cbei_0250*) encodes a *trans*-AT binding site (with a *trans*-AT domain encoded by *cbei_0257*) unlike the *B. amyloliquefaciens* cluster where these domains are absent. Furthermore, the *C. beijerinckii* gene cluster encodes a stand-alone A domain, which is most likely linked to the C domain of Cbei_0251.

5.1.2 Cellulolytic clostridia. The cellulolytic clostridia are a group of well-studied organisms, mainly due to their ability to degrade cellulose rich feed stock and convert it into valuable solvents, such as acetone, butanol and ethanol, also known as ABE fermentation.[102] Although the main interest in these organisms is in engineering strains for better cellulolytic activity and higher ethanol yields, they also possess a variety of interesting secondary metabolite biosynthesis gene clusters, some of which show interspecies homology.

Clostridium cellulovorans – Clostridium cellulovorans 743B, isolated from a woodchip pile, produces an extracellular enzyme complex, known as a cellulosome, which efficiently

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Table 3 PKS, NRPS and hybrids encoded in the *C. cellulovorans* genome

Cluster No.	Type of modules	No. modules	Size (kb)
1	NRPS	2 NRPS	6.35
2	NRPS	2 NRPS	8.63
3	NRPS	1 NRPS	2.49
4	NRPS	3 NRPS	6.59
5	NRPS	18 NRPS	65.71
6	Hybrid (*trans*-AT)	1 PKS, 2 NRPS	9.50
7	Hybrid (*trans*-AT)	1 PKS, 9 NRPS	33.72

degrades plant cell wall polysaccharides and provides an avenue to biomass conversion for biofuel production.[91] Within the 5.26 Mb genome, five NRPS and two PKS-NRPS hybrid gene clusters were detected, making up almost 2.5% of the entire genome.

Several small NRPS gene clusters are present within the *C. cellulovorans* genome. The first of these (cluster 1, Table 3) is a biomodular NRPS gene cluster (*clocel_0725* & *0726*) encoding the domain structure A-T and C-A-T, respectively. A similar gene cluster was also found in *Clostridium thermocellum* strains (see below). Genes surrounding this cluster appear to be involved in carbohydrate metabolism, and it is possible that a sugar moiety may be attached to the peptide, given the presence of two glycosyltransferase genes within the gene cluster.

A further NRPS gene cluster (cluster 2) consists of two NRPS genes, one encoding A-E-C domains (*clocel_2027*) and the other encoding a CAL-C didomain (*clocel_2026*). Surrounding genes appear to encode products involved in carbohydrate metabolism, such as an α-L-arabinofuranosidase (*clocel_2020*), β-galactosidase (*clocel_2022*) and a glycosyltransferase (*clocel_2024*). Other genes in the vicinity encode proteins with cellular stress response functions, suggesting that the resultant metabolite could be involved in a response to environmental stresses – for example in the case of oxidative DNA damage.

C. cellulovorans also contains a single A-T encoding gene (*clocel_2279*, cluster 3), with weak similarity to those present in *C. difficile* and *C. botulinum.* Unusually, two genes within the vicinity (*clocel_2286*, *clocel_2287*) are annotated as containing a TIR2 domain, which is a bacterial homologue of the Toll-like receptor (TLR) domain. In humans, TLRs are responsible for initiating the innate immunity signalling cascade leading to defence against invading pathogens.[103,104] The role of these domains in bacterial proteins has not yet been identified.

A small gene cluster (cluster 4), represented by the genes *clocel_3123*, *clocel_3124* and *clocel_3129*, encodes two standalone A domains, and a T-C-A-T domain structure, respectively. BLAST analysis showed no similarity to any known genes or gene clusters.

A putative peptide producing NRPS system with 18 modules encoded by five genes is also present in *C. cellulovorans* (cluster 5, *clocel_2482*, *clocel_2486*-*clocel_2489*). Homology searches revealed partial similarity to other NRPSs, such as those involved in the biosynthesis of the decapapetides tyrocidine and gramicidin. Multiple transporters and amino acid modifying enzymes are also located within the gene cluster. Amongst the multiple regulators present is an AsnC family transcriptional regulator (Clocel_2477). This ubiquitous family of regulators plays roles in virulence and persistence, and in response to amino acid starvation in a variety of different bacteria.[105]

Two hybrid gene clusters were found within the genome of *C. cellulovorans*. Both clusters putatively code for *trans*-AT NRPS-PKS systems. The first hybrid gene cluster (cluster 6) consists of two NRPS genes, encoding a standalone A-domain (*clocel_2628*), a C-A-T-C domain structure (*clocel_2626*) and a single small PKS component encoded by *clocel_2634*. The cluster encodes genes for multiple small PPTases and multiple acyl-CoA dehydrogenases, however, these genes do not exhibit any similarity to known genes or gene clusters, according to BLAST analysis.

The PKS components of a second hybrid gene cluster (cluster 7) are represented by the genes *clocel_3942* and *clocel_3943* encoding only a single KS and *trans*-AT domain, respectively. Nine NRPS modules are encoded by four further genes (*clocel_3933*; *clocel_3940*; *clocel_3944*; *clocel_3945*). Although a terminal TE domain is absent from the NRPS proteins, two β-lactamase encoding genes (*clocel_3938*; *clocel_3941*) are present within the gene cluster and are most likely responsible for chain release. Upstream of this cluster is an *fkbH* orthologue, *clocel_3946*. FkbH is suspected to play a role in the formation of an unusual PKS extender unit by tethering a glycolytic intermediate to an ACP, in the case of both ascomycin (*Streptomyces hygroscopicus*)[106] and zwittermicin A (*Bacillus cereus*) biosynthesis.[107]

Clostridium thermocellum – *Clostridium thermocellum* is a thermophilic cellulolytic bacterium with one of the highest rates of cellulose hydrolysis reported so far.[99] The cellulosome of this organism has been well-studied, and recent technological advances have shown that it is possible to modify both the cellulosome and the primary metabolism of this organism, with the potential to improve ethanol yields.[108,109] Two strains of *C. thermocellum* have been sequenced, and both share two secondary metabolite gene clusters - one NRPS and one type I PKS gene cluster.

The NRPS gene cluster contains two genes encoding modules with the domain architecture A-T (*clo1313_2580*) and C-A-T-Te with a C-terminal membrane spanning transporter domain (*clo1313_2577*). This cluster is similar to cluster 1 from *C. cellulovorans*. However, while two glycosyltransferase genes are also present in this cluster, as in *C. cellulovorans*, the gene organisation is different. The glycosyltransferase genes in *C. thermocellum* are located between the two NRPS encoding genes and not upstream as is the case in *C. cellulovorans* cluster 1. The remaining surrounding genes also differ from those of *C. cellulovorans* and give little information as to the potential product of this gene cluster in *C. thermocellum*.

The type I PKS gene cluster found in *C. thermocellum* (*clo1313_2097*) encodes a single module with KS-AT-KS-ACP-ACP-KR domains. The gene shows only low similarity to functionally characterised orthologues. Analysis of this cluster shows that it has features of both iterative type 1 PKS and PKS-like polyunsaturated fatty acid synthase gene clusters.[110] Iterative type I PKSs are commonly found in fungal polyketide biosynthesis, but are rare in bacteria, where they have been shown to be involved in enediyne biosynthesis.[16] However, in another study is was anticipated that the gene cluster codes for

 This journal is © The Royal Society of Chemistry 2013

View Article Online

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

polyunsaturated fatty acid (PUFA) biosynthesis (see Section 7.1.1).[110] PKS-like genes in these gene loci have so far not been functionally analysed, and with no compound isolated from *C. thermocellum* that could be linked to the encoded pathway, the jury is still out as to its true nature.

Clostridium clariflavum DSM 19732 - *Clostridium clariflavum* DSM 19732 was isolated from thermophilic anaerobic sludge and is closely related to *C. thermocellum*.[92] As such, it also possesses the ability to degrade lignocellulosic material with high efficiencies.[92] *C. clariflavum* has two secondary metabolite biosynthesis gene clusters. The first gene cluster codes for a type I PKS/PUFA system that is analogous to that found in *C. thermocellum* (described above). The second gene cluster encodes a *trans*-AT PKS-NRPS hybrid. Five core genes encode five PKS modules and two NRPS modules (*clocl_2658* to *clocl_2665*). A *trans*-AT is encoded by *clocl_2664*, and a CAL domain encoding gene is located upstream of the cluster. Homology of individual genes to functionally characterised orthologues is weak, and the cluster as a whole has no functionally characterized counterpart.

Clostridium cellulolyticum H10 – *Clostridium cellulolyticum* H10 is one of the best-studied mesophilic cellulolytic clostridia, due mainly to the presence of a highly active extracellular cellulosome.[111] *C. cellulolyticum* is the producer of closthioamide[9] (Fig. 7), the first antibiotic isolated from an obligate anaerobe, which showed that anaerobes are capable of producing secondary metabolites. Successful induction of this compound was achieved by simulating the natural habitat through the addition of soil extract to the cultures.[9,112] The unusual symmetric thioamide structure of this compound has been verified by total synthesis,[113] and the synthesis of the oxygen analogue, closamide, highlighted the importance of the thioamide moieties for activity against MRSA and VRE.[9,113] The generation of a *C. cellulolyticum* mutant that overexpresses a *nusG* antiterminator gene resulted in constant production of closthioamide and led to the isolation of seven new thioamides (closthioamides B–H), apparently intermediates and shunt products of the biosynthetic pathway.[114] (Fig. 7) Nonetheless, the corresponding biosynthesis gene cluster for this compound remains elusive and is still under investigation.

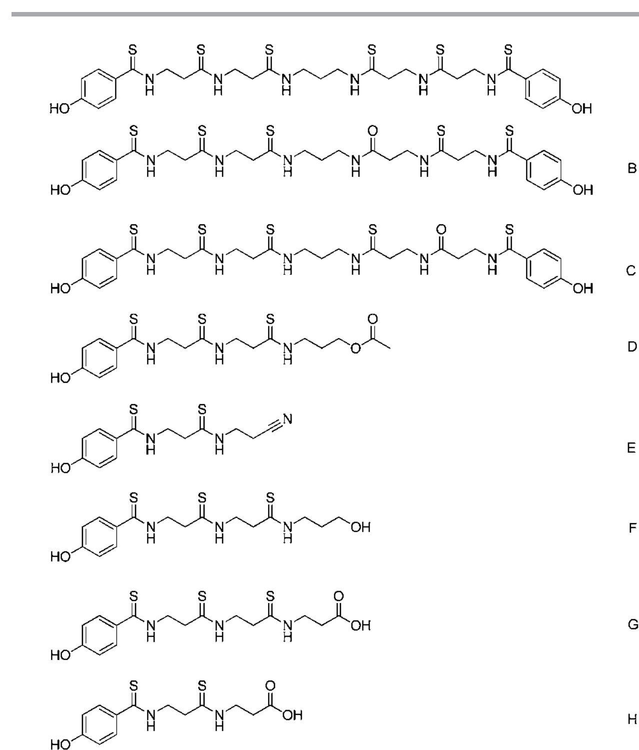


Fig. 7 Structure of closthioamide (top) – the first and only known antibiotic isolated from obligate anaerobic bacteria. Closthioamides B–H are likely intermediates or shunt products of the closthioamide biosynthetic pathway.

Five secondary metabolite biosynthesis gene clusters were identified in *C. cellulolyticum* (Table 4), which occupy nearly five percent of the genome, the largest percentage of all anaerobes analysed in this study. Unlike in other clostridia, the majority of gene clusters in *C. cellulolyticum* are greater than 60 kb in size, with multiple type I PKS or NRPS genes. Furthermore, four of the five clusters contain hybrid PKS-NRPS genes. The secondary metabolite gene clusters of *C. cellulolyticum* show a remarkable level of conservation with gene loci in other clostridia assuming an important role of the corresponding metabolites.

A large hybrid *trans*-AT PKS gene cluster coding for 14 PKS modules and a single NRPS module is located between *ccel_0858* and *ccel_0864* (cluster 1). The gene cluster has some unusual features with three genes for standalone KS domains downstream of the final PKS gene and two single genes for *trans*-ATs (*ccel_0852*, *ccel_0854*) encoded upstream of the PKS and NRPS encoding genes. Additionally, a KSIII is encoded downstream of the PKS genes by *ccel_0874*. Individual gene products show low similarity (∼40%) to PKS genes of *C. papyrosolvens* and functionally characterized genetic loci like those responsible for macrolactin (*mln*) and bacillaene (*bae*) of *Bacillus amyloliquefaciens* FZB42. Multiple radical SAM domain encoding genes are also present within the cluster. Individual genes of this cluster show only weak similarity to other known genes or gene clusters, and the product of the encoded synthase remains unknown.

C. cellulolyticum contains an unusual gene cluster (cluster 3) that consists of four genes, three genes encode one C domain (*ccel_1210*, *ccel_1213* and *ccel_1217*) each and a fourth one adenylation domain (*ccel_1218*). A thiolation domain is absent, however, KSIII (*ccel_1215*) and TE (*ccel_1211*) encoding genes were found in the surrounding area. As with many other clostridial gene clusters, the genes are flanked by a number of transposase and integrase genes, suggesting acquisition by horizontal gene transfer. Interestingly, an orthologous gene

Table 4 PKS, NRPS and hybrid PKS-NRPS encoded in the genome of *C. cellulolyticum*

Cluster no.	Type	No. of modules	Size (kb)
1	Hybrid (*trans*-AT PKS)	14 PKS, 1 NRPS	63.75
2	Hybrid (*trans*-AT PKS)	3 PKS, 8 NRPS	48.10
3	NRPS	1 A, but 3 C domains	5.77
4	Hybrid (*trans*-AT PKS)	2 PKS, 2 NRPS	20.97
5	Hybrid (*trans*-AT PKS)	6 PKS, 9 NRPS	60.98

View Article Online

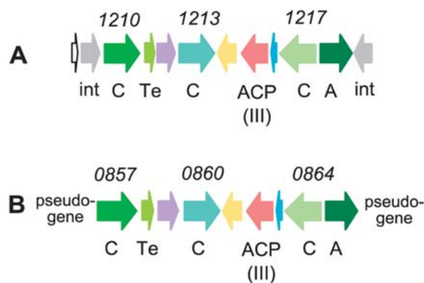


Fig. 8 Conservation of *C. cellulolyticum* cluster 3 **A** (*ccel_xxxx*) in *Clostridium* sp. BNL 1100 **B** (*clo1100_xxxx*). Identical colours indicate similar gene functions, white arrows indicate no similarity. Numbers represent the locus tag for each gene within the genome sequence of each organism. Int, integrase.

cluster is present in the draft genome sequence of *Clostridium* sp. BNL1100 (*clo1100_0857–clo1100_0865*), with all genes being highly conserved and completely syntenic (Fig. 8). At the flanking regions, where transposase and integrase genes are located in the *C. cellulolyticum* genome, pseudogenes were found in *Clostridium* sp. BNL1100. Furthermore, these genes appear to be part of a larger secondary metabolite biosynthesis gene cluster in *Clostridium* sp. BNL1100, suggesting that these genes may be remnants of a truncated gene cluster in *C. cellulolyticum*, or that this cluster has undergone adaptive evolution in *C. cellulolyticum.*

A further gene cluster (cluster 4) is also conserved between *C. cellulolyticum*, *Clostridium* sp. BNL 1100 and *C. papyrosolvens* DSM 2782. Both gene content and organisation are highly conserved between the three gene clusters (Fig. 9), however no similarities to functionally characterised genes were detected by BLAST analysis. The locus consists of three core genes in which *ccel_2330* is a KS-AT-ACP encoding PKS gene, *ccel_2329* is a two module NRPS encoding gene and *ccel_2331* represents a hybrid *trans*-AT-NRPS encoding gene. Upstream of the *Clostridium* sp. BNL 1100 and *C. papyrosolvens ccel_2330* orthogolues is an additional gene encoding an A domain, which is absent in *C. cellulolyticum.*

A large *trans*-AT PKS-NRPS gene cluster (cluster 5) consisting of nine genes encoding six PKS and nine NRPS modules makes up the final secondary metabolite locus in *C. cellulolyticum*. It is located downstream of gene cluster number 4. It remains unclear if these two clusters are independent from each other or if the close proximity suggests a connection. A *trans*-AT is encoded by *ccel_2380* and only a few genes encoding tailoring enzymes surround this region. The product of this gene cluster is unknown, and homology searches failed to give further information about a potential product.

5.1.3 Human-associated clostridia. Secondary metabolite genes within the human pathogenic clostridia are not as common as those in environmental isolates. Although 50% of the analysed clostridial genomes are from human pathogens and they make up 50% of the total clostridial genomic DNA, they account for less than 20% of the total secondary metabolite DNA of the clostridia. As such, the gene clusters within these pathogens are often smaller than those identified within their environmentally isolated relatives. This confirms previous observations showing that pathogenic clostridia contain only limited PKS DNA, compared to non-pathogenic clostridia.[115]

Clostridium difficile – *C. difficile* represents an increasing nosocomial health problem, with an estimated 500 000 cases per year of hospital acquired diarrhoea linked to the pathogen in the US alone.[116] Recent increases in morbidity and mortality due to *C. difficile* infections have spurned a growing interest in sequencing multiple genomes for the purposes of comparative genomics. At present, three *C. difficile* genomes have been completed and published, however, another 13 are in the draft stage and more than 150 are planned. All completed and draft *C. difficile* genomes contain a single A domain encoding gene, which is highly conserved. The region surrounding this A-domain encoding gene exhibits an extremely high level of conservation amongst those strains that possess it (>98% identity and complete synteny), perhaps indicating an important function for this organism.

Clostridium botulinum – *C. botulinum* is a ubiquitous inhabitant of soil and produces one of the world's most lethal toxins.[117] The flaccid paralysis caused by ingestion of the ribosomally produced botulinum toxin can eventually result in death, however, it is this same paralysis-inducing property that has been harnessed by the cosmetic industry.[118] At present, ten *C. botulinum* genomes have been completed and published. All analysed strains, except *C. botulinum* H04402 065, share a similar feature: a single A domain encoding gene – like the *C. difficile* strains – but containing an additional T domain. Amino acid identity of the A-T didomain is high amongst *C. botulinum* strains that possess it (>95%), and the genomic neighbourhood is syntenic amongst these strains. However, the A-T didomain shows limited similarity to A domains encoded by the *C. difficile* strains. Amongst the completed *C. botulinum*

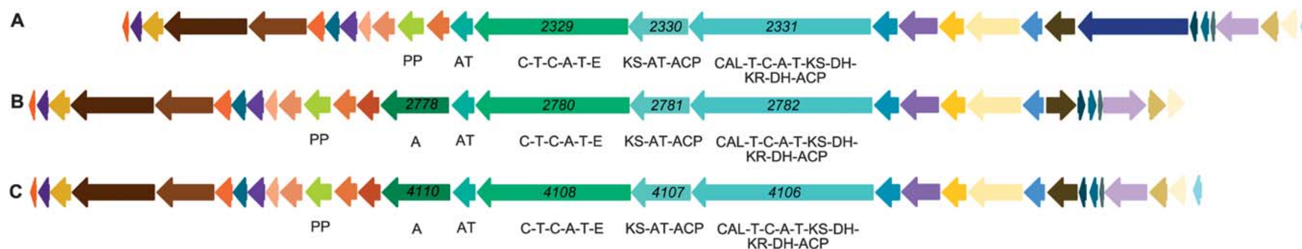


Fig. 9 Comparison of a *C. cellulolyticum* PKS-NRPS hybrid gene cluster **A** (cluster 4, *ccel_xxxx*) conserved in other *Clostridium* BNL 1100 **B** (*clo1100_xxxx*) and *Clostridium papyrosolvens* **C** (*cpap_xxxx*). Identical colours indicate similar gene functions. Numbers represent the locus tag for each gene within the genome sequence of each organism.

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

genomes, strain H04402 065 and strain A2 Kyoto each harbour a large NRPS gene cluster consisting of four NRPS genes coding for eight modules and nine modules, respectively (Fig. 10). The majority of the genes are highly conserved between the two strains, with one major exception being *clm_0375*, which is absent in *C. botulinum* H04402 065 (Fig. 10). Siderophore-related transporters are encoded upstream as well as downstream of the core genes suggesting this polypeptide plays a role in iron acquisition, and maybe a further virulence factor of these pathogenic clostridial strains.

Clostridium sp. SY8519 – *Clostridium* sp. SY8519 was isolated from the intestinal flora of a healthy adult[119] during attempts to identify bacteria that convert daidzein to *O*-desmethylangolensin. *O*-Desmethylangolensin is derived from isoflavonoids present in soybean or soy products as well as in different plants and herbs and has been suggested to have positive health benefits, including inhibition of cancer proliferation and lowering of blood triglyceride levels.[120,121]

The genome of strain SY8519 harbours an NRPS-like gene, which encodes single A and T domains, similar to *C. difficile* and *C. botulinum* strains, but BLAST analysis showed no similarity to those gene clusters. The gene product also contains a TD domain, a terminal reductive domain, believed to be involved in the release of a polyketide-peptide intermediate in myxalamid biosynthesis, with concomitant liberation of an aldehyde and subsequent reduction to an alcohol.[78]

5.1.4 Other Clostridia. *Clostridium kluyveri* DSM 555 – *Clostridium kluyveri* was isolated from the mud of a canal in Delft, Netherlands, and has the unusual ability of being able to grow with only ethanol and acetate as its sole energy sources.[122] Sequencing of the 3.96 Mb genome of *C. kulyveri* revealed four secondary metabolite gene clusters (Table 5).[8] *C. kulyveri* has been shown to produce a putative siderophore under iron-deficient growth conditions, which is thought to be linked to a gene cluster comprising genes *ckl_1504* to *ckl_1518* (cluster 1), due to its strong similarity in several components to gene clusters coding for yersiniabactin **1** biosynthesis (*Yersinia pestis*), pyochelin (*Pseudomonas aeruginosa*) and equibactin **2** (*Streptococcus equi*).[8] The *C. kluyveri* gene cluster also exhibits a high level of similarity to the equibactin biosynthesis gene cluster (*eqb*) in *Streptococcus equi*.[123] Although attempts at heterologous expression of the *eqb* locus have been made in *E. coli*, equibactin has never been isolated and its structure is still only a prediction. It has been suggested that the incorporation of cysteine residues and a subsequent heterocyclization leads to the formation of thiazoline rings as in yersiniabactin.[123]

Table 5 PKS, NRPS and hybrid PKS-NRPS gene clusters present in *C. kluyveri*

Cluster no.	Type	# modules	Size (kb)
1	NRPS	3 NRPS	14.64
2	Hybrid (*cis*-AT PKS)	3 PKS, 1 NRPS	14.91
3	Hybrid (*trans*-AT PKS)	2 PKS, 1 NRPS	11.10
4	Hybrid (*trans*-AT PKS)	1 PKS, 1 NRPS	10.71

Genomic analysis has shown that this gene locus also exists in other anaerobic bacteria (Fig. 11). The core genes consist of two NRPS genes with a different number of modules depending on the host bacterium. Sequence analysis suggests that they all share the incorporation of dihydroxybenzoic acid or salicylic acid as a starter unit and extension with cysteine residues. Interestingly, salicylate was shown to be required for equibactin production.[123] The number of incorporated cysteine residues likely differs from one to three in each strain, as do the surrounding genes, which are predominantly different kinds of transporters.

