Title

Natural Products and the Gene Cluster Revolution

Journal

Trends in Microbiology, 24(12)

ISSN

0966-842X

Author

Jensen, Paul R

Publication Date

2016-12-01

DOI

10.1016/j.tim.2016.07.006

Peer reviewed

Trends in Microbiology

CellPress

Review

Natural Products and the Gene Cluster Revolution

Paul R. Jensen[1,*]

Genome sequencing has created unprecedented opportunities for natural-product discovery and new insight into the diversity and distributions of natural-product biosynthetic gene clusters (BGCs). These gene collectives are highly evolved for horizontal exchange, thus providing immediate opportunities to test the effects of small molecules on fitness. The marine actinomycete genus *Salinispora* maintains extraordinary levels of BGC diversity and has become a useful model for studies of secondary metabolism. Most *Salinispora* BGCs are observed infrequently, resulting in high population-level diversity while conforming to constraints associated with maximum genome size. Comparative genomics is providing a mechanism to assess secondary metabolism in the context of evolution and evidence that some products represent ecotype-defining traits while others appear selectively neutral.

Not So Secondary Metabolism

Natural products, also called secondary or specialized metabolites, are small organic molecules produced by plants, microbes, and invertebrates. They generally confer fitness advantages and are not considered essential for growth or reproduction. Despite their increasingly unpopular 'secondary' moniker, it is clear that these compounds mediate important interactions among organisms and the environment [1]. Natural products are assembled in myriad ways from building blocks generally borrowed from primary metabolism (Box 1). The architectures of the enzymes responsible for producing many of the most commercially important natural products reveal a building-block-type assembly that provides clear opportunities for biosynthetic engineering. Efforts to exploit these systems have been described as 'Lego-ization' [2] and demonstrate a path to produce complex polyketides using *de novo* design and assembly of novel multimodule enzymes [3]. Bioinformatically, many natural-product **biosynthetic gene clusters** (see Glossary) are highly informative of the class of compounds they encode and, in some cases, the precise structures. This has facilitated the development of important computational tools such as antiSMASH [4,5], which can be used to predict natural-product potential from DNA sequence data [6]. While certain BGC types are well designed to accommodate genetic modifications that result in structural changes to the small molecules they encode [7], there