The high number of mobile genetic elements flanking this gene cluster suggests that the encoded pathway has a common origin and the encoded compound plays an important role in iron acquisition in many anaerobic bacteria. Although siderophores are often pathogenicity factors, the occurrence of this putative siderophore biosynthesis gene cluster is independent of the pathogenicity of these organisms. Downstream of this siderophore biosynthesis gene cluster is another secondary metabolite biosynthesis gene cluster (cluster 2, *ckl_1519*-*ckl_1528*) (Fig. 12). It consists of a hybrid NRPS-PKS gene (*ckl_1525*) and a PKS gene (*ckl_1527*). BLAST analysis showed 51–69% identity to several gene clusters present in other anaerobes and facultative anaerobes, respectively. The proximity of this cluster in *C. kluyveri* to the putative siderophore biosynthesis gene cluster described above raises the possibility that these may be linked and may act together to produce one compound. However, the *C. kluyveri* siderophore biosynthesis gene cluster (cluster 1) is absent in other anaerobic bacteria (C. sp DL-VIII, *A. cellulolyticus* CD2, *B. cytotoxicus*) that possess *C. kluyveri* cluster 2 orthologues. The purpose of this cluster and the compound it encodes remain unclear.

A small hybrid *trans*-AT PKS-NRPS gene cluster (cluster 3) is encoded by the genes *ckl_1730* - *ckl_1734*. This cluster contains several unusual elements, including genes for a loading module with a fatty acid attachment (CAL) domain and a gene encoding

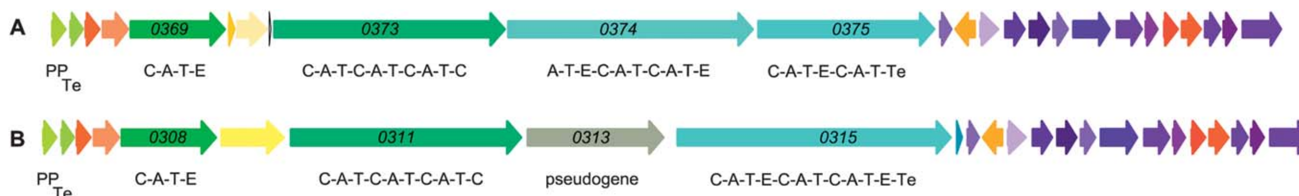


Fig. 10 Gene conservation amongst two NRPS gene clusters in *C. botulinum* strains. **A**, *C. botulinum* A2 str Kyoto (*clm_xxxx*); **B**, *C. botulinum* H04402 065 (*h04402_0xxxx*). Identical colours indicate similar gene functions. Numbers represent the locus tag for each gene within the genome sequence of each organism.

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

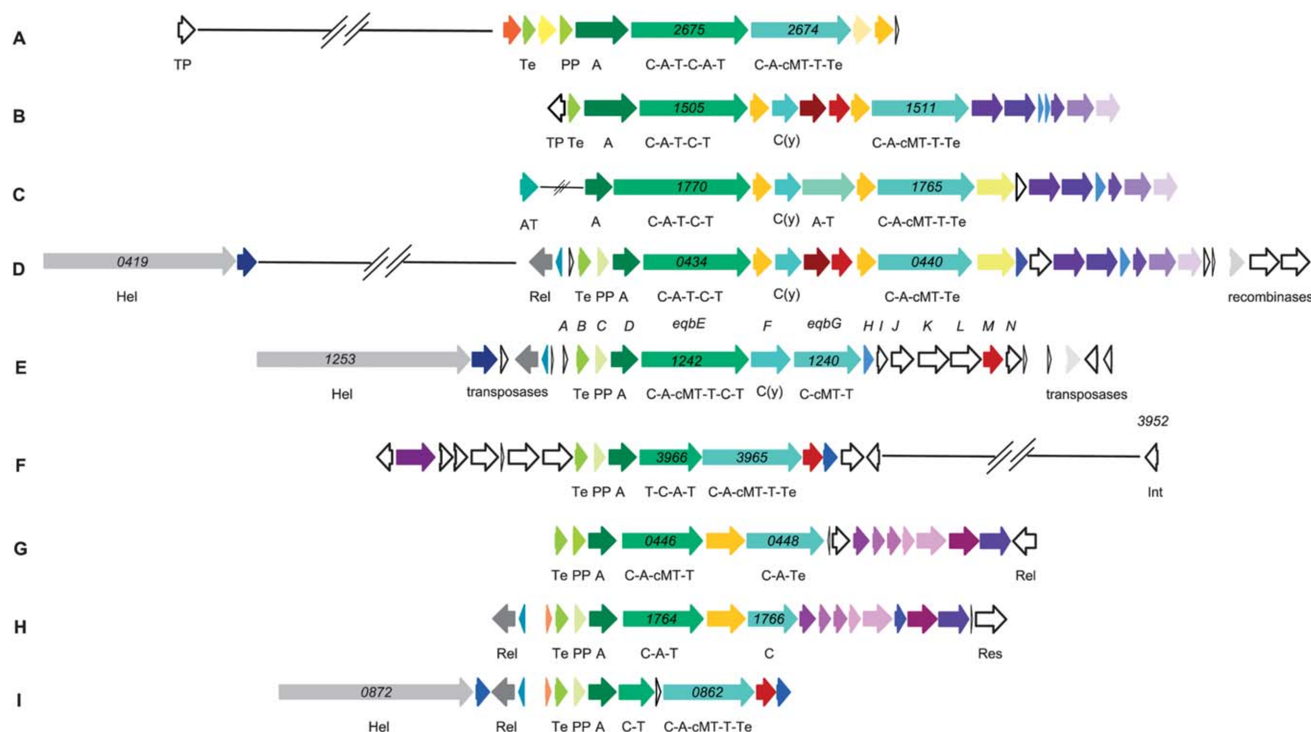


Fig. 11 Conservation of a putative siderophore biosynthesis gene clusters across different bacterial species. **A**, *Desulfotomaculum acetoxidans* DSM 771(*dtox_xxxx*); **B**, *Clostridium kluyveri* DSM 555 (*ckl_xxxx*); **C**, *Clostridium papyrosolvens* DSM 2782 (*cpap_xxxx*); **D**, *Anaerococcus vaginalis* ATCC 51170 (*hmpref0078_xxxx*); **E**, *Streptococcus equi* subsp. equi 4047 (*seq_xxxx*); **F**, *Desulfitobacterium hafniense* DCB-2 (*dhaf_xxxx*); **G**, *Eubacterium hallii* DSM 3353 (*eubhal_0xxxx*); **H**, *Anaerococcus prevotii* PC1 (*apre_xxxx*); **I**, *Finegoldia magna* BVS033A4 (*hmpref9289_xxxx*). Identical colours indicate similar gene functions, white arrows indicate no similarity. Numbers represent the locus tag for each gene within the genome sequence of each organism. TP, transposase; Hel, helicase; Rel, relaxase; Res, resolvase.

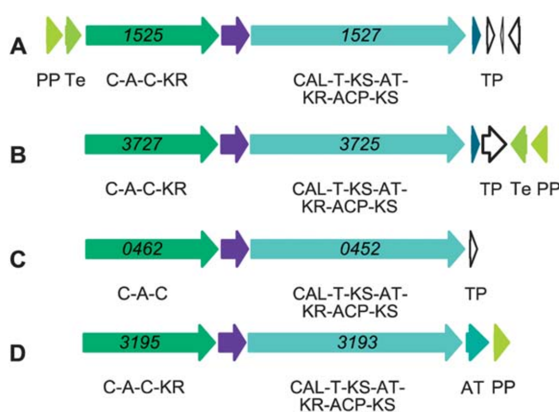
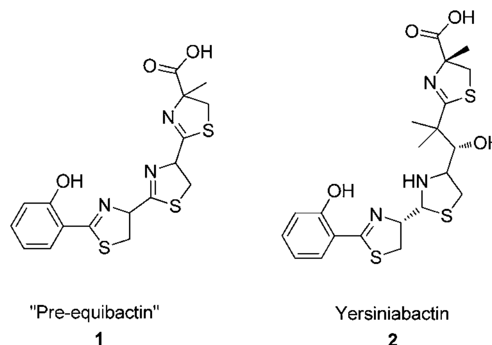


Fig. 12 Conservation of cluster 2 from *C. kluyveri* **A** (*ckl_xxxx*) in different obligate and facultative anaerobes. **B**, *Clostridium* sp DL-VIII (*cdlvIII_xxxx*); **C**, *Acetivibrio cellulolyticus* CD2 (*acelc_02010001xxxx*); **D**, *Bacillus cytotoxicus* NVH 3091-98 (*bcer98_xxxx*). Identical colours indicate similar gene functions, white arrows indicate no similarity. Numbers represent the locus tag for each gene within the genome sequence of each organism. TP, transposase.

a KR-ACP and β-lactamase domain (*ckl_1730*) for which there are no other homologues in the Genbank database.

The *trans*-AT encoding gene is situated in the middle of the cluster (*ckl_1732*), and several other genes, including those for a standalone TE (*ckl_1735*), a PPTase (*ckl_1736*) and efflux pumps (*ckl_1722*, *ckl_1729*) are also present. Particularly striking is the high density of mobile genetic elements located both upstream and downstream of the PKS/NRPS genes. BLAST analysis of the whole PKS/NRPS gene cluster showed no similarity to any known genes or gene cluster.



The fourth gene cluster present in *C. kluyveri* is similar to the gene cluster presented in the preceding paragraph, in that it is also a *trans*-AT PKS-NRPS hybrid cluster of similar length, containing one PKS and one NRPS module spread over three genes (*ckl_2354-ckl_2357*). A *trans*-AT (CKL_2356) is encoded as a part of the cluster, as are several regulators and transporters, suggesting an exported product. The genomic context and cluster organisation provide little information about a potential product, and the cluster shows no overall similarity to any known secondary metabolite gene clusters.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

5.2 Other members of phylum Firmicutes

5.2.1 Other members of the order Clostridiales

5.2.1.1 Family Heliobacteriacae. The Heliobacteriacae are anaerobic anoxygenic phototrophs, comprising four genera with a total of seven species. The unique photosynthetic antenna pigment bacteriochlorophyll (Bchl) g, the ability to form endospores as well as the inability to grow photoautotrophically, separates them from other phototrophs.[124]

Heliobacterium modesticaldum Ice1 was isolated from hot spring volcanic soils in Iceland and represents the first fully sequenced heliobacterial genome.[124] A type III polyketide synthase (HM1_0406) is encoded on the chromosome, possessing 42–48% amino acid identity to BcsA (also called BpsA), a type III PKS from *Bacillus* spp. In *Bacillus subtilis*, *bcsA* is present in an operon with *ypbQ* (also known as *bpsB*) and the products of these genes act together to produce a range of alkylpyrone methyl ethers.[125] An orthologue of *ypbQ* is present in a putative operon with *hm1_0406*, suggesting that this cluster functions in a similar manner to *bcsA-ypdQ* in *Bacillus* species.

5.2.1.2 Clostridiales Family XI incertae sedis*.* This family of anaerobic Gram-positive cocci includes five genera, which have been predominantly isolated from humans.[126] The type strain of the genus *Anaerococcus*, *A. prevotii* PC1, is an opportunistic pathogen, but has been also described as a member of the normal skin flora, as well as the oral cavity and the gut.[127] While the chromosome is devoid of secondary metabolite genes, the plasmid of *A. prevotii* PC1 contains an NRPS gene cluster similar to cluster 1 from *C. kluyveri* (Fig. 11).

5.2.1.3 Family Eubacteriaceae. Eubacterium limosum KIST612 is the type strain of the Eubacteriaceae family and a model organism for biofuel production from syngas (a mixture of H_2, CO_2 and CO). *E. limonsum* produces acetate, butyrate and ethanol from CO, even under high pressures.[128] The 4.2 Mb genome[128] contains one NRPS gene (*eli_3398*) with a C-A-T-C-A-T domain structure and a downstream PPTase encoding gene (*eli_3399*). Unusually, *eli_3399* also encodes a C-terminal short chain dehydrogenase/reductase domain, which form part of a larger family of NAD(P)(H) dependent oxidoreductases.[129] The product of *eli_3398* shares 51% amino acid identity with a similar NRPS from *Syntrophomonas wolfei* subsp. wolfei str. Goettingen. The most closely related functional orthologue is SubA from *Bacillus subtilis* JM4, which is involved in subpeptin production.[130] However, the genomic context is different in *E. limnosum*.

Eubacterium rectale ATCC 33656 is a butyrate-producing bacterium, isolated from human feces.[131] The 3.45 Mb genome[132] contains a two gene NRPS gene cluster encoding a single C (Eubrec_1027) and A domain (Eubrec_1029). The two genes show strong similarity to other uncharacterised C and A domain containing genes within the phylum Firmicutes, however, there are no functionally characterised orthologues.

5.2.1.4 Family Syntrophomonadaceae. As the name suggests, members of the Syntrophomonadaceae are syntrophs - chemoorganotrophs that metabolise fatty acids and live in close association with hydrogen-/formate-using microorganisms as a source of electron donors and acceptors.[133,134] *Syntrophomonas wolfei* subsp. wolfei str. Goettingen was isolated in 1981 from anaerobic digester sludge and has become a model organism for syntrophic fatty acid metabolism.[135] One NRPS gene (*swol_1084*) was identified within its genome, which encodes a C-A-T-C-A-T architecture, exactly the same domain structure as that described above for *Eubacterium limnosum*.

5.2.1.5 Family Lachnospiraceae. The Lachnospiraceae are a diverse group of Gram-positive anaerobes isolated mainly from human and animal intestines. Of the two completed and published genomes from this family, only *Cellulosilyticum lentocellum* contains genes for secondary metabolite biosynthesis.

Cellulosilyticum lentocellum RHM5, formerly known as *Clostridium lentocellum*, is a cellulolytic bacteria isolated from river sediment containing paper mill and domestic effluent.[136] The 4.71 Mb genome[137] contains one NRPS gene cluster in which *clole_3882* encodes an A-PCP-C-A-PCP architecture, and *clole_3880* encodes a stand-alone A domain. Within the cluster, genes for a stand-alone ACP and a PPTase are also present. Clole_3882 shows limited identity to other characterised NRPSs.

5.2.1.6 Family Ruminococcaceae. The Ruminococcaceae family consists of cellulolytic bacteria, mainly found in the intestinal tract of herbivores. To date, *Ruminococcus albus* 7 (ATCC 27210), isolated from cow rumen, is the only completed and published genome from the Ruminococcaceae. The *R. albus* genome[138] contains one NRPS-PKS hybrid gene cluster consisting of eleven genes (*rumal_0869* to *0885*). The encoded NRPS and PKS proteins have unusual structures, in that multiple methyltransferase domains are present within both NRPS and PKS modules and the PKS encoding genes lack recognisable AT domains or *trans*-AT domain binding sites. Furthermore, there are no obvious *trans*-AT genes present within the cluster. Three of seven encoded A-domains likely mediate cysteine incorporation, and a cysteine desulfurase encoding gene (*rumal_0895*) is located downstream of the NRPS genes. Multiple encoded thioesterases (Rumal_0865, 0874, 0880) and a PPTase gene (Rumal_0886) also form a part of this cluster. Neither individual genes, nor the cluster as a whole, show significant homology to any previously characterised clusters.

5.2.1.7 Family Oscillospiraceae. The Oscillospiraceae family contains a group of strict anaerobes with 16s rRNA homology to other strains inhabiting the alimentary tract of various animals.[139] The type strain of the genus *Oscillibacter*, *Oscillibacter valericigenes* SPN1, is the only sequenced member of this family so far. No secondary metabolite gene clusters were identified in the genome of *O. valericigenes* SPN1.

5.2.1.8 Family Peptococcaceae. The family Peptococcaceae was originally constructed to house a heterogenous group of Gram-positive anaerobic cocci, including the genera *Peptococcus*, *Peptostreptococcus* and *Ruminococcus*.[140] The family has since expanded to contain 14 recognised genera, including both cocci and rods found in environments ranging from soil to the human navel cavity.[141]

Desulfotomaculum acetoxidans strains are predominantly intestinal bacteria, which have been isolated from habitats such as animal manure, rumen content and dung-contaminated freshwater. The 4.5 Mb genome of strain DSM 771,[142] isolated from piggery waste in Germany, contains two secondary

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

metabolite biosynthesis gene clusters. An NRPS gene cluster (*dtox_2674-2676*) exhibits similar gene organisation, however, relatively weak amino acid identity, to the putative siderophore biosynthesis cluster from *Clostridium kluyveri* (Fig. 11). Furthermore, *D. acetoxidans* also possesses a PKS gene (*dtox_1322*) with KS-AT-DH-KR-ACP domain structure, which shows limited similarity to other functionally characterised genes.

Desulfitobacterium hafniense is another member of the Peptococcaceae family. This bacterium is able to dehalogenate organic compounds, and has been proposed as a potential bioremediation agent.[143] Strain DCP-2 and Y51 were isolated from municipal sludge and soil contaminated with tetrachloroethene, respectively.[144,145] Both strains contain an orthologous cluster to cluster 1 of *C. kluyveri*, as for *Desulfotomaculum acetoxidans.* (Fig. 11).

Syntrophobotulus glycolicus DSM 8271 was isolated from anoxic sewage sludge in Germany and is able to oxidize glyoxylate under anaerobic conditions.[146] The genome contains two NRPS genes, each encoding one module. Sgly_0842 shows a C-A-T domain structure and a polyprenyl synthetase (Sgly_0840) is encoded downstream of this NRPS. Similar genes could be found in *Desulfosporosinus* strains (Peptococcaceae family) with identities up to 46%. Sgly_3165 contains a C-A-T and a C-terminal N-terminal short chain dehydrogenase/reductase domain, which is most closely related to the final NRPS module of NcpB from *Nostoc* sp. ATCC 53789, involved in nostocyclopeptide synthesis.[147] However, the nostocyclopeptide biosynthesis gene cluster is far larger than the small gene locus found in *S. glycolicus*, and the resultant product of Sgly_3165 remains unknown.

5.2.2 Order Thermoanaerobacterales. Bacteria within the order Thermoanaerobacterales are of great interest due to their ability to convert higher sugars into ethanol. All of these organisms are thermophiles, with members of the genus *Caldicellulosiruptor* being the most thermophilic cellulose degraders known.[148] Within this order, the genomes of four *Thermoanaerobacter* species, eight *Caldicellulosiruptor* species, and single representatives from the genera *Moorella*, *Carboxydothermus*, *Thermosediminibacter*, and *Mahella* have been completed (Table 5). Despite their relatively close relationship to *Clostridium*, no secondary metabolite biosynthesis gene clusters were identified in any of the analysed genomes from this order.

5.2.3 Order Natranaerobiales. Only one genome has so far been completed from this group of polyextremophiles, which thrive in the combination of hot, alkaline and saline environments.[149] No secondary metabolite biosynthesis gene clusters were found in the only sequenced member of this order, *Natranaerobius thermophilus* JW/NM-WN-LF (Table 5).

5.2.4 Order Halanaerobiales. This order comprises a group of halophilic anaerobes, which favour salt concentrations up to 20%. Three genera are represented by the four sequenced genomes within this order, however, no secondary metabolite genes were found in Halanaerobiales (Table 6).

5.2.5 Other Firmicutes. Members of the class Negativicutes represent a group of taxonomically controversial anaerobes that possess a double-bilayer, Gram-negative cell wall.[150] Despite their phenotypic differences, 16s rRNA sequencing has shown these organisms to be more closely related to Gram-positive bacteria, hence their current position within the phylum Firmicutes.[150] To date, three genomes from representatives of the two families within the Negativicutes have been completed and published, although no secondary metabolite biosynthesis gene clusters were detected in these genomes (Table 6).

6 Phylum Bacteroidetes

The phylum Bacteroidetes contains a large and diverse range of Gram-negative, non sporulating, anaerobic rods with habitats that vary from soil to the mammalian intestine. Fourteen organisms from this phylum have had their genomes sequenced, ranging from 2.34–6.29 Mb in size, and include members of the families Bacteroidaceae (opportunistic pathogens that inhabit human and animal intestines), Porphyromonadaceae and Prevotellaceae (both of which contain bacteria that colonise the human oral cavity, occasionally causing periodontal disease). Based on the analysis conducted here, secondary metabolic potential appears very limited amongst Bacteroidetes (Table 7).

Bacteroides fragilis strains contain an unusual chromosomally located gene (*bf2961* in *B. fragilis* YCH46), which encodes an A-T didomain protein, in conjunction with putative siderophore transporter genes and a number of regulatory genes. Conservation of this A-T encoding gene amongst two *B. fragilis* isolates was previously shown[3] and investigation of this region in completed and draft *Bacteroides* genomes shows highly conserved amino acid sequences of the encoded proteins amongst *B. fragilis* and several other *Bacteroides* species. Region synteny is well conserved in *B. salanitronis*, although amino acid identity is significantly lower than amongst *B. fragilis* strains. No other clues are given to the function of this gene cluster from *in silico* analyses. However, the location of a putative siderophore receptor gene within the cluster suggests that the product may play a role in iron binding or sequestration.

A further putative A-domain encoding gene in *B. fragilis* strains has also been previously identified.[3] *B. fragilis* genomes contain four such genes and conserved domain searches using the encoded proteins show, in each case, higher homology to long-chain acyl-CoA synthetases involved in lipid biosynthesis than to true A-domain containing proteins. Thus, it is likely that these proteins are involved in fatty acid biosynthesis rather than secondary metabolite production.

7 Phylum Proteobacteria

Sequenced anaerobe genomes occur only in the delta (δ-), epsilon (ε-) and gamma (γ-) subdivisions of the phylum Proteobacteria. In total, there are currently 24 completed and published genomes, of which 14 possess secondary metabolite biosynthesis gene clusters. Several orthologous clusters predicted to be involved in polyunsaturated fatty acid (PUFA) biosyntheis *via* a PKS-like mechanism are found in diverse members of this phylum and will be discussed together, following which only

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Table 6 Analysed genomes from the phylum Firmicutes (other than the genus *Clostridium*)

Strain[a]	Family	Genome size (Mb)	PKS	NRPS	Hybrid	Size (kb) (%) of genome	References
Heliobacterium modesticaldum str. Ice1	Heliobacteriacae	3.08	1			1.13 (0.04)	124
Finegoldia magna ATCC 29328[+]	Clostridiales Family XI. *Incertae Sedis*	1.99				—	151
Anaerococcus prevotii PC1 DSM 20548	Clostridiales Family XI. *Incertae Sedis*	1.88		1		8.85 (0.47)	126
Symbiobacterium thermophilum IAM 14863[+]	Clostridiales Family XVIII. *Incertae Sedis*	3.57				—	152
Eubacterium eligens ATCC 27750[+]	Eubacteriaceae	2.83				—	132
Eubacterium limosum KIST612	Eubacteriaceae	4.28		1		7.33 (0.17)	128
Eubacterium rectale ATCC 33656	Eubacteriaceae	3.45	1			2.93 (0.08)	132
Syntrophothermus lipocalidus DSM 12680[+]	Syntrophomonadaceae	2.41				—	153
Syntrophomonas wolfei subsp. wolfei str. Goettingen	Syntrophomonadaceae	2.94		1		7.92 (0.27)	133
Butyrivibrio proteoclasticum B316[+]	Lachnospiraceae	4.4				—	154
Cellulosilyticum lentocellum RHM5, DSM 5427	Lachnospiraceae	4.71		1		7.93 (0.17)	137
Ruminococcus albus 7 (ATCC 27210)	Ruminococcaceae	3.69			1	35.16 (0.95)	138
Oscillibacter valericigenes Sjm18-20	Oscillospiraceae	4.47				—	155
Desulfotomaculum reducens MI-1[+]	Peptococcaceae	3.61				—	156
Desulfotomaculum acetoxidans DSM 771	Peptococcaceae	4.55	1	1		20.67 (0.45)	142
Candidatus Desulforudis audaxviator MP104C[+]	Peptococcaceae	2.35				—	157
Desulfitobacterium hafniense DCP-2	Peptococcaceae	5.28		1		10.87 (0.21)	143
Desulfitobacterium hafniense Y51	Peptococcaceae	5.73		1		10.72 (0.19)	158
Pelotomaculum thermopropionicum SI[+]	Peptococcaceae	3.03				—	159
Syntrophobotulus glycolicus FIGlyR, DSM 8271	Peptococcaceae	3.41		2		8.67 (0.25)	146
Thermincola potens JR[+]	Peptococcaceae	3.16				—	160
Thermoanaerobacterium thermosaccharolyticum DSM 571[+]	Thermoanaerobacteraceae	2.79				—	90
Thermoanaerobacter pseudethanolicus 39E[+]	Thermoanaerobacteraceae	2.36				—	161
Thermoanaerobacter sp. X514 [+]	Thermoanaerobacteraceae	2.46				—	90
Thermoanaerobacter italicus Ab9, DSM 9252	Thermoanaerobacteraceae	2.45				—	90
Thermoanaerobacter mathranii mathranii A3, DSM11426[+]	Thermoanaerobacteraceae	2.31				—	90
Thermoanaerobacter sp. X513[+]	Thermoanaerobacteraceae	2.46				—	90
Thermoanaerobacter brockii Ako-1, DSM 3389[+]	Thermoanaerobacteraceae	2.34				—	90
Thermoanaerobacter tengcongensis MB4T[+]	Thermoanaerobacteraceae	2.69				—	162
Caldicellulosiruptor lactoaceticus 6A, DSM 9545[+]	Thermoanaerobacteraceae	2.63				—	148
Caldicellulosiruptor bescii Z-1320, DSM 6725[+]	Thermoanaerobacteraceae	2.93				—	163
Caldicellulosiruptor saccharolyticus DSM 8903[+]	Thermoanaerobacteraceae	2.97				—	164
Caldicellulosiruptor obsidiansis OB47[+]	Thermoanaerobacteraceae	2.53				—	165
Caldicellulosiruptor hydrothermalis 108[+]	Thermoanaerobacteraceae	2.77				—	148
Caldicellulosiruptor owensensis 108[+]	Thermoanaerobacteraceae	2.43				—	148
Caldicellulosiruptor kristjanssonii 177R1B, DSM 12137[+]	Thermoanaerobacteraceae	2.80				—	148
Caldicellulosiruptor kronotskyensis 2002[+]	Thermoanaerobacteraceae	2.84				—	148
Moorella thermoacetica ATCC 39073[+]	Thermoanaerobacteraceae	2.63				—	166
Carboxydothermus hydrogenoformans DSM 6008 [+]	Thermoanaerobacteraceae	2.40				—	167
Thermosediminibacter oceani JW/IW-1228P, DSM 16646[+]	Thermoanaerobacterales Family III. *Incertae Sedis*	2.28				—	168
Mahella australiensis 50-1 BON, DSM 15567[+]	Thermoanaerobacterales Family IV. *Incertae Sedis*	3.14				—	169
Natranaerobius thermophilus JW/NM-WN-LF[+]	Natranaerobiaceae	3.19				—	170
Halanaerobium hydrogenoformans	Halanaerobiaceae	2.61				—	171
Halanaerobium praevalens GSL, DSM 2228[+]	Halanaerobiaceae	2.31				—	172
Halothermothrix orenii H 168[+]	Halanaerobiaceae	2.58				—	173
Acetohalobium arabaticum Z-7288, DSM 550[+]	Halobacteroidaceae	2.47				—	174
Acidaminococcus fermentans VR4, DSM 20731[+]	Acidaminococcaceae	2.33				—	175
Acidaminococcus intestini RYC-MR95[+]	Acidaminococcaceae	2.49				—	176
Veillonella parvula Te3, DSM 2008[+]	Veillonellaceae	2.13				—	177

[a] [+]No NRPS or PKS gene clusters detected.

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Table 7 Analysed genomes from the phylum Bacteroidetes

Strain[a]	Family	Genome size (Mb)	PKS	NRPS	Hybrid	size (kb) (%) of genome	Reference
Bacteroides fragilis 638R	*Bacteroidaceae*	5.37		1		2.43 (0.05)	178
Bacteroides thetaiotaomicron VPI-5482+	*Bacteroidaceae*	6.29				—	179
Bacteroides fragilis YCH46	*Bacteroidaceae*	5.28		1		2.43 (0.05)	180
Bacteroides fragilis NCTC 9343	*Bacteroidaceae*	5.21		1		2.43 (0.05)	181
Bacteroides vulgatus ATCC 8482+	*Bacteroidaceae*	5.16				—	182
Bacteroides helcogenes P 36-108+	*Bacteroidaceae*	3.99				—	183
Bacteroides salanitronis BL78	*Bacteroidaceae*	4.24		1		2.48 (0.06)	184
Porphyromonas gingivalis ATCC 33277+	*Porphyromonadaceae*	2.35				—	185
Porphyromonas gingivalis TDC60+	*Porphyromonadaceae*	2.34				—	186
Porphyromonas gingivalis W83+	*Porphyromonadaceae*	2.34				—	187
Parabacteroides distasonis ATCC 8503+	*Porphyromonadaceae*	4.81				—	182
Odoribacter splanchnicus 1651/6+	*Porphyromonadaceae*	4.39				—	188
Paludibacter propionicigenes WB4+	*Porphyromonadaceae*	3.69				—	189
Prevotella ruminicola 23+	*Prevotellaceae*	3.62				—	190
Chlorobium tepidum TLS+	*Chlorobiaceae*	2.15				—	191

[a] +No NRPS or PKS gene clusters detected.

those bacteria containing PKS or NRPS gene clusters will be discussed. A complete list of analysed genomes from the phylum Proteobacteria is presented in Table 8.

7.1 Conserved PUFA biosynthesis gene clusters amongst anaerobic Proteobacteria

Bacteria were once thought to be unable to produce PUFAs, however, since the 1990s it has become apparent that many marine bacteria produce omega-3 fatty acids, such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA).[192,193] More recent studies have shown that PUFA biosynthesis gene clusters appear most often in the δ-proteobacteria, although further genomic analysis of PUFA production suggests that it is more widespread than previously estimated.[110] The genes responsible for PUFA production are designated *pfaA–E*, although some systems have merged the enzymatic activities from multiple genes into one.

Table 8 Analysed genomes from the phylum Proteobacteria

Strain[a]	Class	Order	Genome size (Mb)	PKS	NRPS	Hybrid	size (kb) (%) of genome	Reference
Hippea maritima MH2+	δ-proteobacteria	Desulfurellales	1.69				—	197
Desulfarculus baarsii 2st14+	δ-proteobacteria	Desulfarculales	3.66				—	198
Geobacter uraniireducens R_f 4	δ-proteobacteria	Desulfuromonadales	5.13	2			63.20 (1.23)	199
Geobacter lovleyi SZ	δ-proteobacteria	Desulfuromonadales	3.99	2			21.2 (0.53)	200
Geobacter bemidjiensis Bem	δ-proteobacteria	Desulfuromonadales	4.62	2	1		16.28 (0.35)	201
Geobacter sulfurreducens KN400	δ-proteobacteria	Desulfuromonadales	3.71		1		2.47 (0.07)	202
Geobacter sulfurreducens PCA	δ-proteobacteria	Desulfuromonadales	3.81		1		2.47 (0.07)	203
Geobacter metallireducens GS-15	δ-proteobacteria	Desulfuromonadales	4.00		1		2.47 (0.07)	204
Pelobacter carbinolicus DSM 2380+	δ-proteobacteria	Desulfuromonadales	3.67				—	5
Pelobacter propionicus DSM 2379	δ-proteobacteria	Desulfuromonadales	4.01	1			13.81 (0.34)	204
Syntrophus aciditrophicus SB+	δ-proteobacteria	Syntrophobacterales	3.18				—	205
Desulfobacca acetoxidans ASRB2	δ-proteobacteria	Syntrophobacterales	3.28		1		6.67 (0.20)	206
Desulfomicrobium baculatum X	δ-proteobacteria	Desulfovibrionales	3.94	1			1.09 (0.03)	207
Desulfovibrio vulgaris DP4+	δ-proteobacteria	Desulfovibrionales	3.66				—	208
Desulfovibrio magneticus RS-1	δ-proteobacteria	Desulfovibrionales	5.25		1		3.95 (0.08)	209
Desulfovibrio vulgaris Hildenborough+	δ-proteobacteria	Desulfovibrionales	3.77				—	210
Desulfovibrio alaskensis G20	δ-proteobacteria	Desulfovibrionales	3.70			1	19.44 (0.53)	211
Desulfohalobium retbaense HR100, DSM 5692+	δ-proteobacteria	Desulfovibrionales	2.91				—	212
Desulfatibacillum alkenivorans AK-01	δ-proteobacteria	Desulfobacterales	6.52	1			13.98 (0.21)	213
Desulfobacterium autotrophicum HRM2	δ-proteobacteria	Desulfobacterales	5.59	3			28.85 (0.52)	196
Desulfobulbus propionicus 1pr3	δ-proteobacteria	Desulfobacterales	3.85		1		4.80 (0.12)	214
Desulfotalea psychrophila LSv54+	δ-proteobacteria	Desulfobacterales	3.66				—	215
Nautilia profundicola Am-H+	ε-Proteobacteria	Nautiliales	1.68				—	216
Dichelobacter nodosus VCS1703A+	γ-Proteobacteria	Cardiobacteriales	1.39				—	217

[a] +No NRPS or PKS gene clusters detected.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

The core set of enzymatic activities required for PUFA production in bacteria is shared with PKSs and fatty acid synthases.[193] The PUFA biosynthesis gene clusters described below are often found in conjunction with *ole* genes (for *ole*fin synthesis), which have been shown to be involved in the head-to-head condensation of two fatty acids to give hydrocarbons with chain lengths of C23 to C33.[194] The co-occurrence of *ole* and *pfa* genes has led to the suggestion that, in these instances, the *pfa* genes may provide the precursors necessary for olefin biosynthesis.[110] While our focus is on the PKS type PUFA synthases, we consider here that where both *ole* and *pfa* genes are present in a strain, they may act in concert and have therefore considered them to be part of the same cluster. All of the clusters described here were previously identified,[110] however, the fragmented nature of the *Desulfobacterium autotrophicum* has not been previously mentioned (see below). Here we have performed a more in-depth analysis of a limited set of isolates, and the main features of each of these clusters are summarised below (Fig. 13).

Six of the nine genomes from the genus *Geobacter* and *Pelobacter propionicus* share a type I PKS-like PUFA biosynthesis gene cluster (from *gbem_2106* to *gbem_2098* in *Geobacter bemdjiensis* Bem). Whilst this cluster is relatively well conserved amongst these organisms, there are some subtle differences. For example, *Geobacter lovleyi* is missing the PPTase gene from this cluster and the *ole* genes appear to have undergone rearrangement in *Pelobacter propionicus* (Fig. 13). Indeed, the *G. bemdjiensis* PUFA biosynthesis gene cluster can be linked to the production of hentriacontanonaene (C31:9), a head-to-head olefinic hydrocarbon, whilst *Geobacter sulfurreducens*, which does not contain this gene cluster, does not produce unsaturated lipids.[195]Although both *Desulfatibacillum alkenivorans* AK-01 and *Desulfobacterium autotrophicum* HRM2 contain PUFA biosynthesis gene clusters, they differ from those in the other anaerobes analysed.

Three genes in the *ole* locus (*oleA*, *C*, and *D*) are absent in *Desulfatibacillum alkenivorans*, meaning that it should not be able to produce olefinic PUFAs, however, this has not yet

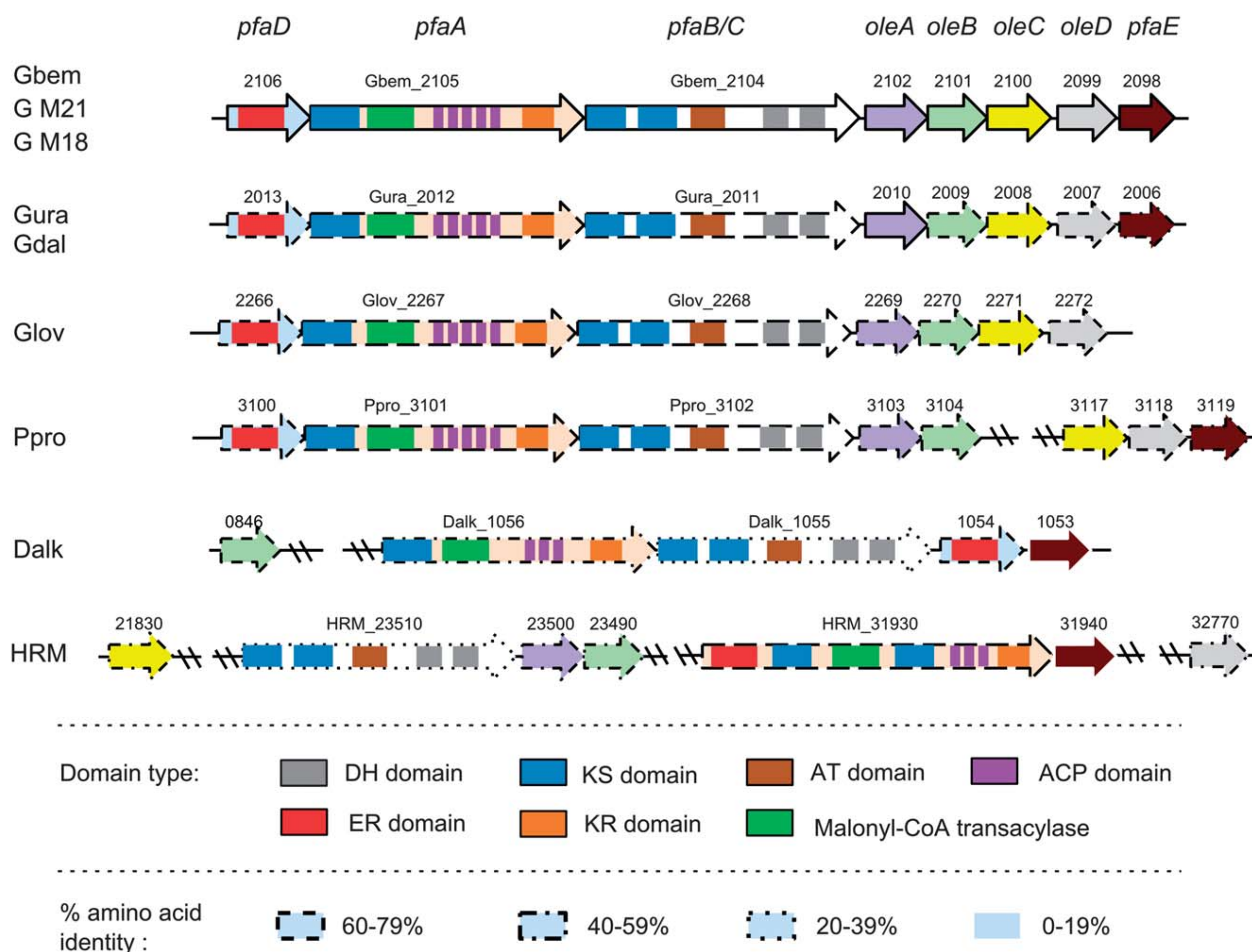


Fig. 13 Conservation of PUFA biosynthesis gene clusters in anaerobic bacteria. Each arrow represents an open reading frame (ORF) and lines surrounding each arrow represent the percentage conservation of encoded proteins, relative to *G. bemidjiensis* orthologues, according to the figure legend. Domain colours are outlined in the figure legend. ORFs of the same colour are predicted to perform the same function. Numbers above each gene represent the locus tag for each gene within the genome sequence of each organism, and the corresponding *pfa* orthologue is marked in the top row. Gbem, *Geobacter bemidjiensis* Bem; G M21, *Geobacter* sp. M21; G M18, *Geobacter* sp. M18; Gura, *Geobacter uraniireducens* R_f 4; Gdal, *Geobacter daltonii* FRC-32; Glov, *Geobacter lovleyi* SZ; Ppro, *Pelobacter propionicus* DSM 2379; Dalk, *Desulfatibacillum alkenivorans* AK-01; HRM, *Desulfobacterium autotrophicum* HRM2.

View Article Online

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

been experimentally verified. *Desulfobacterium autotrophicum* HRM2 appears to contain all of the genes necessary for PUFA biosynthesis, however, these are not located in a contiguous gene cluster, as in the other systems described above. This may be due to the high number of repetitive elements contained within the genome (422 repeat elements per Mb of DNA), which promote genetic rearrangements.[196] Another unusual facet of this system is the putative PPTase, encoded by *hrm2_31940*. Conserved domain analysis shows that this gene encodes a highly unusual tandem PPTase domain structure, which is unique amongst sequences in the Genbank database.

Despite genomic evidence that these gene clusters are involved in PUFA biosynthesis, definitive genetic studies (knock out or heterologous expression) have yet to be performed of any of the strains described above.

7.2 Order Desulfuromonadales

7.2.1 Genus *Geobacter*. The genus *Geobacter* represents a group of organisms that inhabit subsurface environments and can use Fe(III) and Mn(IV) as a terminal electron acceptors to oxidise organic compounds to CO_2.[218] These organisms also possess the ability to reductively precipitate uranium and similar contaminants, leading to great interest in their use for bioremediation.[219] Furthermore, some *Geobacter* species are also capable or producing electricity from organic waste material.[203] The six *Geobacter* genomes that have been completed and published share a number of secondary metabolite biosynthesis gene clusters (Fig. 14). A further three completed, yet unpublished, *Geobacter* genomes have also been analysed for comparison purposes.

Three of the nine *Geobacter* genomes share a type III PKS gene cluster (Fig. 14), which is conserved with complete synteny and

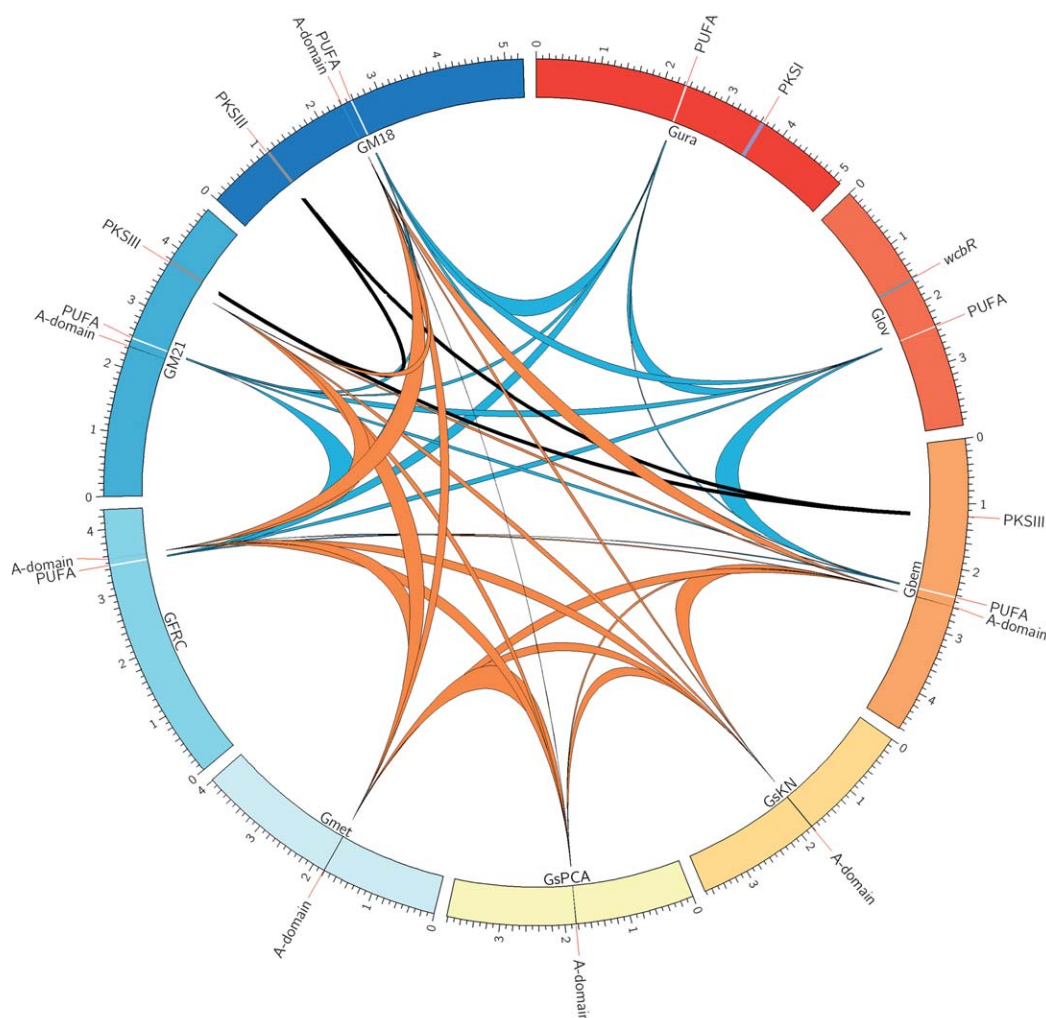


Fig. 14 Conservation of secondary metabolite biosynthesis gene clusters in *Geobacter* spp. Each *Geobacter* genome is represented by a different coloured block in the outer ring (numbers represent genome size in Mb). Secondary metabolite biosynthesis gene clusters mentioned in the text are labelled on the diagram, according to their corresponding position in each genome. Links represent homology between gene clusters, with identical coloured links between conserved clusters. Figure compiled with Circos.[40] Gura, *Geobacter uraniireducens* R_f 4; Glov, *Geobacter lovleyi* SZ; Gbem, *Geobacter bemidjiensis* Bem; GsKN, *Geobacter sulfurreducens* KN400; GsPCA, *Geobacter sulfurreducens* PCA; Gms1, *Geobacter metallireducens* GS-15; GFRC, *Geobacter daltonii* FRC-32; GM21, *Geobacter* sp. M21; GM18, *Geobacter* sp. M18.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

>60% amino acid identity amongst the three isolates. The gene cluster also contains two genes encoding radical SAM containing enzymes, a peptidase and a methyltransferase. The type III PKS (Gbem_1028 in *G. bemidjiensis* Bem) does not show any significant homology to characterised bacterial chalcone and stilbene synthetase-like enzymes, with only 30% identity to Pks11 of *Mycobacterium marinum*. Pks11 in *Mycobacterium tuberculosis* has been shown to be essential for growth and involved in the production of phthiocerol dimycocerosate waxes.[220–222]

The majority of analysed *Geobacter* genomes also contain a gene encoding a single NRPS adenylation domain, with >58% amino acid identity across the species. Conserved domain searches suggest that the encoded protein also possesses ACP and AT functionality and may be involved in glycerophospholipid biosynthesis, however, BLAST searches did not reveal any functionally validated orthologues. This gene appears in two different strain dependent genomic contexts; however, no obvious function can be gleaned from the surrounding genes.

Geobacter lovleyi SZ contains a unique gene (*glov_1541*), encoding a type I PKS with KS-AT-DH-ER-KR-ACP domains. This protein is most homologous to WcbR or RkpA from various bacteria, which have been shown to be involved in capsule production in the pathogens *Sinorhizobium meliloti* and in *Burkholderia pseudomallei* (Fig. 15).[223,224] Despite low levels of amino acid identity (<50%) between WcbR or RpkA homologues, the organisation of gene clusters containing *wcbR* or *rpkA* genes hints at similar biological functions. Fig. 15 shows the organisation of this gene cluster in *G. lovleyi* compared to its closest orthologues. It should be noted, however, that *G. lovleyi* has not been shown to produce a capsule, therefore, the true role of this gene cluster in this strain remains unknown.

G. uraniireducens also contains a unique PKS gene cluster that is not present in any of the other sequenced *Geobacter* genomes. This gene cluster consists of five genes, which together encode a multimodular type I *trans*-AT PKS (*gura_3094* to *gura_3098*). Two *trans*-AT enzymes (Gura_3093 and Gura_3105), a PPTase (Gura_3107) and β-lactamase (Gura_3108) are also encoded by genes within this cluster. Transcriptional studies have shown that the PKS genes from this gene cluster were upregulated between two to nine-fold in cultures grown in subsurface sediments, compared with defined culture medium.[225] However, the cultures were not analysed for secondary metabolite production and as yet, no product has been assigned to this gene cluster.

7.2.2 Genus Pelobacter. Although related to *Geobacter*, *Pelobacter* species do not possess many of the phenotypic properties of members of the genus *Geobacter*. *Pelobacter* spp. are incapable of electron transfer to metal or electrodes and inhabit methanogenic environments, where they can live in syntrophy with methanogens by degrading organic compounds.[5] Two *Pelobacter* genomes have been published, however, only *Pelobacter propionicus* possesses secondary metabolite genes; a single cluster (*ppro_3100* to *ppro_3119*, Fig. 13) orthologous to the PUFA biosynthesis gene locus present in *Geobacter* species (described above).

7.3 Order Syntrophobacterales

The order Syntrophobacterales comprises strictly anaerobic organisms that exist as part of larger syntrophic communities and that play an essential role in the recycling of organic matter to methane and carbon dioxide.[205] As so few of these organisms

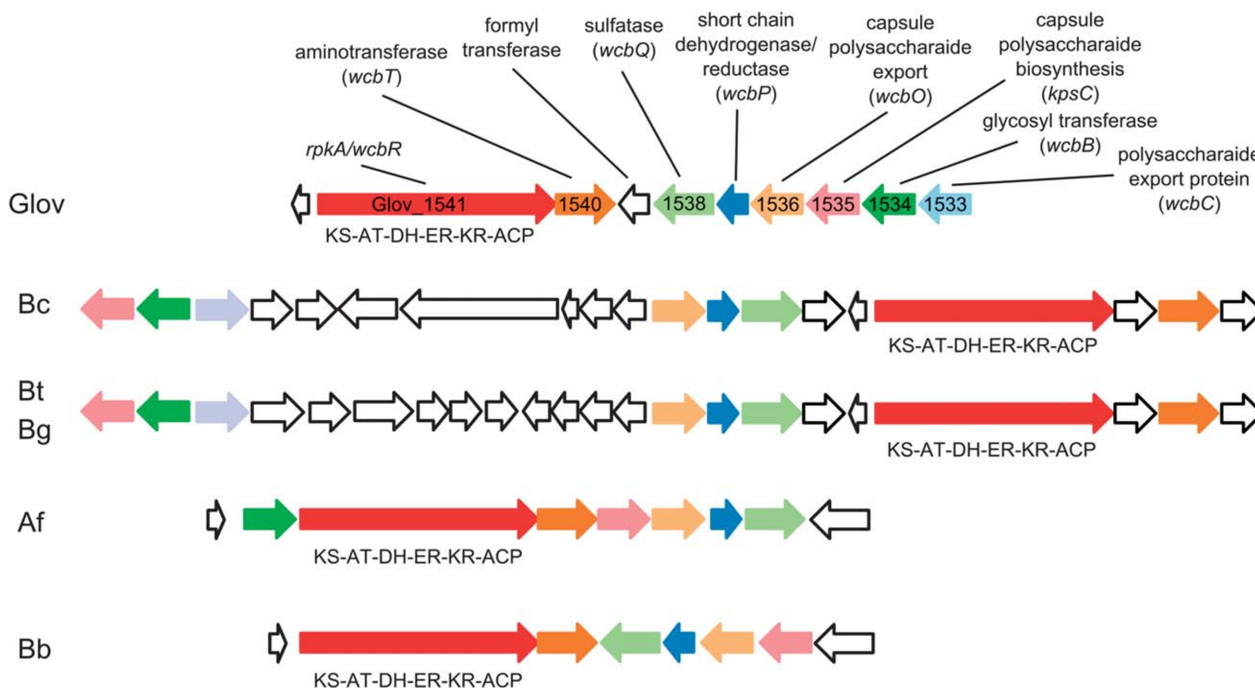


Fig. 15 Gene conservation amongst various species carrying the *wcb/rpk* locus. Arrows of the same colour represent ORFs encoding proteins of the same function. White arrows represent genes unique to each strain. Glov, *Geobacter lovleyi* SZ; Bc, *Burkholderia cenocepacia* MC0-3; Bt, *Burkholderia thailandensis* E264; Bg, *Burkholderia gladioli* BSR3; Af, *Acidithiobacillus ferrooxidans* ATCC 53993; Bb, *Bordetella bronchiseptica* RB50.

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

have been isolated in pure culture, their metabolism is not well understood. To date, two completed genomes from this order exist, namely *Syntrophus aciditrophicus* SB and *Desulfobacca acetoxidans* ASRB2.[142,205]

Only *Desulfobacca acetoxidans*, the type and sole species of the genus *Desulfobacca*, possess a cryptic secondary metabolite biosynthesis gene cluster. This organism was isolated from granular sludge of laboratory scale anaerobic sludge reactor in 1999.[226] This strain contains a NRPS gene (*desac_2459*) encoding the domain structure C-A-T. BLASTP searches revealed the most closely related proteins to be OciC from *Planktothrix rubescens* NIVA-CYA 98 (45% identity) and McnE from *Microcystis aeruginosa* PCC 9701 (41% identity), which are involved in the production of the non-ribosomal peptides cyanopeptolin and microcystin, respectively. However, the genomic context of *desac_2459* does not support a role in the production of either or the above-mentioned peptides and its function remains unknown.

A further gene (*desac_2817*) encoding a didomain A-T was identified. Conserved domain searches suggest, however, that this gene is most likely to be involved in lipid biosynthesis, as a putative fatty acid ligase.

7.4 Order Desulfovibrionales

The Desulfovibrionales represent a group of Gram-negative, sulfate reducing anaerobes. Sequenced genomes from the genus *Desulfovibrio* predominate within this order, although the genera *Desulfomicrobium* and *Desulfohalobium* are represented by one isolate each. The presence of secondary metabolite gene clusters is relatively infrequent within this order and each gene cluster appears to be unique to the strain that possesses it.

Desulfomicrobium baculatum X is the type species of the genus *Desulfomicrobium*, whose 3.9 Mb genome[207] contains a single type III PKS gene cluster of unknown function. Homology searches revealed that type III PKS genes of *Geobacter* spp. are the most closely related, however, amino acid identity is less than 40% amongst homologous proteins.

Desulfovibrio magneticus RS-1 is a magnetotactic bacterium, with the ability to synthesize magnetite crystals.[227] Within the 5.2 Mb chromosome and two circular plasmids[209] is a single chromosomally located NRPS gene (*dmr_01570*), encoding an A-T didomain architecture. The closest orthologue is a predicted peptide synthetase from *Agrobacterium tumefaciens* (Atu3072, 44% amino acid identity), for which no known product has been assigned.

Desulfovibrio alaskensis G20 is capable of reducing both toxic metals and radionuclides, such as uranium and chromium, to less toxic forms.[228] Formerly known as *Desulfovibrio desulfuricans* G20, this organism is also capable of corroding ferrous metals.[229] The 3.7 Mb genome[211] of *Desulfovibrio alaskensis* contains a two gene cassette coding for a hybrid NRPS-PKS. The first gene in the small gene cluster (*dde_1609*) encodes a single PKS module followed by two NRPS modules, whilst the second gene (*dde_1608*) encodes a single PKS module (including a domain with weak similarity to methyltransferases) and a single NRPS module with terminal TE activity. A PPTase encoding gene is also located downstream of *dde_1608*. The NRPS adenylation domains encoded within both of these genes appear to be specific for cysteine incorporation. A gene encoding a protein of similar length and domain structure to Dde_1609 is present in *Legionella longbeacheae* NSW150, one of the pathogens responsible for Legionaire's disease, however, its product has not been identified.

7.5 Order Desulfobacterales

The Desulfobacterales are a further group of sulfate-reducing anaerobes, however, their ecophysiology differs from that of the Desulfovirbrionales and they tend to predominate in habitats where Desulfovirbrionales are mostly absent.[230] Four genomes from the order *Desulfobacterales* have been sequenced (Table 8). Apart from *Desulfatibacillum alkenivorans* AK-01 and *Desulfobacterium autotrophicum* HRM2, which have been discussed above in conjunction with PUFA biosynthesis gene clusters, only *Desulfobulbus propionicus* 1pr3 contains an NRPS-like gene (*despr_1895*). The encoded protein possesses an ACP-A didomain structure, combined with a C-terminal lysophospholipid acyltransferase (LPLAT) domain. Proteins containing the LPLAT domain are usually involved in the biogenesis or remodelling of membrane glycerophospholipids.[231] A further A-domain encoding gene is found downstream of *despr_1895*, although this appears to be a long-chain acyl-CoA synthetase. Taken together, these results suggest that this cluster may code for membrane lipid biosynthesis.

8 Phylum Actinobacteria

Many of the most renowned secondary metabolite producers belong to the phylum Actinobacteria, most notably the Streptomycetes.[2] In stark contrast to their aerobic cousins, anaerobic Actinobacteria, grouped within the Bifidobacteriaceae and Propionibacteriaceae families, appear to have very limited potential for secondary metabolite production. Of the 34 sequenced and published genomes within this phylum, only nine contain NRPS or PKS genes, and seven of these strains all share a single NRPS gene cluster (Table 9).

8.1 Family Bifidobacteriaceae

The family Bifidobacteriaceae contains four genera, however, only *Bifidobacterium* species are obligate anaerobes. Members of the genus *Bifidobacterium* are common inhabitants of the gastrointestinal tract of animals and humans, and many bifidobacterial species are thought to have a positive impact on the intestinal health of their hosts.[232] From a genomic perspective, *Bifidobacterium* species isolated from the human intestine have gained the most attention, due to the potential for use in probiotics and functional foods. Of the 15 analysed *Bifidobacterium* isolates, only one strain (*Bifidobacterium breve* UCC2003) contains PKS genes.

Bifidobacterium breve UCC2003 was isolated from infant faeces and within its 2.4 Mb genome[233] contains two PKS-like genes. These genes (*bbr_0204 - 0205*) encode two proteins with AT-ACP-ACP-KR-DH and KS-KS domain structure, respectively.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Table 9 Analysed genomes from the phylum Actinobacteria

Strain[a]	Family	Genome size (Mb)	PKS	NRPS	Hybrid	Size (kb) (%) of genome	Reference
*Bifidobacterium bifi*dum S17+	Bifidobacteriaceae	2.18				—	237
Bifidobacterium bifidum PRL2010+	Bifidobacteriaceae	2.21				—	238
Bifidobacterium longum infantis 157F-NC+	Bifidobacteriaceae	2.40				—	239
Bifidobacterium longum infantis JCM 1217+	Bifidobacteriaceae	2.39				—	239
Bifidobacterium longum infantis JCM 1222+	Bifidobacteriaceae	2.83				—	239
Bifidobacterium breve UCC2003	Bifidobacteriaceae	2.42	1			7.17 (0.30)	233
Bifidobacterium animalis lactis CNCM I-2494+	Bifidobacteriaceae	1.94				—	240
Bifidobacterium longum NCC2705+	Bifidobacteriaceae	2.26				—	241
Bifidobacterium longum DJO10A+	Bifidobacteriaceae	2.38				—	242
Bifidobacterium longum infantis ATCC 15697+	Bifidobacteriaceae	2.83				—	243
Bifidobacterium longum subsp. longum JDM30+	Bifidobacteriaceae	2.48				—	244
Bifidobacterium longum subsp. longum BBMN68+	Bifidobacteriaceae	2.27				—	245
Bifidobacterium longum longum KACC 91563+	Bifidobacteriaceae	2.40				—	246
Bifidobacterium animalis lactis AD011+	Bifidobacteriaceae	1.93				—	247
Bifidobacterium animalis lactis Bl-04+	Bifidobacteriaceae	1.94				—	248
Bifidobacterium animalis lactis DSM 10140+	Bifidobacteriaceae	1.94				—	248
Bifidobacterium dentium Bd+	Bifidobacteriaceae	2.64				—	249
Bifidobacterium animalis lactis BB-12+	Bifidobacteriaceae	1.94				—	250
Bifidobacterium animalis lactis V9+	Bifidobacteriaceae	1.94				—	251
Bifidobacterium animalis lactis BLC1+	Bifidobacteriaceae	1.94				—	252
Bifidobacterium animalis subsp. animalis ATCC 25527+	Bifidobacteriaceae	1.93				—	253
Propionibacterium acnes KPA171202	Propionibacteriaceae	2.56		1		5.06 (0.20)	254
Propionibacterium freudenreichii shermanii CIRM-BIA1+	Propionibacteriaceae	2.62				—	255
Propionibacterium acnes TypeIA2 P.acn17	Propionibacteriaceae	2.52		1		2.85 (0.11)	256
Propionibacterium acnes TypeIA2 P.acn31	Propionibacteriaceae	2.50		1		2.78 (0.11)	256
Propionibacterium acnes TypeIA2 P.acn33	Propionibacteriaceae	2.49		1		2.78 (0.11)	256
Propionibacterium acnes 266	Propionibacteriaceae	2.49		1		5.06 (0.20)	235
Propionibacterium acnes 6609	Propionibacteriaceae	2.56		1		2.93 (0.11)	257
Propionibacterium acnes ATCC 11828	Propionibacteriaceae	2.49		1		2.94 (0.12)	258
Slackia heliotrinireducens RHS 1	Coriobacteriaceae	3.17		1		2.81 (0.09)	259
Eggerthella lenta VPI 0255+	Coriobacteriaceae	3.63				—	260
Cryptobacterium curtum 12-3T+	Coriobacteriaceae	1.62				—	261
Atopobium parvulum IPP 1246+	Coriobacteriaceae	1.54				—	262
Olsenella uli VPI, DSM 7084+	Coriobacteriaceae	2.05				—	263

[a] +No NRPS or PKS gene clusters detected.

Furthermore, a PPTase gene (*bbr_0209*) is also found as part of this cluster. Investigation of draft *Bifidobacterium* genome sequences shows that this cluster also exists in partial form in *Bifidobacterium breve* ACS-071-V-Sch8b, although it appears to have been deleted from *Bifidobacterium breve* DSM 20213. The multiple KS and ACP domains encoded by these genes are reminiscent of PUFA synthesis systems described above, however, both proteins show only limited identity (<35%) to those potentially involved in PUFA biosynthesis from *Desulfatibacillum alkenivorans* and *Desulfobacterium autotrophicum*. Despite their partial similarity to PUFA synthases, no EHA or DHA have been detected from *Bifidobacterium* species, although several *Bifidobacterium breve* isolates have been shown to produce conjugated linoleic acids (CLAs), another class of PUFAs.[234] CLAs are likely produced by an alternate enzymatic mechanism not related to PKSs, however, a genetic basis for the production of CLAs in *Bifidobacterium* species has not been verified experimentally. With no currently identified product linked to this cluster, the question remains open as to whether these clusters code for the biosynthesis of PUFAs.

8.2 Family Propionibacteriaceae

The family Propionibacteriaceae comprises non-spore forming, pleomorphic anaerobic rods, the majority of which are normally found as harmless commensals on human skin. However,

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Propionibacterium. acnes, in particular, is also an opportunistic pathogen. Currently there are eight completed and published genomes from the Propionibacteriaceae family, whilst many others are in the draft phase. Genomic sequencing has focused on human pathogenic *P. acnes* isolates (seven of eight genomes), whilst one *Propionibacterium freudenreichii* strain has also been sequenced.

All of the *P. acnes* isolates sequenced to date contain a two gene NRPS gene cluster, however, at least one of these genes appears truncated in a number of strains. This genomic region has previously been identified in two comparative genomic studies.[3,235] Two genes, *ppa1288* and *ppa1287* (*P. acnes* KPA171202 locus tags), encode A-T-C and A-T domain architectures, respectively, however *ppa1288* orthologues are truncated in a number of *P. acnes* strains. Genes upstream and downstream of the two NRPS genes are still intact and syntenic amongst *P. acnes* strains, indicating that just the NRPS genes are affected. The presence of a full-length gene cluster does not appear to be related to virulence as some strains associated with moderate to severe acne, based on a recent multi-locus sequence typing scheme,[236] also possess a partially truncated gene cluster.

Due to shared homology with NRPS involved in the production of cyclic lipopeptides, such as surfactin, it was suggested that the NRPS gene cluster encodes the biosynthesis of a similar molecule.[235] However, the limited number of modules would prevent a large surfactin-like molecule from being produced and our analysis shows only limited amino acid identity (<30%) to characterised proteins involved in cyclic lipopeptide biosynthesis, such as tyrocidine or gramicidin.

Transcriptional analysis of two strains that contain the full length genes showed that this region was expressed in both strain KPA171202 and strain 266 under standard culture conditions, but at a higher level in strain KPA171202.[235]

8.3 Family Coriobacteriaceae

The family Coriobacteriaceae comprises high-GC, Gram-positive anaerobes from a number of different genera. To date, five genomes have been completed and published, however, only *Slackia heliotrinireducens* RHS 1 contains a single NRPS-like gene. *Slackia heliotrinireducens* was isolated from the rumen of an Australian sheep and has the ability to degrade pyrrolizidine alkaloids. The NRPS gene (*shel_23800*) encodes a N-terminal short chain dehydrogenase/reductase domain and a C-terminal A-domain of unknown function.

9 Phylum Thermotogae

The phylum Thermotogae predominantly comprises organisms that thrive in temperatures above 70 °C, although some mesophilic *Thermotoga* sp. have recently been identified.[264,265] There are currently seven completed and published genomes, however, none of these possess PKS or NRPS genes (Table 10).

Given that comparative genomic studies have revealed that the *Thermotogae* have received many of their genes *via* horizontal gene transfer (HGT) from Firmicutes or Archea, it is interesting that no clostridial secondary metabolite biosynthesis gene clusters have been passed on to *Thermotoga* strains. This may be because many HGT events in *Thermotoga* spp. have arisen from thermophilic Firmicutes, such as *Thermoanaerobacter*,[265] which are commonly devoid of secondary metabolite biosynthesis gene clusters. Interestingly, Thermotogae often inhabit the same high-temperature and pressure environments as *Thermoanaerobacter* and *Desulfotomaculum* spp.,[266,267] both of which also have limited ability to produce secondary metabolites. It is interesting to speculate that the environment in which these organisms live either does not require secondary metabolites, due to it being a highly restrictive niche, or is not permissive for their production.

Table 10 Analysed genomes from the phylum Thermotogae

Strain[a]	Genome size (Mb)	Reference
Thermotoga petrophila RKU-1[+]	1.82	265
Thermotoga lettingae TMO[+]	2.13	265
Thermotoga sp. RQ2[+]	1.88	268
Thermotoga neapolitana DSM 4359[+]	1.88	269
Thermotoga naphthophila RKU-10[+]	1.81	268
Thermotoga maritima MSB8[+]	1.86	270
Thermosipho melanesiensis BI429[+]	1.92	265
Thermosipho africanus TCF52B[+]	2.02	271
Fervidobacterium nodosum Rt17-B1[+]	1.95	265

[a] [+]No NRPS or PKS gene clusters detected.

10 Phylum Spirochaetes

The phylum *Spirochaetes* consists of 21 completed and published genomes from anaerobic strains (Table 11). The sequenced genomes are biased towards pathogenic isolates, mainly from the genera *Treponema* (consisting mainly of syphilis-causing *Treponema pallidum* isolates) and *Brachyspira* (pathogens of swine and fowl). Regardless of the site of isolation, genus or species, all analysed genomes of the Spirochaetes were devoid of PKS and NRPS genes. These observations are in agreement with those previously found for two Spirochaete genomes[3] and this larger data set suggests that all members of the phylum Spirocheaetes are incapable of producing polyketide or non-ribosomal peptide secondary metabolites.

In a similar situation to that of the Thermotogae, comparative genomic studies have shown that HGT has significantly shaped the genome of *Sphaerocheata* spp., an unusual group of organisms, which align with the Spirochetes by 16s rDNA sequencing. However, they do not have axial flagella nor a spiral shape, the classical morphological features of members of this phylum.[272] *Sphaerocheata* spp. share many genes involved in carbohydrate metabolism with the Clostridia and it is interesting that these strains have not acquired any secondary metabolite biosynthesis genes, given their prevalence in Clostridia with potentially overlapping habitats.

 This journal is © The Royal Society of Chemistry 2013

Table 11 Analysed genomes from the phylum Spirochaetes

Strain[a]	Family	Genome size (Mb)	Reference
Treponema pallidum pallidum SS14+	Spirochaetaceae	1.14	273
Treponema pallidum pallidum Chicago+	Spirochaetaceae	1.14	274
Treponema succinifaciens 6091+	Spirochaetaceae	2.90	275
Treponema azotonutricium ZAS-9+	Spirochaetaceae	3.86	276
Treponema primitia ZAS-2+	Spirochaetaceae	4.06	276
Treponema paraluiscuniculi Cuniculi A+	Spirochaetaceae	1.13	277
Treponema pallidum pallidum Dallas1+	Spirochaetaceae	1.14	278
Treponema pallidum pertenue CDC2+	Spirochaetaceae	1.14	278
Treponema pallidum pertenue Gauthier+	Spirochaetaceae	1.14	278
Treponema pallidum pertenue SamoaD+	Spirochaetaceae	1.14	278
Treponema pallidum pallidum Nichol+	Spirochaetaceae	1.14	279
Treponema denticola ATCC 35405+	Spirochaetaceae	2.84	280
Spirochaeta smaragdinae SEBR 4228+	Spirochaetaceae	4.65	281
Spirochaeta thermophila DSM 6192+	Spirochaetaceae	2.47	282
Sphaerochaeta globosa Buddy+	Spirochaetaceae	3.31	272
Sphaerochaeta pleomorpha Grapes+	Spirochaetaceae	3.59	272
Sphaerochaeta coccoides SPN1+	Spirochaetaceae	2.22	283
Brachyspira hyodysenteriae WA1+	Brachyspiraceae	3.04	284
Brachyspira pilosicoli 95/1000+	Brachyspiraceae	2.59	285
Brachyspira intermedia PWS/A+	Brachyspiraceae	3.30	286
Brachyspira murdochii 56-150+	Brachyspiraceae	3.24	287

[a] +No NRPS or PKS gene clusters detected.

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

11 Other phyla

The remaining strains, comprising 26 sequenced genomes (Table 12), represent nine diverse phyla, as well as two unclassified bacteria. Anaerobe genomes from the phyla Chloroflexi (six strains), Deferribacteres (four strains), Chrysiogenetes (one strain), Thermodesulfobacteria (one strain), Synergistetes (three strains) and Deinococcus-Thermus (one strain) and from *Akkermansia muciniphila* ATCC BAA-835 do not contain secondary metabolite biosynthesis gene clusters. The remaining ten genomes will be discussed in the following sections.

11.1 Phylum Verrucomicrobia

The Verrucomicrobia have been found in aquatic and terrestrial environments, where they are estimated to comprise up to 10% of the total bacteria in soil.[288] However, difficulty isolating these organisms has meant that there is little experimental information available, and consequently, precious little is known about their metabolism or ecology.[289] This phylum contains two sequenced members (Table 12), of which only *Opitutus terrae* contains secondary metabolite biosynthesis genes.

Opitutus terrae PB90-1 was isolated from rice paddy soil and within its 5.96 Mb genome[290] contains a large hybrid NRPS-PKS gene cluster (*oter_1961* to *oter_1973*). This gene cluster contains 11 genes coding for 11 NRPS and 6 *cis*-AT PKS modules, comprising almost 70 kb of DNA. Homology searches have shown that this gene cluster is unique to strain PB90-1 and is not present in any of the other draft genomes from the family Opitutaceae, and only weak homologues exist in other bacteria, none of which live an anaerobic lifestyle. Several genes surrounding the hybrid gene cluster are, however, present in the other Opitutaceae genomes, suggesting that this cluster may have once been present but has been lost through deletion events, or that this cluster has been acquired by strain PB90-1 from an unknown source.

A putative type III PKS gene (*oter_3635*) with unknown function was also detected in the *Opitutus terrae* genome. Oter_3635 has limited (<30%) amino acid identity to many uncharacterised type III PKS, however, homology is higher (>60%) to orthologues within the Opitutaceae family.

11.2 Phylum Elusimicrobia

The phylum Elusimicrobia was previously known as "termite group 1" and consists mainly of uncultivated bacteria living within the intestines of insects and in soil. In 2009 the first representative of this phylum was isolated in pure culture and was given the epithet *Elusimicrobium minutum*.[291] The genome of the organism was duly sequenced and was shown to be 1.69 Mb.[292]

Within the relatively small genome of *Elusimicrobium minutum* are two secondary metabolite biosynthesis gene clusters. The first of these is a two-gene NRPS gene cluster, where the NRPS genes are temporally separated by approximately 15 kb of DNA. The two NRPS genes, *emin_0995* and *emin_1012*, encode proteins with A-C-T-TE and single A domain, respectively. Emin_0995 has ~30% identity to cyclic peptide related NRPS proteins from algae, including NosD (*Nostoc* spp., nostopeptolide), OciC (*Planktothrix* spp., cyanopeptolin) and McnE (*Microcystis* spp., cyanopeptolin/micropeptin).[293–296]

Upstream and in a putative operon with *emin_0995* is a gene encoding a phosphoribosylglycinamide synthetase, which catalyses the second step in purine biosynthesis.[297,298] Emin_1012

View Article Online

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

Table 12 Analysed genomes from other phyla

Strain[a]	Phylum	Order	Genome size (Mb)	PKS	NRPS	Hybrid	Size (kb) (%) of genome	Reference
Dehalococcoides ethenogenes 195[+]	Chloroflexi	—	1.47				—	299
Dehalococcoides sp. CBDB1[+]	Chloroflexi	—	1.40				—	300
Dehalococcoides sp. BAV1[+]	Chloroflexi	—	1.34				—	301
Dehalococcoides sp. VS[+]	Chloroflexi	—	1.41				—	301
Dehalogenimonas lykanthroporepellens BL-DC-9[+]	Chloroflexi	—	1.67				—	302
Chloroflexus aurantiacus J-10-fl[+]	Chloroflexi	Chloroflexales	5.26				—	303
Deferribacter desulfuricans SSM1[+]	Deferribacteres	Deferribacterales	2.54				—	304
Calditerrivibrio nitroreducens Yu37-1[+]	Deferribacteres	Deferribacterales	2.21				—	305
Denitrovibrio acetiphilus N2460[+]	Deferribacteres	Deferribacterales	3.22				—	306
Flexistipes sinusarabici MAS10[+]	Deferribacteres	Deferribacterales	2.53				—	307
Desulfurispirillum indicum S5[+]	Chrysiogenetes	Chrysiogenales	2.93				—	308
Thermodesulfatator indicus CIR 29812[+]	Thermodesulfobacteria	Thermodesulfobacteriales	2.32				—	309
Thermanaerovibrio acidaminovorans Su883[+]	Synergistetes	Synergistales	1.84				—	310
Aminobacterium colombiense ALA-1[+]	Synergistetes	Synergistales	1.98				—	311
Thermovirga lienii Cas60314[+]	Synergistetes	Synergistales	2.00				—	312
Opitutus terrae PB90-1	Verrucomicrobia	Opitutales	5.96			1	68.88 (1.16)	290
Akkermansia muciniphila ATCC BAA-835[+]	Verrucomicrobia	Verrucomicrobiales	2.66				—	313
Elusimicrobium minutum Pei 191	Elusimicrobia	Elusimicrobiales	1.64	1	1		16.28 (0.99)	292
Leptotrichia buccalis C-1013-b, DSM 1135	Fusobacteria	Fusobacteriales	2.47		1		2.48 (0.10)	314
Sebaldella termitidis ATCC 33386	Fusobacteria	Fusobacteriales	4.49		1		2.46 (0.05)	315
Ilyobacter polytropus CuHBu1, DSM 2926	Fusobacteria	Fusobacteriales	3.12		1		4.94 (0.16)	316
Fusobacterium nucleatum ATCC 25586	Fusobacteria	Fusobacteriales	2.17		1		2.50 (0.12)	317
Fibrobacter succinogenes S85	Fibrobacteres	Fibrobacterales	3.84		1		2.97 (0.08)	318
Thermus scotoductus SA-01[+]	Deinococcus-Thermus	Thermales	2.36				—	319
Candidatus Cloacamonas acidaminovorans*[+]	—	—	2.25				—	320
Candidatus Methylomirabilis oxyfera*	—	—	2.75	1			1.03 (0.04)	321

[a] [+]No NRPS or PKS gene clusters detected. *These bacteria have not been assigned to a phylum as yet.

consists of an N-terminal LPLAT domain (see above for *Desulfobulbus propionicus*), combined with a C-terminal A domain. The combination of these two domains suggests a role in membrane biogenesis and remodelling, as may be the case with a similar enzyme from *Desulfobulbus propionicus* (see above).

The second gene cluster contains a single PKS gene (*emin_1101*) encoding the domain structure KS-AT-KS-KR. The encoded protein does not contain an ACP domain or TE activity, and there are no obvious genes encoding ACP or TE-like functionalities within the vicinity of this cluster. There is, however, a putative PPTase gene located directly downstream of *emin_1101*. The genomic context of this gene gives limited information about its function and its product remains unknown.

11.3 Phylum Fusobacteria

Fusobacteria are rod-shaped, Gram-negative, non-sporulating anaerobes, similar to Bacteroidetes. Several *Fusobacterium* species are considered pathogenic, and can cause necrotic tissue infections and septicemia, although infections with these bacteria are relatively rare.[322] Other members of the Fusobacteria, such as the *Leptotrichia*, are commensals of the human oral cavity.

Four genomes from the families Leptotrichiaceae and Fusobacteriaceae have been completed (Table 12), and all contain a single NRPS gene cluster. Each gene cluster within each respective organism contains only a single NRPS gene, with the encoded domain structure A-ACP-LPLAT, except for *Ilyobacter polytropus*, which contains two of these genes (*ilyop_2170* and *ilyop_2171*) located consecutively on a plasmid.

Although their deduced products possess the same structure, each gene has a differing level of homology to each other; even Ilyop_2170 and Ilyop_2171 share only 50% amino acid identity. The genomic context of each gene is different within each species and, although the LPLAT domain suggests a role in membrane biogenesis, the function of each of these genes remains unclear.

11.4 Phylum Fibrobacteres

The phylum Fibrobacteres contains a number of organisms known to play a role in the degradation of cellulose in the rumen. The phylum contains only one genus, with only two

This journal is © The Royal Society of Chemistry 2013

View Article Online

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

recognised species. As such, only one genome has been completed and published from this phylum.

Fibrobacter succinogenes was first isolated in 1947 from bovine rumen and it has the unusual ability to degrade cellulose by a cellulosome independent method.[323] The 3.84 Mb genome[318] contains only a single NRPS gene (*fisuc_2434*) encoding the domain structure A-T-C. There are no genes encoding PPTases or chain release enzymes in the immediate vicinity, making this a stand-alone NRPS module of unknown function.

11.5 Unclassified bacteria

'*Candidatus* Methylomirabilis oxyfera' represents an anaerobic organism with a novel method of producing oxygen. It is capable of oxidizing methane with oxygen generated from the reduction of nitrite to dinitrogen gas without a nitrous oxide reductase.[321] The genome of the organism was reconstructed from metagenomic sequencing of enriched sediment cultures and it has yet to be cultivated. '*Candidatus* Methylomirabilis oxyfera' contains one type III PKS gene. BLASTP analysis of the encoded protein showed only weak similarity (<35%) to other PKS III proteins, including those that have been characterised.

12 Conclusions

The search for new natural products is a continual process and due to the high rediscovery rate in previously investigated sources, the investigation of novel genera provides a potential avenue to new products. Here we have applied bioinformatic techniques to identify and catalogue secondary metabolite biosynthesis gene clusters in anaerobic bacteria. This is the first study to use genomics to investigate the biosynthetic capability of anaerobes and shows that, contrary to popular belief, many anaerobes are capable of producing natural products. Compared to prolific natural product producers, such as *Streptomyces* spp., at first glance it may appear that anaerobes have rather limited ability to produce natural products. However, when one considers the limited energy that an anaerobic lifestyle provides, the diversion of precious nutrients into secondary metabolic pathways suggests that these compounds must play an important role in the lifestyle of these organisms.

At present, both Firmicutes of the genus *Clostridium* and anaerobic δ-proteobacteria appear to be the most promising sources of novel natural products from the anaerobic world. However, there are limited numbers of genomes sequenced from several phyla, which show some promise, including the Verrucomicrobia and Elusimicrobia. Only time (and more genomes) will tell if secondary metabolite genes are common features in both of these phyla.

What does this mean for the future of secondary metabolite research in anaerobes? Certainly, the future looks brighter now than it ever has in the past. We have shown that there are many potential new molecules waiting to be identified from the "anaerob-ome" and we hope that this investigation will increase the attention given to secondary metabolite research in anaerobes.

13 References

1 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2012, **75**, 311–335.
2 M. Nett, H. Ikeda and B. S. Moore, *Nat. Prod. Rep.*, 2009, **26**, 1362–1384.
3 S. Donadio, P. Monciardini and M. Sosio, *Nat. Prod. Rep.*, 2007, **24**, 1073–1109.
4 J. Zhou, Q. He, C. L. Hemme, A. Mukhopadhyay, K. Hillesland, A. Zhou, Z. He, J. D. Van Nostrand, T. C. Hazen, D. A. Stahl, J. D. Wall and A. P. Arkin, *Nat. Rev. Microbiol.*, 2011, **9**, 452–466.
5 J. E. Butler, N. D. Young and D. R. Lovley, *BMC Genomics*, 2009, **10**, 103.
6 P. Durre, *Ann. N. Y. Acad. Sci.*, 2008, **1125**, 353–362.
7 D. L. Stevens, M. J. Aldape and A. E. Bryant, *Anaerobe*, 2012, **18**, 254–259.
8 H. Seedorf, W. F. Fricke, B. Veith, H. Bruggemann, H. Liesegang, A. Strittmatter, M. Miethke, W. Buckel, J. Hinderberger, F. Li, C. Hagemeier, R. K. Thauer and G. Gottschalk, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 2128–2133.
9 T. Lincke, S. Behnken, K. Ishida, M. Roth and C. Hertweck, *Angew. Chem., Int. Ed.*, 2010, **49**, 2011–2013.
10 J. T. Heap, S. A. Kuehne, M. Ehsaan, S. T. Cartman, C. M. Cooksley, J. C. Scott and N. P. Minton, *J. Microbiol. Methods*, 2010, **80**, 49–55.
11 M. V. Coppi, C. Leang, S. J. Sandler and D. R. Lovley, *Appl. Environ. Microbiol.*, 2001, **67**, 3180–3187.
12 J. B. Rollefson, C. E. Levar and D. R. Bond, *J. Bacteriol.*, 2009, **191**, 4207–4217.
13 M. Ichimura, H. Nakayama-Imaohji, S. Wakimoto, H. Morita, T. Hayashi and T. Kuwahara, *Appl. Environ. Microbiol.*, 2010, **76**, 3325–3332.
14 D. W. Udwary, E. A. Gontang, A. C. Jones, C. S. Jones, A. W. Schultz, J. M. Winter, J. Y. Yang, N. Beauchemin, T. L. Capson, B. R. Clark, E. Esquenazi, A. S. Eustaquio, K. Freel, L. Gerwick, W. H. Gerwick, D. Gonzalez, W. T. Liu, K. L. Malloy, K. N. Maloney, M. Nett, J. K. Nunnery, K. Penn, A. Prieto-Davo, T. L. Simmons, S. Weitz, M. C. Wilson, L. S. Tisa, P. C. Dorrestein and B. S. Moore, *Appl. Environ. Microbiol.*, 2011, **77**, 3617–3625.
15 E. A. Emmert, A. K. Klimowicz, M. G. Thomas and J. Handelsman, *Appl. Environ. Microbiol.*, 2004, **70**, 104–113.
16 C. Hertweck, *Angew. Chem., Int. Ed.*, 2009, **48**, 4688–4716.
17 K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925–936.
18 M. B. Austin and J. P. Noel, *Nat. Prod. Rep.*, 2003, **20**, 79–110.
19 J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.
20 J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.
21 J. Grunewald and M. A. Marahiel, *Microbiol. Mol. Biol. Rev.*, 2006, **70**, 121–146.
22 A. Koglin and C. T. Walsh, *Nat. Prod. Rep.*, 2009, **26**, 987–1000.
23 M. Strieker, A. Tanovic and M. A. Marahiel, *Curr. Opin. Struct. Biol.*, 2010, **20**, 234–240.

This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

24 L. Du, C. Sanchez and B. Shen, *Metab. Eng.*, 2001, **3**, 78–95.
25 F. Lipmann, *Science*, 1971, **173**, 875–884.
26 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493–505.
27 G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211–224.
28 G. Yadav, R. S. Gokhale and D. Mohanty, *J. Mol. Biol.*, 2003, **328**, 335–363.
29 H. Tae, E. B. Kong and K. Park, *BMC Bioinformatics*, 2007, **8**, 327.
30 Y. Minowa, M. Araki and M. Kanehisa, *J. Mol. Biol.*, 2007, **368**, 1500–1517.
31 N. Ziemert, S. Podell, K. Penn, J. H. Badger, E. Allen and P. R. Jensen, *PLoS One*, 2012, **7**, e34064.
32 S. Anand, M. V. Prasad, G. Yadav, N. Kumar, J. Shehara, M. Z. Ansari and D. Mohanty, *Nucleic Acids Res.*, 2010, **38**, W487–496.
33 M. H. Li, P. M. Ung, J. Zajkowski, S. Garneau-Tsodikova and D. H. Sherman, *BMC Bioinformatics*, 2009, **10**, 185.
34 T. Weber, C. Rausch, P. Lopez, I. Hoof, V. Gaykova, D. H. Huson and W. Wohlleben, *J. Biotechnol.*, 2009, **140**, 13–17.
35 C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben and D. H. Huson, *Nucleic Acids Res.*, 2005, **33**, 5799–5808.
36 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225–233.
37 A. Starcevic, J. Zucko, J. Simunkovic, P. F. Long, J. Cullum and D. Hranueli, *Nucleic Acids Res.*, 2008, **36**, 6882–6892.
38 P. Kamra, R. S. Gokhale and D. Mohanty, *Nucleic Acids Res.*, 2005, **33**, W220–225.
39 D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chem. Biol.*, 2008, **15**, 1231–1240.
40 M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones and M. A. Marra, *Genome Res.*, 2009, **19**, 1639–1645.
41 M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano and R. Breitling, *Nucleic Acids Res.*, 2011, **39**, W339–346.
42 S. J. Moss, C. J. Martin and B. Wilkinson, *Nat. Prod. Rep.*, 2004, **21**, 575–593.
43 D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik and R. Muller, *J. Am. Chem. Soc.*, 2008, **130**, 14234–14243.
44 N. Traitcheva, H. Jenke-Kodama, J. He, E. Dittmann and C. Hertweck, *ChemBioChem*, 2007, **8**, 1841–1849.
45 C. Olano, B. Wilkinson, S. J. Moss, A. F. Brana, C. Mendez, P. F. Leadlay and J. A. Salas, *Chem. Commun.*, 2003, 2780–2782.
46 I. Pagani, K. Liolios, J. Jansson, I. M. Chen, T. Smirnova, B. Nosrat, V. M. Markowitz and N. C. Kyrpides, *Nucleic Acids Res.*, 2012, **40**, D571–579.
47 H. P. Klenk and M. Goker, *Syst. Appl. Microbiol.*, 2010, **33**, 175–182.
48 D. Wu, P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N. N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B. J. Tindall, S. D. Hooper, A. Pati, A. Lykidis, S. Spring, I. J. Anderson, P. D'Haeseleer, A. Zemla, M. Singer, A. Lapidus, M. Nolan, A. Copeland, C. Han, F. Chen, J. F. Cheng, S. Lucas, C. Kerfeld, E. Lang, S. Gronow, P. Chain, D. Bruce, E. M. Rubin, N. C. Kyrpides, H. P. Klenk and J. A. Eisen, *Nature*, 2009, **462**, 1056–1060.
49 S. G. Van Lanen and B. Shen, *Curr. Opin. Microbiol.*, 2006, **9**, 252–260.
50 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.
51 H. Gross, *Curr. Opin. Drug. Discov. Devel.*, 2009, **12**, 207–219.
52 J. M. Winter, S. Behnken and C. Hertweck, *Curr. Opin. Chem. Biol.*, 2011, **15**, 22–31.
53 M. Zerikly and G. L. Challis, *ChemBioChem*, 2009, **10**, 625–633.
54 R. Teta, M. Gurgui, E. J. Helfrich, S. Kunne, A. Schneider, G. Van Echten-Deckert, A. Mangoni and J. Piel, *ChemBioChem*, 2010, **11**, 2506–2512.
55 S. Bergmann, J. Schumann, K. Scherlach, C. Lange, A. A. Brakhage and C. Hertweck, *Nat. Chem. Biol.*, 2007, **3**, 213–217.
56 M. Gressler, C. Zaehle, K. Scherlach, C. Hertweck and M. Brock, *Chem. Biol.*, 2011, **18**, 198–209.
57 Y. M. Chiang, E. Szewczyk, A. D. Davidson, N. Keller, B. R. Oakley and C. C. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 2965–2970.
58 D. Pistorius and R. Muller, *ChemBioChem*, 2012, **13**, 416–426.
59 L. Laureti, L. Song, S. Huang, C. Corre, P. Leblond, G. L. Challis and B. Aigle, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6258–6263.
60 B. Frank, S. C. Wenzel, H. B. Bode, M. Scharfe, H. Blocker and R. Muller, *J. Mol. Biol.*, 2007, **374**, 24–38.
61 J. E. Loper, M. D. Henkels, B. T. Shaffer, F. A. Valeriote and H. Gross, *Appl. Environ. Microbiol.*, 2008, **74**, 3085–3093.
62 I. de Bruijn, M. J. de Kock, M. Yang, P. de Waard, T. A. van Beek and J. M. Raaijmakers, *Mol. Microbiol.*, 2007, **63**, 417–428.
63 N. Brendel, L. P. Partida-Martinez, K. Scherlach and C. Hertweck, *Org. Biomol. Chem.*, 2007, **5**, 2211–2213.
64 K. Ishida, T. Lincke, S. Behnken and C. Hertweck, *J. Am. Chem. Soc.*, 2010, **132**, 13966–13968.
65 J. B. Biggins, M. A. Ternei and S. F. Brady, *J. Am. Chem. Soc.*, 2012, **134**, 13192–13195.
66 J. Franke, K. Ishida and C. Hertweck, *Angew. Chem., Int. Ed.*, 2012, **51**, 11611–11615.
67 E. Mahenthiralingam, L. Song, A. Sass, J. White, C. Wilmot, A. Marchbank, O. Boaisha, J. Paine, D. Knight and G. L. Challis, *Chem. Biol.*, 2011, **18**, 665–677.
68 M. He, M. Sebaihia, T. D. Lawley, R. A. Stabler, L. F. Dawson, M. J. Martin, K. E. Holt, H. M. Seth-Smith, M. A. Quail, R. Rance, K. Brooks, C. Churcher, D. Harris, S. D. Bentley, C. Burrows, L. Clark, C. Corton, V. Murray, G. Rose, S. Thurston, A. van Tonder, D. Walker, B. W. Wren,

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

G. Dougan and J. Parkhill, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 7527–7532.

69 G. S. Myers, D. A. Rasko, J. K. Cheung, J. Ravel, R. Seshadri, R. T. DeBoy, Q. Ren, J. Varga, M. M. Awad, L. M. Brinkac, S. C. Daugherty, D. H. Haft, R. J. Dodson, R. Madupu, W. C. Nelson, M. J. Rosovitz, S. A. Sullivan, H. Khouri, G. I. Dimitrov, K. L. Watkins, S. Mulligan, J. Benton, D. Radune, D. J. Fisher, H. S. Atkins, T. Hiscox, B. H. Jost, S. J. Billington, J. G. Songer, B. A. McClane, R. W. Titball, J. I. Rood, S. B. Melville and I. T. Paulsen, *Genome Res.*, 2006, **16**, 1031–1040.

70 H. Skarin, T. Hafstrom, J. Westerberg and B. Segerman, *BMC Genomics*, 2011, **12**, 185.

71 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.

72 R. K. Pettit, *Mar. Biotechnol.*, 2011, **13**, 1–11.

73 Z. E. Wilson and M. A. Brimble, *Nat. Prod. Rep.*, 2009, **26**, 44–71.

74 H. Bruggemann and G. Gottschalk, *Ann. N. Y. Acad. Sci.*, 2008, **1125**, 73–81.

75 M. Q. Wei, A. Mengesha, D. Good and J. Anne, *Cancer Lett.*, 2008, **259**, 16–27.

76 R. M. Ryan, J. Green and C. E. Lewis, *BioEssays*, 2006, **28**, 84–94.

77 N. Gaitatzis, B. Silakowski, B. Kunze, G. Nordsiek, H. Blocker, G. Hofle and R. Muller, *J. Biol. Chem.*, 2002, **277**, 13082–13090.

78 B. Silakowski, G. Nordsiek, B. Kunze, H. Blocker and R. Muller, *Chem. Biol.*, 2001, **8**, 59–69.

79 S. Keis, C. F. Bennett, V. K. Ward and D. T. Jones, *Int. J. Syst. Bacteriol.*, 1995, **45**, 693–705.

80 Y. Shi, Y. X. Li and Y. Y. Li, *BMC Bioinformatics*, 2010, **11**, S9.

81 X. H. Chen, A. Koumoutsi, R. Scholz, A. Eisenreich, K. Schneider, I. Heinemeyer, B. Morgenstern, B. Voss, W. R. Hess, O. Reva, H. Junge, B. Voigt, P. R. Jungblut, J. Vater, R. Sussmuth, H. Liesegang, A. Strittmatter, G. Gottschalk and R. Borriss, *Nat. Biotechnol.*, 2007, **25**, 1007–1014.

82 X. H. Chen, A. Koumoutsi, R. Scholz, K. Schneider, J. Vater, R. Sussmuth, J. Piel and R. Borriss, *J. Biotechnol.*, 2009, **140**, 27–37.

83 G. Bao, R. Wang, Y. Zhu, H. Dong, S. Mao, Y. Zhang, Z. Chen, Y. Li and Y. Ma, *J. Bacteriol.*, 2011, **193**, 5007–5008.

84 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.

85 S. Hu, H. Zheng, Y. Gu, J. Zhao, W. Zhang, Y. Yang, S. Wang, G. Zhao, S. Yang and W. Jiang, *BMC Genomics*, 2011, **12**, 93.

86 T. J. Smith, K. K. Hill, B. T. Foley, J. C. Detter, A. C. Munk, D. C. Bruce, N. A. Doggett, L. A. Smith, J. D. Marks, G. Xie and T. S. Brettin, *PLoS One*, 2007, **2**, e1271.

87 R.-M. Tian, T. Li, X.-J. Hou, Q. Wang, K. Cai, Y.-N. Liu, X. Gao, H. Liu, L. Xiao, W. Tu, J. Shi, W.-C. Cao and H. Wang, *Genome*, 2011, **54**, 546–554.

88 A. T. Carter, B. M. Pearson, L. C. Crossman, N. Drou, D. Heavens, D. Baker, M. Febrer, M. Caccamo, K. A. Grant and M. W. Peck, *J. Bacteriol.*, 2011, **193**, 2351–2352.

89 M. Sebaihia, M. W. Peck, N. P. Minton, N. R. Thomson, M. T. Holden, W. J. Mitchell, A. T. Carter, S. D. Bentley, D. R. Mason, L. Crossman, C. J. Paul, A. Ivens, M. H. Wells-Bennik, I. J. Davis, A. M. Cerdeno-Tarraga, C. Churcher, M. A. Quail, T. Chillingworth, T. Feltwell, A. Fraser, I. Goodhead, Z. Hance, K. Jagels, N. Larke, M. Maddison, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, B. White, S. Whithead and J. Parkhill, *Genome Res.*, 2007, **17**, 1082–1092.

90 C. L. Hemme, H. Mouttaki, Y. J. Lee, G. Zhang, L. Goodwin, S. Lucas, A. Copeland, A. Lapidus, T. Glavina del Rio, H. Tice, E. Saunders, T. Brettin, J. C. Detter, C. S. Han, S. Pitluck, M. L. Land, L. J. Hauser, N. Kyrpides, N. Mikhailova, Z. He, L. Wu, J. D. Van Nostrand, B. Henrissat, Q. He, P. A. Lawson, R. S. Tanner, L. R. Lynd, J. Wiegel, M. W. Fields, A. P. Arkin, C. W. Schadt, B. S. Stevenson, M. J. McInerney, Y. Yang, H. Dong, D. Xing, N. Ren, A. Wang, R. L. Huhnke, J. R. Mielenz, S. Y. Ding, M. E. Himmel, S. Taghavi, D. van der Lelie, E. M. Rubin and J. Zhou, *J. Bacteriol.*, 2010, **192**, 6494–6496.

91 Y. Tamaru, H. Miyake, K. Kuroda, A. Nakanishi, Y. Kawade, K. Yamamoto, M. Uemura, Y. Fujita, R. H. Doi and M. Ueda, *J. Bacteriol.*, 2010, **192**, 901–902.

92 J. A. Izquierdo, L. Goodwin, K. W. Davenport, H. Teshima, D. Bruce, C. Detter, R. Tapia, S. Han, M. Land, L. Hauser, C. D. Jeffries, J. Han, S. Pitluck, M. Nolan, A. Chen, M. Huntemann, K. Mavromatis, N. Mikhailova, K. Liolios, T. Woyke and L. R. Lynd, *Stand. Genomic Sci.*, 2012, **6**, 104–115.

93 M. Sebaihia, B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R. Thomson, A. P. Roberts, A. M. Cerdeno-Tarraga, H. Wang, M. T. Holden, A. Wright, C. Churcher, M. A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S. Moule, K. Mungall, C. Price, E. Rabbinowitsch, S. Sharp, M. Simmonds, K. Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell and J. Parkhill, *Nat. Genet.*, 2006, **38**, 779–786.

94 R. A. Stabler, M. He, L. Dawson, M. Martin, E. Valiente, C. Corton, T. D. Lawley, M. Sebaihia, M. A. Quail, G. Rose, D. N. Gerding, M. Gibert, M. R. Popoff, J. Parkhill, G. Dougan and B. W. Wren, *Genome Biol.*, 2009, **10**, R102.

95 M. Kopke, C. Held, S. Hujer, H. Liesegang, A. Wiezer, A. Wollherr, A. Ehrenreich, W. Liebl, G. Gottschalk and P. Durre, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13087–13092.

96 C. Bettegowda, X. Huang, J. Lin, I. Cheong, M. Kohli, S. A. Szabo, X. Zhang, L. A. Diaz, Jr., V. E. Velculescu, G. Parmigiani, K. W. Kinzler, B. Vogelstein and S. Zhou, *Nat. Biotechnol.*, 2006, **24**, 1573–1580.

97 T. Shimizu, K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori,

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

S. Kuhara and H. Hayashi, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 996–1001.
98 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.
99 S. D. Brown, B. Raman, C. K. McKeown, S. P. Kale, Z. He and J. R. Mielenz, *Appl. Biochem. Biotechnol.*, 2007, **137–140**, 663–674.
100 N. Fonknechten, S. Chaussonnerie, S. Tricot, A. Lajus, J. R. Andreesen, N. Perchat, E. Pelletier, M. Gouyvenoux, V. Barbe, M. Salanoubat, D. Le Paslier, J. Weissenbach, G. N. Cohen and A. Kreimeyer, *BMC Genomics*, 2010, **11**, 555.
101 S. Yokoyama, K. Oshima, I. Nomura, M. Hattori and T. Suzuki, *J. Bacteriol.*, 2011, **193**, 5568–5569.
102 R. Gheshlaghi, J. M. Scharer, M. Moo-Young and C. P. Chou, *Biotechnol. Adv.*, 2009, **27**, 764–781.
103 S. Akira, *Semin. Immunol.*, 2004, **16**, 1–2.
104 L. A. O'Neill and A. G. Bowie, *Nat. Rev. Immunol.*, 2007, **7**, 353–364.
105 W. Deng, H. Wang and J. Xie, *J. Cell. Biochem.*, 2011, **112**, 2655–2662.
106 K. Wu, L. Chung, W. P. Revill, L. Katz and C. D. Reeves, *Gene*, 2000, **251**, 81–90.
107 Y. A. Chan, M. T. Boyne, 2nd, A. M. Podevels, A. K. Klimowicz, J. Handelsman, N. L. Kelleher and M. G. Thomas, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 14349–14354.
108 D. G. Olson, S. A. Tripathi, R. J. Giannone, J. Lo, N. C. Caiazza, D. A. Hogsett, R. L. Hettich, A. M. Guss, G. Dubrovsky and L. R. Lynd, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17727–17732.
109 D. A. Argyros, S. A. Tripathi, T. F. Barrett, S. R. Rogers, L. F. Feinberg, D. G. Olson, J. M. Foden, B. B. Miller, L. R. Lynd, D. A. Hogsett and N. C. Caiazza, *Appl. Environ. Microbiol.*, 2011, **77**, 8288–8294.
110 C. N. Shulse and E. E. Allen, *PLoS One*, 2011, **6**, e20146.
111 M. Desvaux, *FEMS Microbiol. Rev.*, 2005, **29**, 741–764.
112 E. Petitdemange, F. Caillet, J. Giallo and C. Gaudin, *Int. J. Syst. Bacteriol.*, 1984, **34**, 155–159.
113 F. Kloss, T. Lincke and C. Hertweck, *Eur. J. Org. Chem.*, 2011, 1429–1431.
114 S. Behnken, T. Lincke, F. Kloss, K. Ishida and C. Hertweck, *Angew. Chem., Int. Ed.*, 2011, **51**, 2425–2428.
115 S. Behnken and C. Hertweck, *PLoS One*, 2011, **7**, e29609.
116 L. Heinlen and J. D. Ballard, *Am. J. Med. Sci.*, 2010, **340**, 247–252.
117 J. Sobel, *Clin. Infect. Dis.*, 2005, **41**, 1167–1173.
118 P. Katona, *Anaerobe*, 2012, **18**, 240–243.
119 S. Yokoyama, T. Niwa, T. Osawa and T. Suzuki, *Arch. Microbiol.*, 2010, **192**, 15–22.
120 C. Atkinson, C. L. Frankenfeld and J. W. Lampe, *Exp. Biol. Med.*, 2005, **230**, 155–170.
121 C. L. Frankenfeld, *Adv. Nutr.*, 2011, **2**, 317–324.
122 H. A. Barker and S. M. Taha, *J. Bacteriol.*, 1942, **43**, 347–363.
123 Z. Heather, M. T. Holden, K. F. Steward, J. Parkhill, L. Song, G. L. Challis, C. Robinson, N. Davis-Poynter and A. S. Waller, *Mol. Microbiol.*, 2008, **70**, 1274–1292.
124 W. M. Sattley, M. T. Madigan, W. D. Swingley, P. C. Cheung, K. M. Clocksin, A. L. Conrad, L. C. Dejesa, B. M. Honchak, D. O. Jung, L. E. Karbach, A. Kurdoglu, S. Lahiri, S. D. Mastrian, L. E. Page, H. L. Taylor, Z. T. Wang, J. Raymond, M. Chen, R. E. Blankenship and J. W. Touchman, *J. Bacteriol.*, 2008, **190**, 4687–4696.
125 C. Nakano, H. Ozawa, G. Akanuma, N. Funa and S. Horinouchi, *J. Bacteriol.*, 2009, **191**, 4916–4923.
126 K. Labutti, R. Pukall, K. Steenblock, T. Glavina Del Rio, H. Tice, A. Copeland, J. F. Cheng, S. Lucas, F. Chen, M. Nolan, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, P. Chain, E. Saunders, T. Brettin, J. C. Detter, C. Han, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and A. Lapidus, *Stand. Genomic Sci.*, 2009, **1**, 159–165.
127 T. Ezaki, Y. Kawamura, N. Li, Z. Y. Li, L. Zhao and S. Shu, *Int. J. Sys. Evol. Microbiol.*, 2001, **51**, 1521–1528.
128 H. Roh, H. J. Ko, D. Kim, D. G. Choi, S. Park, S. Kim, I. S. Chang and I. G. Choi, *J. Bacteriol.*, 2011, **193**, 307–308.
129 K. L. Kavanagh, H. Jornvall, B. Persson and U. Oppermann, *Cell. Mol. Life Sci.*, 2008, **65**, 3895–3906.
130 S. Wu, J. Zhong and L. Huan, *Biochem. Biophys. Res. Commun.*, 2006, **344**, 1147–1154.
131 A. Barcenilla, S. E. Pryde, J. C. Martin, S. H. Duncan, C. S. Stewart, C. Henderson and H. J. Flint, *Appl. Environ. Microbiol.*, 2000, **66**, 1654–1661.
132 M. A. Mahowald, F. E. Rey, H. Seedorf, P. J. Turnbaugh, R. S. Fulton, A. Wollam, N. Shah, C. Wang, V. Magrini, R. K. Wilson, B. L. Cantarel, P. M. Coutinho, B. Henrissat, L. W. Crock, A. Russell, N. C. Verberkmoes, R. L. Hettich and J. I. Gordon, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5859–5864.
133 J. R. Sieber, D. R. Sims, C. Han, E. Kim, A. Lykidis, A. L. Lapidus, E. McDonnald, L. Rohlin, D. E. Culley, R. Gunsalus and M. J. McInerney, *Environ. Microbiol.*, 2010, **12**, 2289–2301.
134 M. J. McInerney, C. G. Struchtemeyer, J. Sieber, H. Mouttaki, A. J. Stams, B. Schink, L. Rohlin and R. P. Gunsalus, *Ann. N. Y. Acad. Sci.*, 2008, **1125**, 58–72.
135 M. J. McInerney, M. P. Bryant, R. B. Hespell and J. W. Costerton, *Appl. Environ. Microbiol.*, 1981, **41**, 1029–1039.
136 W. D. Murray, L. Hoffman, N. L. Campbell and R. H. Madden, *Syst. Appl. Microbiol.*, 1986, **8**, 181–184.
137 D. A. Miller, G. Suen, D. Bruce, A. Copeland, J. F. Cheng, C. Detter, L. A. Goodwin, C. S. Han, L. J. Hauser, M. L. Land, A. Lapidus, S. Lucas, L. Meincke, S. Pitluck, R. Tapia, H. Teshima, T. Woyke, B. G. Fox, E. R. Angert and C. R. Currie, *J. Bacteriol.*, 2011, **193**, 2357–2358.
138 G. Suen, D. M. Stevenson, D. C. Bruce, O. Chertkov, A. Copeland, J. F. Cheng, C. Detter, J. C. Detter, L. A. Goodwin, C. S. Han, L. J. Hauser, N. N. Ivanova,

This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

N. C. Kyrpides, M. L. Land, A. Lapidus, S. Lucas, G. Ovchinnikova, S. Pitluck, R. Tapia, T. Woyke, J. Boyum, D. Mead and P. J. Weimer, *J. Bacteriol.*, 2011, **193**, 5574–5575.

139 T. Iino, K. Mori, K. Tanaka, K. Suzuki and S. Harayama, *Int. J. Syst. Evol. Microbiol.*, 2007, **57**, 1840–1845.

140 M. Rogosa, *Int. J. Syst. Evol. Microbiol.*, 1971, **21**, 234–237.

141 T. Ezaki, in *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes.*, ed. P. De Vos, Garrity, Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K.-H. Schleifer and W. B. Whitman, 2009.

142 S. Spring, A. Lapidus, M. Schroder, D. Gleim, D. Sims, L. Meincke, T. Glavina Del Rio, H. Tice, A. Copeland, J. F. Cheng, S. Lucas, F. Chen, M. Nolan, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, P. Chain, E. Saunders, T. Brettin, J. C. Detter, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and C. Han, *Stand. Genomic Sci.*, 2009, **1**, 242–253.

143 S. H. Kim, C. Harzman, J. K. Davis, R. Hutcheson, J. B. Broderick, T. L. Marsh and J. M. Tiedje, *BMC Microbiol.*, 2012, **12**, 21.

144 T. Madsen and D. Licht, *Appl. Environ. Microbiol.*, 1992, **58**, 2874–2878.

145 A. Suyama, R. Iwakiri, K. Kai, T. Tokunaga, N. Sera and K. Furukawa, *Biosci., Biotechnol., Biochem.*, 2001, **65**, 1474–1481.

146 C. Han, R. Mwirichia, O. Chertkov, B. Held, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, L. Goodwin, S. Pitluck, M. Huntemann, K. Liolios, N. Ivanova, I. Pagani, K. Mavromatis, G. Ovchinikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, E. M. Brambilla, M. Rohde, S. Spring, J. Sikorski, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and J. C. Detter, *Stand. Genomic Sci.*, 2011, **4**, 371–380.

147 J. E. Becker, R. E. Moore and B. S. Moore, *Gene*, 2004, **325**, 35–42.

148 S. E. Blumer-Schuette, I. Ozdemir, D. Mistry, S. Lucas, A. Lapidus, J. F. Cheng, L. A. Goodwin, S. Pitluck, M. L. Land, L. J. Hauser, T. Woyke, N. Mikhailova, A. Pati, N. C. Kyrpides, N. Ivanova, J. C. Detter, K. Walston-Davenport, S. Han, M. W. Adams and R. M. Kelly, *J. Bacteriol.*, 2011, **193**, 1483–1484.

149 N. M. Mesbah, D. B. Hedrick, A. D. Peacock, M. Rohde and J. Wiegel, *Int. J. Syst. Evol. Microbiol.*, 2007, **57**, 2507–2512.

150 H. Marchandin, C. Teyssier, J. Campos, H. Jean-Pierre, F. Roger, B. Gay, J. P. Carlier and E. Jumas-Bilak, *Int. J. Syst. Evol. Microbiol.*, 2010, **60**, 1271–1279.

151 T. Goto, A. Yamashita, H. Hirakawa, M. Matsutani, K. Todo, K. Ohshima, H. Toh, K. Miyamoto, S. Kuhara, M. Hattori, T. Shimizu and S. Akimoto, *DNA Res.*, 2008, **15**, 39–47.

152 K. Ueda, A. Yamashita, J. Ishikawa, M. Shimada, T. O. Watsuji, K. Morimura, H. Ikeda, M. Hattori and T. Beppu, *Nucleic Acids Res.*, 2004, **32**, 4937–4944.

153 O. D. Djao, X. Zhang, S. Lucas, A. Lapidus, T. G. Del Rio, M. Nolan, H. Tice, J. F. Cheng, C. Han, R. Tapia, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, G. Ovchinnikova, A. Pati, E. Brambilla, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, M. Rohde, J. Sikorski, S. Spring, M. Goker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **3**, 268–275.

154 W. J. Kelly, S. C. Leahy, E. Altermann, C. J. Yeoman, J. C. Dunne, Z. Kong, D. M. Pacheco, D. Li, S. J. Noel, C. D. Moon, A. L. Cookson and G. T. Attwood, *PLoS One*, 2010, **5**, e11942.

155 Y. Katano, S. Fujinami, A. Kawakoshi, H. Nakazawa, S. Oji, T. Iino, A. Oguchi, A. Ankai, S. Fukui, Y. Terui, S. Kamata, T. Harada, S. Tanikawa, K. Suzuki and N. Fujita, *Stand. Genomic Sci.*, 2012, **6**, 406–414.

156 P. Junier, T. Junier, S. Podell, D. R. Sims, J. C. Detter, A. Lykidis, C. S. Han, N. S. Wigginton, T. Gaasterland and R. Bernier-Latmani, *Environ. Microbiol.*, 2010, **12**, 2738–2754.

157 D. Chivian, E. L. Brodie, E. J. Alm, D. E. Culley, P. S. Dehal, T. Z. DeSantis, T. M. Gihring, A. Lapidus, L. H. Lin, S. R. Lowry, D. P. Moser, P. M. Richardson, G. Southam, G. Wanger, L. M. Pratt, G. L. Andersen, T. C. Hazen, F. J. Brockman, A. P. Arkin and T. C. Onstott, *Science*, 2008, **322**, 275–278.

158 H. Nonaka, G. Keresztes, Y. Shinoda, Y. Ikenaga, M. Abe, K. Naito, K. Inatomi, K. Furukawa, M. Inui and H. Yukawa, *J. Bacteriol.*, 2006, **188**, 2262–2274.

159 T. Kosaka, S. Kato, T. Shimoyama, S. Ishii, T. Abe and K. Watanabe, *Genome Res.*, 2008, **18**, 442–448.

160 K. G. Byrne-Bailey, K. C. Wrighton, R. A. Melnyk, P. Agbo, T. C. Hazen and J. D. Coates, *J. Bacteriol.*, 2010, **192**, 4078–4079.

161 C. L. Hemme, M. W. Fields, Q. He, Y. Deng, L. Lin, Q. Tu, H. Mouttaki, A. Zhou, X. Feng, Z. Zuo, B. D. Ramsay, Z. He, L. Wu, J. Van Nostrand, J. Xu, Y. J. Tang, J. Wiegel, T. J. Phelps and J. Zhou, *Appl. Environ. Microbiol.*, 2011, **77**, 7998–8008.

162 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.

163 I. A. Kataeva, S. J. Yang, P. Dam, F. L. Poole, 2nd, Y. Yin, F. Zhou, W. C. Chou, Y. Xu, L. Goodwin, D. R. Sims, J. C. Detter, L. J. Hauser, J. Westpheling and M. W. Adams, *J. Bacteriol.*, 2009, **191**, 3760–3761.

164 H. J. van de Werken, M. R. Verhaart, A. L. VanFossen, K. Willquist, D. L. Lewis, J. D. Nichols, H. P. Goorissen, E. F. Mongodin, K. E. Nelson, E. W. van Niel, A. J. Stams, D. E. Ward, W. M. de Vos, J. van der Oost, R. M. Kelly and S. W. Kengen, *Appl. Environ. Microbiol.*, 2008, **74**, 6720–6729.

165 J. G. Elkins, A. Lochner, S. D. Hamilton-Brehm, K. W. Davenport, M. Podar, S. D. Brown, M. L. Land,

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

L. J. Hauser, D. M. Klingeman, B. Raman, L. A. Goodwin, R. Tapia, L. J. Meincke, J. C. Detter, D. C. Bruce, C. S. Han, A. V. Palumbo, R. W. Cottingham, M. Keller and D. E. Graham, *J. Bacteriol.*, 2010, **192**, 6099–6100.

166 E. Pierce, G. Xie, R. D. Barabote, E. Saunders, C. S. Han, J. C. Detter, P. Richardson, T. S. Brettin, A. Das, L. G. Ljungdahl and S. W. Ragsdale, *Environ. Microbiol.*, 2008, **10**, 2550–2573.

167 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.

168 S. Pitluck, M. Yasawong, C. Munk, M. Nolan, A. Lapidus, S. Lucas, T. Glavina Del Rio, H. Tice, J. F. Cheng, D. Bruce, C. Detter, R. Tapia, C. Han, L. Goodwin, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, M. Rohde, S. Spring, J. Sikorski, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **3**, 108–116.

169 J. Sikorski, H. Teshima, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, M. Huntemann, K. Mavromatis, G. Ovchinikova, A. Pati, R. Tapia, C. Han, L. Goodwin, A. Chen, K. Palaniappan, M. Land, L. Hauser, O. D. Ngatchou-Djao, M. Rohde, R. Pukall, S. Spring, B. Abt, M. Goker, J. C. Detter, T. Woyke, J. Bristow, V. Markowitz, P. Hugenholtz, J. A. Eisen, N. C. Kyrpides, H. P. Klenk and A. Lapidus, *Stand. Genomic Sci.*, 2011, **4**, 331–341.

170 B. Zhao, N. M. Mesbah, E. Dalin, L. Goodwin, M. Nolan, S. Pitluck, O. Chertkov, T. S. Brettin, J. Han, F. W. Larimer, M. L. Land, L. Hauser, N. Kyrpides and J. Wiegel, *J. Bacteriol.*, 2011, **193**, 4023–4024.

171 S. D. Brown, M. B. Begemann, M. R. Mormile, J. D. Wall, C. S. Han, L. A. Goodwin, S. Pitluck, M. L. Land, L. J. Hauser and D. A. Elias, *J. Bacteriol.*, 2011, **193**, 3682–3683.

172 N. Ivanova, J. Sikorski, O. Chertkov, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, S. Pitluck, M. Huntemann, K. Liolios, I. Pagani, K. Mavromatis, G. Ovchinikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, E. M. Brambilla, K. P. Kannan, M. Rohde, B. J. Tindall, M. Goker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and A. Lapidus, *Stand. Genomic Sci.*, 2011, **4**, 312–321.

173 K. Mavromatis, N. Ivanova, I. Anderson, A. Lykidis, S. D. Hooper, H. Sun, V. Kunin, A. Lapidus, P. Hugenholtz, B. Patel and N. C. Kyrpides, *PLoS One*, 2009, **4**, e4192.

174 J. Sikorski, A. Lapidus, O. Chertkov, S. Lucas, A. Copeland, T. Glavina Del Rio, M. Nolan, H. Tice, J. F. Cheng, C. Han, E. Brambilla, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, D. Bruce, C. Detter, R. Tapia, L. Goodwin, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, M. Rohde, M. Goker, S. Spring, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **3**, 57–65.

175 Y. J. Chang, R. Pukall, E. Saunders, A. Lapidus, A. Copeland, M. Nolan, T. Glavina Del Rio, S. Lucas, F. Chen, H. Tice, J. F. Cheng, C. Han, J. C. Detter, D. Bruce, L. Goodwin, S. Pitluck, N. Mikhailova, K. Liolios, A. Pati, N. Ivanova, K. Mavromatis, A. Chen, K. Palaniappan, M. Land, L. Hauser, C. D. Jeffries, T. Brettin, M. Rohde, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **3**, 1–14.

176 G. D'Auria, J. C. Galan, M. Rodriguez-Alcayna, A. Moya, F. Baquero and A. Latorre, *J. Bacteriol.*, 2011, **193**, 7008–7009.

177 S. Gronow, S. Welnitz, A. Lapidus, M. Nolan, N. Ivanova, T. Glavina Del Rio, A. Copeland, F. Chen, H. Tice, S. Pitluck, J. F. Cheng, E. Saunders, T. Brettin, C. Han, J. C. Detter, D. Bruce, L. Goodwin, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, A. Pati, K. Mavromatis, N. Mikhailova, A. Chen, K. Palaniappan, P. Chain, M. Rohde, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and S. Lucas, *Stand. Genomic Sci.*, 2009, **2**, 57–65.

178 S. Patrick, G. W. Blakely, S. Houston, J. Moore, V. R. Abratt, M. Bertalan, A. M. Cerdeno-Tarraga, M. A. Quail, N. Corton, C. Corton, A. Bignell, A. Barron, L. Clark, S. D. Bentley and J. Parkhill, *Microbiology*, 2010, **156**, 3255–3269.

179 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.

180 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.

181 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.

182 J. Xu, M. A. Mahowald, R. E. Ley, C. A. Lozupone, M. Hamady, E. C. Martens, B. Henrissat, P. M. Coutinho, P. Minx, P. Latreille, H. Cordum, A. Van Brunt, K. Kim, R. S. Fulton, L. A. Fulton, S. W. Clifton, R. K. Wilson, R. D. Knight and J. I. Gordon, *PLoS Biol.*, 2007, **5**, e156.

183 A. Pati, S. Gronow, A. Zeytun, A. Lapidus, M. Nolan, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, K. Mavromatis, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, J. C. Detter, E. Brambilla, M. Rohde, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and S. Lucas, *Stand. Genomic Sci.*, 2011, **4**, 45–53.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

184 S. Gronow, B. Held, S. Lucas, A. Lapidus, T. G. Del Rio, M. Nolan, H. Tice, S. Deshpande, J. F. Cheng, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, K. Mavromatis, A. Pati, R. Tapia, C. Han, L. Goodwin, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, E. M. Brambilla, M. Rohde, M. Goker, J. C. Detter, T. Woyke, J. Bristow, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and J. A. Eisen, *Stand. Genomic Sci.*, 2011, **4**, 191–199.

185 M. Naito, H. Hirakawa, A. Yamashita, N. Ohara, M. Shoji, H. Yukitake, K. Nakayama, H. Toh, F. Yoshimura, S. Kuhara, M. Hattori, T. Hayashi and K. Nakayama, *DNA Res.*, 2008, **15**, 215–225.

186 T. Watanabe, F. Maruyama, T. Nozawa, A. Aoki, S. Okano, Y. Shibata, K. Oshima, K. Kurokawa, M. Hattori, I. Nakagawa and Y. Abiko, *J. Bacteriol.*, 2011, **193**, 4259–4260.

187 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.

188 M. Goker, S. Gronow, A. Zeytun, M. Nolan, S. Lucas, A. Lapidus, N. Hammon, S. Deshpande, J. F. Cheng, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, K. Mavromatis, G. Ovchinikova, A. Pati, R. Tapia, C. Han, L. Goodwin, A. Chen, K. Palaniappan, M. Land, L. Hauser, C. D. Jeffries, E. M. Brambilla, M. Rohde, J. C. Detter, T. Woyke, J. Bristow, V. Markowitz, P. Hugenholtz, J. A. Eisen, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2011, **4**, 200–209.

189 S. Gronow, C. Munk, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, E. Brambilla, M. Rohde, M. Goker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2011, **4**, 36–44.

190 J. Purushe, D. E. Fouts, M. Morrison, B. A. White, R. I. Mackie, P. M. Coutinho, B. Henrissat and K. E. Nelson, *Microb. Ecol.*, 2010, **60**, 721–729.

191 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.

192 H. Okuyama, Y. Orikasa, T. Nishida, K. Watanabe and N. Morita, *Appl. Environ. Microbiol.*, 2007, **73**, 665–670.

193 U. Kaulmann and C. Hertweck, *Angew. Chem., Int. Ed.*, 2002, **41**, 1866–1869.

194 D. J. Sukovich, J. L. Seffernick, J. E. Richman, K. A. Hunt, J. A. Gralnick and L. P. Wackett, *Appl. Environ. Microbiol.*, 2010, **76**, 3842–3849.

195 D. J. Sukovich, J. L. Seffernick, J. E. Richman, J. A. Gralnick and L. P. Wackett, *Appl. Environ. Microbiol.*, 2010, **76**, 3850–3862.

196 A. W. Strittmatter, H. Liesegang, R. Rabus, I. Decker, J. Amann, S. Andres, A. Henne, W. F. Fricke, R. Martinez-Arias, D. Bartels, A. Goesmann, L. Krause, A. Puhler, H. P. Klenk, M. Richter, M. Schuler, F. O. Glockner, A. Meyerdierks, G. Gottschalk and R. Amann, *Environ. Microbiol.*, 2009, **11**, 1038–1055.

197 M. Huntemann, M. Lu, M. Nolan, A. Lapidus, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, G. Ovchinikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, C. D. Jeffries, J. C. Detter, E. M. Brambilla, M. Rohde, S. Spring, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and K. Mavromatis, *Stand. Genomic Sci.*, 2011, **4**, 303–311.

198 H. Sun, S. Spring, A. Lapidus, K. Davenport, T. G. Del Rio, H. Tice, M. Nolan, A. Copeland, J. F. Cheng, S. Lucas, R. Tapia, L. Goodwin, S. Pitluck, N. Ivanova, I. Pagani, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, L. Hauser, Y. J. Chang, C. D. Jeffries, J. C. Detter, C. Han, M. Rohde, E. Brambilla, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and M. Land, *Stand. Genomic Sci.*, 2010, **3**, 276–284.

199 H. T. Tran, J. Krushkal, F. M. Antommattei, D. R. Lovley and R. M. Weis, *BMC Genomics*, 2008, **9**, 471.

200 D. D. Wagner, L. A. Hug, J. K. Hatt, M. A. Spitzmiller, E. Padilla-Crespo, K. M. Ritalahti, E. A. Edwards, K. T. Konstantinidis and F. E. Loffler, *BMC Genomics*, 2012, **13**, 200.

201 M. Aklujkar, N. D. Young, D. Holmes, M. Chavan, C. Risso, H. E. Kiss, C. S. Han, M. L. Land and D. R. Lovley, *BMC Genomics*, 2010, **11**, 490.

202 H. Nagarajan, J. E. Butler, A. Klimes, Y. Qiu, K. Zengler, J. Ward, N. D. Young, B. A. Methe, B. O. Palsson, D. R. Lovley and C. L. Barrett, *PLoS One*, 2010, **5**, e10922.

203 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.

204 J. E. Butler, N. D. Young and D. R. Lovley, *BMC Genomics*, 2010, **11**, 40.

205 M. J. McInerney, L. Rohlin, H. Mouttaki, U. Kim, R. S. Krupp, L. Rios-Hernandez, J. Sieber, C. G. Struchtemeyer,

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

A. Bhattacharyya, J. W. Campbell and R. P. Gunsalus, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7600–7605.
206 M. Goker, H. Teshima, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, S. Pitluck, M. Huntemann, K. Liolios, N. Ivanova, I. Pagani, K. Mavromatis, G. Ovchinikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, E. M. Brambilla, M. Rohde, S. Spring, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2011, **4**, 393–401.
207 A. Copeland, S. Spring, M. Goker, S. Schneider, A. Lapidus, T. G. Del Rio, H. Tice, J. F. Cheng, F. Chen, M. Nolan, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavrommatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. C. Jeffries, L. Meincke, D. Sims, T. Brettin, J. C. Detter, C. Han, P. Chain, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and S. Lucas, *Stand. Genomic Sci.*, 2009, **1**, 29–37.
208 C. B. Walker, S. Stolyar, D. Chivian, N. Pinel, J. A. Gabster, P. S. Dehal, Z. He, Z. K. Yang, H. C. Yen, J. Zhou, J. D. Wall, T. C. Hazen, A. P. Arkin and D. A. Stahl, *Environ. Microbiol.*, 2009, **11**, 2244–2252.
209 H. Nakazawa, A. Arakaki, S. Narita-Yamada, I. Yashiro, K. Jinno, N. Aoki, A. Tsuruyama, Y. Okamura, S. Tanikawa, N. Fujita, H. Takeyama and T. Matsunaga, *Genome Res.*, 2009, **19**, 1801–1808.
210 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.
211 L. J. Hauser, M. L. Land, S. D. Brown, F. Larimer, K. L. Keller, B. J. Rapp-Giles, M. N. Price, M. Lin, D. C. Bruce, J. C. Detter, R. Tapia, C. S. Han, L. A. Goodwin, J. F. Cheng, S. Pitluck, A. Copeland, S. Lucas, M. Nolan, A. L. Lapidus, A. V. Palumbo and J. D. Wall, *J. Bacteriol.*, 2011, **193**, 4268–4269.
212 S. Spring, M. Nolan, A. Lapidus, T. Glavina Del Rio, A. Copeland, H. Tice, J. F. Cheng, S. Lucas, M. Land, F. Chen, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, L. Hauser, Y. J. Chang, C. D. Jeffries, C. Munk, H. Kiss, P. Chain, C. Han, T. Brettin, J. C. Detter, E. Schuler, M. Goker, M. Rohde, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **2**, 38–48.
213 A. V. Callaghan, B. E. Morris, I. A. Pereira, M. J. McInerney, R. N. Austin, J. T. Groves, J. J. Kukor, J. M. Suflita, L. Y. Young, G. J. Zylstra and B. Wawrik, *Environ. Microbiol.*, 2012, **14**, 101–113.
214 I. Pagani, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, O. Chertkov, K. Davenport, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, K. Mavromatis, N. Ivanova, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, J. C. Detter, E. Brambilla, K. P. Kannan, O. D. Djao, M. Rohde, R. Pukall, S. Spring, M. Goker, J. Sikorski, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2011, **4**, 100–110.
215 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.
216 B. J. Campbell, J. L. Smith, T. E. Hanson, M. G. Klotz, L. Y. Stein, C. K. Lee, D. Wu, J. M. Robinson, H. M. Khouri, J. A. Eisen and S. C. Cary, *PLoS Genet.*, 2009, **5**, e1000362.
217 G. S. Myers, D. Parker, K. Al-Hasani, R. M. Kennan, T. Seemann, Q. Ren, J. H. Badger, J. D. Selengut, R. T. Deboy, H. Tettelin, J. D. Boyce, V. P. McCarl, X. Han, W. C. Nelson, R. Madupu, Y. Mohamoud, T. Holley, N. Fedorova, H. Khouri, S. P. Bottomley, R. J. Whittington, B. Adler, J. G. Songer, J. I. Rood and I. T. Paulsen, *Nat. Biotechnol.*, 2007, **25**, 569–575.
218 D. R. Lovley, D. E. Holmes and K. P. Nevin, *Adv. Microb. Physiol.*, 2004, **49**, 219–286.
219 D. R. Lovley, T. Ueki, T. Zhang, N. S. Malvankar, P. M. Shrestha, K. A. Flanagan, M. Aklujkar, J. E. Butler, L. Giloteaux, A. E. Rotaru, D. E. Holmes, A. E. Franks, R. Orellana, C. Risso and K. P. Nevin, *Adv. Microb. Physiol.*, 2011, **59**, 1–100.
220 S. J. Waddell, G. A. Chung, K. J. Gibson, M. J. Everett, D. E. Minnikin, G. S. Besra and P. D. Butcher, *Lett. Appl. Microbiol.*, 2005, **40**, 201–206.
221 P. Saxena, G. Yadav, D. Mohanty and R. S. Gokhale, *J. Biol. Chem.*, 2003, **278**, 44780–44790.
222 C. M. Sassetti, D. H. Boyd and E. J. Rubin, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 12712–12717.
223 M. G. Muller, L. S. Forsberg and D. H. Keating, *J. Bacteriol.*, 2009, **191**, 6988–7000.
224 J. Cuccui, A. Easton, K. K. Chu, G. J. Bancroft, P. C. Oyston, R. W. Titball and B. W. Wren, *Infect. Immun.*, 2007, **75**, 1186–1195.
225 D. E. Holmes, R. A. O'Neil, M. A. Chavan, L. A. N'Guessan, H. A. Vrionis, L. A. Perpetua, M. J. Larrahondo, R. DiDonato, A. Liu and D. R. Lovley, *ISME J.*, 2009, **3**, 216–230.
226 S. J. Oude Elferink, W. M. Akkermans-van Vliet, J. J. Bogte and A. J. Stams, *Int. J. Syst. Bacteriol.*, 1999, **49 Pt 2**, 345–350.
227 T. Sakaguchi, A. Arakaki and T. Matsunaga, *Int. J. Sys. Evol. Microbiol.*, 2002, **52**, 215–221.
228 D. R. Lovley and E. J. Phillips, *Appl. Environ. Microbiol.*, 1992, **58**, 850–856.
229 W. P. Iverson, *Nature*, 1968, **217**, 1265–1267.

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

230 A. Dhillon, A. Teske, J. Dillon, D. A. Stahl and M. L. Sogin, *Appl. Environ. Microbiol.*, 2003, **69**, 2765–2772.
231 H. Shindou, D. Hishikawa, T. Harayama, K. Yuki and T. Shimizu, *J. Lipid Res.*, 2009, **50**, S46–51.
232 M. Ventura, S. O'Flaherty, M. J. Claesson, F. Turroni, T. R. Klaenhammer, D. van Sinderen and P. W. O'Toole, *Nat. Rev. Microbiol.*, 2009, **7**, 61–71.
233 M. O'Connell Motherway, A. Zomer, S. C. Leahy, J. Reunanen, F. Bottacini, M. J. Claesson, F. O'Brien, K. Flynn, P. G. Casey, J. A. Munoz, B. Kearney, A. M. Houston, C. O'Mahony, D. G. Higgins, F. Shanahan, A. Palva, W. M. de Vos, G. F. Fitzgerald, M. Ventura, P. W. O'Toole and D. van Sinderen, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11217–11222.
234 L. Gorissen, K. Raes, S. Weckx, D. Dannenberger, F. Leroy, L. De Vuyst and S. De Smet, *Appl. Microbiol. Biotechnol.*, 2010, **87**, 2257–2266.
235 E. Brzuszkiewicz, J. Weiner, A. Wollherr, A. Thurmer, J. Hupeden, H. B. Lomholt, M. Kilian, G. Gottschalk, R. Daniel, H. J. Mollenkopf, T. F. Meyer and H. Bruggemann, *PLoS One*, 2011, **6**, e21581.
236 H. B. Lomholt and M. Kilian, *PLoS One*, 2010, **5**, e12277.
237 D. Zhurina, A. Zomer, M. Gleinser, V. F. Brancaccio, M. Auchter, M. S. Waidmann, C. Westermann, D. van Sinderen and C. U. Riedel, *J. Bacteriol.*, 2011, **193**, 301–302.
238 F. Turroni, F. Bottacini, E. Foroni, I. Mulder, J. H. Kim, A. Zomer, B. Sanchez, A. Bidossi, A. Ferrarini, V. Giubellini, M. Delledonne, B. Henrissat, P. Coutinho, M. Oggioni, G. F. Fitzgerald, D. Mills, A. Margolles, D. Kelly, D. van Sinderen and M. Ventura, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 19514–19519.
239 S. Fukuda, H. Toh, K. Hase, K. Oshima, Y. Nakanishi, K. Yoshimura, T. Tobe, J. M. Clarke, D. L. Topping, T. Suzuki, T. D. Taylor, K. Itoh, J. Kikuchi, H. Morita, M. Hattori and H. Ohno, *Nature*, 2011, **469**, 543–547.
240 C. Chervaux, C. Grimaldi, A. Bolotin, B. Quinquis, S. Legrain-Raspaud, J. E. van Hylckama Vlieg, G. Denariaz and T. Smokvina, *J. Bacteriol.*, **193**, 5560–5561.
241 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.
242 J. H. Lee, V. N. Karamychev, S. A. Kozyavkin, D. Mills, A. R. Pavlov, N. V. Pavlova, N. N. Polouchine, P. M. Richardson, V. V. Shakhova, A. I. Slesarev, B. Weimer and D. J. O'Sullivan, *BMC Genomics*, 2008, **9**, 247.
243 D. A. Sela, J. Chapman, A. Adeuya, J. H. Kim, F. Chen, T. R. Whitehead, A. Lapidus, D. S. Rokhsar, C. B. Lebrilla, J. B. German, N. P. Price, P. M. Richardson and D. A. Mills, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 18964–18969.
244 Y. X. Wei, Z. Y. Zhang, C. Liu, Y. Z. Zhu, Y. Q. Zhu, H. Zheng, G. P. Zhao, S. Wang and X. K. Guo, *J. Bacteriol.*, 2010, **192**, 4076–4077.
245 Y. Hao, D. Huang, H. Guo, M. Xiao, H. An, L. Zhao, F. Zuo, B. Zhang, S. Hu, S. Song, S. Chen and F. Ren, *J. Bacteriol.*, 2011, **193**, 787–788.
246 J. S. Ham, T. Lee, M. J. Byun, K. T. Lee, M. K. Kim, G. S. Han, S. G. Jeong, M. H. Oh, D. H. Kim and H. Kim, *J. Bacteriol.*, 2011, **193**, 5044.
247 J. F. Kim, H. Jeong, D. S. Yu, S. H. Choi, C. G. Hur, M. S. Park, S. H. Yoon, D. W. Kim, G. E. Ji, H. S. Park and T. K. Oh, *J. Bacteriol.*, 2009, **191**, 678–679.
248 R. Barrangou, E. P. Briczinski, L. L. Traeger, J. R. Loquasto, M. Richards, P. Horvath, A. C. Coute-Monvoisin, G. Leyer, S. Rendulic, J. L. Steele, J. R. Broadbent, T. Oberg, E. G. Dudley, S. Schuster, D. A. Romero and R. F. Roberts, *J. Bacteriol.*, 2009, **191**, 4144–4151.
249 M. Ventura, F. Turroni, A. Zomer, E. Foroni, V. Giubellini, F. Bottacini, C. Canchaya, M. J. Claesson, F. He, M. Mantzourani, L. Mulas, A. Ferrarini, B. Gao, M. Delledonne, B. Henrissat, P. Coutinho, M. Oggioni, R. S. Gupta, Z. Zhang, D. Beighton, G. F. Fitzgerald, P. W. O'Toole and D. van Sinderen, *PLoS Genet.*, 2009, **5**, e1000785.
250 C. Garrigues, E. Johansen and M. B. Pedersen, *J. Bacteriol.*, 2010, **192**, 2467–2468.
251 Z. Sun, X. Chen, J. Wang, P. Gao, Z. Zhou, Y. Ren, T. Sun, L. Wang, H. Meng, W. Chen and H. Zhang, *J. Bacteriol.*, 2010, **192**, 4080–4081.
252 F. Bottacini, F. Dal Bello, F. Turroni, C. Milani, S. Duranti, E. Foroni, A. Viappiani, F. Strati, D. Mora, D. van Sinderen and M. Ventura, *J. Bacteriol.*, 2011, **193**, 6387–6388.
253 J. R. Loquasto, R. Barrangou, E. G. Dudley and R. F. Roberts, *J. Dairy Sci.*, 2011, **94**, 5864–5870.
254 H. Bruggemann, A. Henne, F. Hoster, H. Liesegang, A. Wiezer, A. Strittmatter, S. Hujer, P. Durre and G. Gottschalk, *Science*, 2004, **305**, 671–673.
255 H. Falentin, S. M. Deutsch, G. Jan, V. Loux, A. Thierry, S. Parayre, M. B. Maillard, J. Dherbecourt, F. J. Cousin, J. Jardin, P. Siguier, A. Couloux, V. Barbe, B. Vacherie, P. Wincker, J. F. Gibrat, C. Gaillardin and S. Lortal, *PLoS One*, 2010, **5**, e11748.
256 A. Voros, B. Horvath, J. Hunyadkurti, A. McDowell, E. Barnard, S. Patrick and I. Nagy, *J. Bacteriol.*, 2012, **194**, 1621–1622.
257 J. Hunyadkurti, Z. Feltoti, B. Horvath, M. Nagymihaly, A. Voros, A. McDowell, S. Patrick, E. Urban and I. Nagy, *J. Bacteriol.*, 2011, **193**, 4561–4562.
258 B. Horvath, J. Hunyadkurti, A. Voros, C. Fekete, E. Urban, L. Kemeny and I. Nagy, *J. Bacteriol.*, 2012, **194**, 202–203.
259 R. Pukall, A. Lapidus, M. Nolan, A. Copeland, T. Glavina Del Rio, S. Lucas, F. Chen, H. Tice, J. F. Cheng, O. Chertkov, D. Bruce, L. Goodwin, C. Kuske, T. Brettin, J. C. Detter, C. Han, S. Pitluck, A. Pati, K. Mavrommatis, N. Ivanova, G. Ovchinnikova, A. Chen, K. Palaniappan, S. Schneider, M. Rohde, P. Chain, P. D'Haeseleer, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, N. C. Kyrpides, H. P. Klenk and P. Hugenholtz, *Stand. Genomic Sci.*, 2009, **1**, 234–241.
260 E. Saunders, R. Pukall, B. Abt, A. Lapidus, T. Glavina Del Rio, A. Copeland, H. Tice, J. F. Cheng, S. Lucas, F. Chen, M. Nolan, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang,

View Article Online

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

C. D. Jeffries, P. Chain, L. Meincke, D. Sims, T. Brettin, J. C. Detter, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and C. Han, *Stand. Genomic Sci.*, 2009, **1**, 174–182.

261 K. Mavrommatis, R. Pukall, C. Rohde, F. Chen, D. Sims, T. Brettin, C. Kuske, J. C. Detter, C. Han, A. Lapidus, A. Copeland, T. Glavina Del Rio, M. Nolan, S. Lucas, H. Tice, J. F. Cheng, D. Bruce, L. Goodwin, S. Pitluck, G. Ovchinnikova, A. Pati, N. Ivanova, A. Chen, K. Palaniappan, P. Chain, P. D'Haeseleer, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, M. Rohde, H. P. Klenk and N. C. Kyrpides, *Stand. Genomic Sci.*, 2009, **1**, 93–100.

262 A. Copeland, J. Sikorski, A. Lapidus, M. Nolan, T. G. Del Rio, S. Lucas, F. Chen, H. Tice, S. Pitluck, J. F. Cheng, R. Pukall, O. Chertkov, T. Brettin, C. Han, J. C. Detter, C. Kuske, D. Bruce, L. Goodwin, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Chen, K. Palaniappan, P. Chain, M. Rohde, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and J. C. Detter, *Stand. Genomic Sci.*, 2009, **1**, 166–173.

263 M. Goker, B. Held, S. Lucas, M. Nolan, M. Yasawong, T. Glavina Del Rio, H. Tice, J. F. Cheng, D. Bruce, J. C. Detter, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, M. Rohde, J. Sikorski, R. Pukall, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and A. Lapidus, *Stand. Genomic Sci.*, 2010, **3**, 76–84.

264 A. D. Frock, J. S. Notey and R. M. Kelly, *Environ. Technol.*, 2010, **31**, 1169–1181.

265 O. Zhaxybayeva, K. S. Swithers, P. Lapierre, G. P. Fournier, D. M. Bickhart, R. T. DeBoy, K. E. Nelson, C. L. Nesbo, W. F. Doolittle, J. P. Gogarten and K. M. Noll, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5865–5870.

266 E. A. Bonch-Osmolovskaya, M. L. Miroshnichenko, A. V. Lebedinsky, N. A. Chernyh, T. N. Nazina, V. S. Ivoilov, S. S. Belyaev, E. S. Boulygina, Y. P. Lysov, A. N. Perov, A. D. Mirzabekov, H. Hippe, E. Stackebrandt, S. L'Haridon and C. Jeanthon, *Appl. Environ. Microbiol.*, 2003, **69**, 6143–6151.

267 H. Dahle, F. Garshol, M. Madsen and N. K. Birkeland, *Antonie van Leeuwenhoek*, 2008, **93**, 37–49.

268 K. S. Swithers, J. L. DiPippo, D. C. Bruce, C. Detter, R. Tapia, S. Han, E. Saunders, L. A. Goodwin, J. Han, T. Woyke, S. Pitluck, L. Pennacchio, M. Nolan, N. Mikhailova, A. Lykidis, M. L. Land, T. Brettin, K. O. Stetter, K. E. Nelson, J. P. Gogarten and K. M. Noll, *J. Bacteriol.*, 2011, **193**, 5869–5870.

269 A. D. Frock, S. R. Gray and R. M. Kelly, *Appl. Environ. Microbiol.*, 2012, **78**, 1978–1986.

270 K. E. Nelson, J. A. Eisen and C. M. Fraser, *Methods Enzymol.*, 2001, **330**, 169–180.

271 C. L. Nesbo, E. Bapteste, B. Curtis, H. Dahle, P. Lopez, D. Macleod, M. Dlutek, S. Bowman, O. Zhaxybayeva, N. K. Birkeland and W. F. Doolittle, *J. Bacteriol.*, 2009, **191**, 1974–1978.

272 A. Caro-Quintero, K. M. Ritalahti, K. D. Cusick, F. E. Loffler and K. T. Konstantinidis, *mBio*, 2012, **3**, DOI: 10.1128/mBio.00025-12.

273 P. Matejkova, M. Strouhal, D. Smajs, S. J. Norris, T. Palzkill, J. F. Petrosino, E. Sodergren, J. E. Norton, J. Singh, T. A. Richmond, M. N. Molla, T. J. Albert and G. M. Weinstock, *BMC Microbiol.*, 2008, **8**, 76.

274 L. Giacani, B. M. Jeffrey, B. J. Molini, H. T. Le, S. A. Lukehart, A. Centurion-Lara and D. D. Rockey, *J. Bacteriol.*, 2011, **192**, 2645–2646.

275 C. Han, S. Gronow, H. Teshima, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, A. Zeytun, R. Tapia, L. Goodwin, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, K. Mavromatis, N. Mikhailova, M. Huntemann, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, E. M. Brambilla, M. Rohde, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and J. C. Detter, *Stand. Genomic Sci.*, 2011, **4**, 361–370.

276 A. Z. Rosenthal, E. G. Matson, A. Eldar and J. R. Leadbetter, *ISME J.*, 2011, **5**, 1133–1142.

277 D. Smajs, M. Zobanikova, M. Strouhal, D. Cejkova, S. Dugan-Rocha, P. Pospisilova, S. J. Norris, T. Albert, X. Qin, K. Hallsworth-Pepin, C. Buhay, D. M. Muzny, L. Chen, R. A. Gibbs and G. M. Weinstock, *PLoS One*, 2012, **6**, e20415.

278 D. Cejkova, M. Zobanikova, L. Chen, P. Pospisilova, M. Strouhal, X. Qin, L. Mikalova, S. J. Norris, D. M. Muzny, R. A. Gibbs, L. L. Fulton, E. Sodergren, G. M. Weinstock and D. Smajs, *PLoS Neglected Trop. Dis.*, 2012, **6**, e1471.

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, T. 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 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.

281 K. Mavromatis, M. Yasawong, O. Chertkov, A. Lapidus, S. Lucas, M. Nolan, T. G. Del Rio, H. Tice, J. F. Cheng, S. Pitluck, K. Liolios, N. Ivanova, R. Tapia, C. Han, D. Bruce, L. Goodwin, A. Pati, A. Chen, K. Palaniappan,

 This journal is © The Royal Society of Chemistry 2013

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, J. C. Detter, M. Rohde, E. Brambilla, S. Spring, M. Goker, J. Sikorski, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, H. P. Klenk and N. C. Kyrpides, *Stand. Genomic. Sci.*, 2010, **3**, 136–144.
282 A. Angelov, S. Liebl, M. Ballschmiter, M. Bomeke, R. Lehmann, H. Liesegang, R. Daniel and W. Liebl, *J. Bacteriol.*, 2010, **192**, 6492–6493.
283 B. Abt, C. Han, C. Scheuner, M. Lu, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, L. A. Goodwin, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, K. Mavromatis, N. Mikhailova, M. Huntemann, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, E. M. Brambilla, M. Rohde, S. Spring, S. Gronow, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and J. C. Detter, *Stand. Genomic Sci.*, 2012, **6**, 194–209.
284 M. I. Bellgard, P. Wanchanthuek, T. La, K. Ryan, P. Moolhuijzen, Z. Albertyn, B. Shaban, Y. Motro, D. S. Dunn, D. Schibeci, A. Hunter, R. Barrero, N. D. Phillips and D. J. Hampson, *PLoS One*, 2009, **4**, e4641.
285 P. Wanchanthuek, M. I. Bellgard, T. La, K. Ryan, P. Moolhuijzen, B. Chapman, M. Black, D. Schibeci, A. Hunter, R. Barrero, N. D. Phillips and D. J. Hampson, *PLoS One*, 2010, **5**, e11455.
286 T. Hafstrom, D. S. Jansson and B. Segerman, *BMC Genomics*, 2011, **12**, 395.
287 A. Pati, J. Sikorski, S. Gronow, C. Munk, A. Lapidus, A. Copeland, T. Glavina Del Tio, M. Nolan, S. Lucas, F. Chen, H. Tice, J. F. Cheng, C. Han, J. C. Detter, D. Bruce, R. Tapia, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, S. Spring, M. Rohde, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2012, **2**, 260–269.
288 M. Wagner and M. Horn, *Curr. Opin. Biotechnol.*, 2006, **17**, 241–249.
289 M. Martinez-Garcia, D. M. Brazel, B. K. Swan, C. Arnosti, P. S. Chain, K. G. Reitenga, G. Xie, N. J. Poulton, M. Lluesma Gomez, D. E. Masland, B. Thompson, W. K. Bellows, K. Ziervogel, C. C. Lo, S. Ahmed, C. D. Gleasner, C. J. Detter and R. Stepanauskas, *PLoS One*, 2012, **7**, e35314.
290 M. W. van Passel, R. Kant, A. Palva, A. Copeland, S. Lucas, A. Lapidus, T. Glavina del Rio, S. Pitluck, E. Goltsman, A. Clum, H. Sun, J. Schmutz, F. W. Larimer, M. L. Land, L. Hauser, N. Kyrpides, N. Mikhailova, P. P. Richardson, P. H. Janssen, W. M. de Vos and H. Smidt, *J. Bacteriol.*, 2010, **193**, 2367–2368.
291 O. Geissinger, D. P. Herlemann, E. Morschel, U. G. Maier and A. Brune, *Appl. Environ. Microbiol.*, 2009, **75**, 2831–2840.
292 D. P. Herlemann, O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz and A. Brune, *Appl. Environ. Microbiol.*, 2009, **75**, 2841–2849.
293 D. Hoffmann, J. M. Hevel, R. E. Moore and B. S. Moore, *Gene*, 2003, **311**, 171–180.
294 T. B. Rounge, T. Rohrlack, A. J. Nederbragt, T. Kristensen and K. S. Jakobsen, *BMC Genomics*, 2009, **10**, 396.
295 A. Tooming-Klunderud, T. Rohrlack, K. Shalchian-Tabrizi, T. Kristensen and K. S. Jakobsen, *Microbiology*, 2007, **153**, 1382–1393.
296 T. Nishizawa, A. Ueda, T. Nakano, A. Nishizawa, T. Miura, M. Asayama, K. Fujii, K. Harada and M. Shirai, *J. Biochem.*, 2011, **149**, 475–485.
297 D. A. Goldthwait, G. R. Greenberg and R. A. Peabody, *J. Biol. Chem.*, 1956, **221**, 569–577.
298 A. Aiba and K. Mizobuchi, *J. Biol. Chem.*, 1989, **264**, 21239–21246.
299 R. Seshadri, L. Adrian, D. E. Fouts, J. A. Eisen, A. M. Phillippy, B. A. Methe, N. L. Ward, W. C. Nelson, R. T. Deboy, H. M. Khouri, J. F. Kolonay, R. J. Dodson, S. C. Daugherty, L. M. Brinkac, S. A. Sullivan, R. Madupu, K. E. Nelson, K. H. Kang, M. Impraim, K. Tran, J. M. Robinson, H. A. Forberger, C. M. Fraser, S. H. Zinder and J. F. Heidelberg, *Science*, 2005, **307**, 105–108.
300 M. Kube, A. Beck, S. H. Zinder, H. Kuhl, R. Reinhardt and L. Adrian, *Nat. Biotechnol.*, 2005, **23**, 1269–1273.
301 P. J. McMurdie, S. F. Behrens, J. A. Muller, J. Goke, K. M. Ritalahti, R. Wagner, E. Goltsman, A. Lapidus, S. Holmes, F. E. Loffler and A. M. Spormann, *PLoS Genet.*, 2009, **5**, e1000714.
302 S. Siddaramappa, J. F. Challacombe, S. F. Delano, L. D. Green, H. Daligault, D. Bruce, C. Detter, R. Tapia, S. Han, L. Goodwin, J. Han, T. Woyke, S. Pitluck, L. Pennacchio, M. Nolan, M. Land, Y. J. Chang, N. C. Kyrpides, G. Ovchinnikova, L. Hauser, A. Lapidus, J. Yan, K. S. Bowman, M. S. da Costa, F. A. Rainey and W. M. Moe, *Stand. Genomic Sci.*, 2012, **6**, 251–264.
303 K. H. Tang, K. Barry, O. Chertkov, E. Dalin, C. S. Han, L. J. Hauser, B. M. Honchak, L. E. Karbach, M. L. Land, A. Lapidus, F. W. Larimer, N. Mikhailova, S. Pitluck, B. K. Pierson and R. E. Blankenship, *BMC Genomics*, 2011, **12**, 334.
304 Y. Takaki, S. Shimamura, S. Nakagawa, Y. Fukuhara, H. Horikawa, A. Ankai, T. Harada, A. Hosoyama, A. Oguchi, S. Fukui, N. Fujita, H. Takami and K. Takai, *DNA Res.*, 2010, **17**, 123–137.
305 S. Pitluck, J. Sikorski, A. Zeytun, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, K. Liolios, I. Pagani, N. Ivanova, K. Mavromatis, A. Pati, A. Chen, K. Palaniappan, L. Hauser, Y. J. Chang, C. D. Jeffries, J. C. Detter, E. Brambilla, O. D. Djao, M. Rohde, S. Spring, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and M. Land, *Stand. Genomic Sci.*, 2011, **4**, 54–62.
306 H. Kiss, E. Lang, A. Lapidus, A. Copeland, M. Nolan, T. Glavina Del Rio, F. Chen, S. Lucas, H. Tice, J. F. Cheng, C. Han, L. Goodwin, S. Pitluck, K. Liolios, A. Pati, N. Ivanova, K. Mavromatis, A. Chen, K. Palaniappan,

Published on 21 December 2012. Downloaded by Pennsylvania State University on 17/09/2016 10:56:17.

M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, J. C. Detter, T. Brettin, S. Spring, M. Rohde, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **2**, 270–279.

307 A. Lapidus, O. Chertkov, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, I. Pagani, N. Ivanova, M. Huntemann, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, E. M. Brambilla, M. Rohde, B. Abt, S. Spring, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and T. Woyke, *Stand. Genomic Sci.*, 2011, **5**, 86–96.

308 I. Rauschenbach, P. Narasingarao and M. M. Haggblom, *Int. J. Syst. Evol. Microbiol.*, 2011, **61**, 654–658.

309 I. Anderson, E. Saunders, A. Lapidus, M. Nolan, S. Lucas, H. Tice, T. G. D. Rio, J.-F. Cheng, C. Han, R. Tapia, L. A. Goodwin, S. Pitluck, K. Liolios, K. Mavromatis, I. Pagani, N. Ivanova, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, C. D. Jeffries, Y.-j. Chang, E.-M. Brambilla, M. Rohde, S. Spring, M. Göker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H.-P. Klenk, *Stand. Genomic Sci.*, 2012, **6**, 155–164.

310 M. Chovatia, J. Sikorski, M. Schroder, A. Lapidus, M. Nolan, H. Tice, T. Glavina Del Rio, A. Copeland, J. F. Cheng, S. Lucas, F. Chen, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, P. Chain, E. Saunders, J. C. Detter, T. Brettin, M. Rohde, M. Goker, S. Spring, J. Bristow, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and J. A. Eisen, *Stand. Genomic Sci.*, 2009, **1**, 254–261.

311 O. Chertkov, J. Sikorski, E. Brambilla, A. Lapidus, A. Copeland, T. Glavina Del Rio, M. Nolan, S. Lucas, H. Tice, J. F. Cheng, C. Han, J. C. Detter, D. Bruce, R. Tapia, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, S. Spring, M. Rohde, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **2**, 280–289.

312 M. Göker, E. Saunders, A. Lapidus, M. Nolan, S. Lucas, N. Hammon, S. Deshpande, J.-F. Cheng, C. Han, R. Tapia, L. A. Goodwin, S. Pitluck, K. Liolios, K. Mavromatis, I. Pagani, N. Ivanova, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, Y.-j. Chang, C. D. Jeffries, E.-M. Brambilla, M. Rohde, S. Spring, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H.-P. Klenk, *Stand. Genomic Sci.*, 2012, **6**, 230–239.

313 M. W. van Passel, R. Kant, E. G. Zoetendal, C. M. Plugge, M. Derrien, S. A. Malfatti, P. S. Chain, T. Woyke, A. Palva, W. M. de Vos and H. Smidt, *PLoS One*, 2011, **6**, e16876.

314 N. Ivanova, S. Gronow, A. Lapidus, A. Copeland, T. Glavina Del Rio, M. Nolan, S. Lucas, F. Chen, H. Tice, J. F. Cheng, E. Saunders, D. Bruce, L. Goodwin, T. Brettin, J. C. Detter, C. Han, S. Pitluck, N. Mikhailova, A. Pati, K. Mavrommatis, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, P. Chain, C. Rohde, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2009, **1**, 126–132.

315 M. Harmon-Smith, L. Celia, O. Chertkov, A. Lapidus, A. Copeland, T. Glavina Del Rio, M. Nolan, S. Lucas, H. Tice, J. F. Cheng, C. Han, J. C. Detter, D. Bruce, L. Goodwin, S. Pitluck, A. Pati, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, T. Brettin, M. Goker, B. Beck, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and F. Chen, *Stand. Genomic Sci.*, 2010, **2**, 220–227.

316 J. Sikorski, O. Chertkov, A. Lapidus, M. Nolan, S. Lucas, T. G. Del Rio, H. Tice, J. F. Cheng, R. Tapia, C. Han, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, E. Brambilla, M. Yasawong, M. Rohde, R. Pukall, S. Spring, M. Goker, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides and H. P. Klenk, *Stand. Genomic Sci.*, 2010, **3**, 304–314.

317 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.

318 G. Suen, P. J. Weimer, D. M. Stevenson, F. O. Aylward, J. Boyum, J. Deneke, C. Drinkwater, N. N. Ivanova, N. Mikhailova, O. Chertkov, L. A. Goodwin, C. R. Currie, D. Mead and P. J. Brumm, *PLoS One*, 2011, **6**, e18814.

319 K. Gounder, E. Brzuszkiewicz, H. Liesegang, A. Wollherr, R. Daniel, G. Gottschalk, O. Reva, B. Kumwenda, M. Srivastava, C. Bricio, J. Berenguer, E. van Heerden and D. Litthauer, *BMC Genomics*, 2011, **12**, 577.

320 E. Pelletier, A. Kreimeyer, S. Bocs, Z. Rouy, G. Gyapay, R. Chouari, D. Riviere, A. Ganesan, P. Daegelen, A. Sghir, G. N. Cohen, C. Medigue, J. Weissenbach and D. Le Paslier, *J. Bacteriol.*, 2008, **190**, 2572–2579.

321 K. F. Ettwig, M. K. Butler, D. Le Paslier, E. Pelletier, S. Mangenot, M. M. Kuypers, F. Schreiber, B. E. Dutilh, J. Zedelius, D. de Beer, J. Gloerich, H. J. Wessels, T. van Alen, F. Luesken, M. L. Wu, K. T. van de Pas-Schoonen, H. J. Op den Camp, E. M. Janssen-Megens, K. J. Francoijs, H. Stunnenberg, J. Weissenbach, M. S. Jetten and M. Strous, *Nature*, 2010, **464**, 543–548.

322 P. J. Huggan and D. R. Murdoch, *J. Infect.*, 2008, **57**, 283–289.

323 E. Ransom-Jones, D. L. Jones, A. J. McCarthy and J. E. McDonald, *Microb. Ecol.*, 2012, **63**, 267–281.

This journal is © The Royal Society of Chemistry 2013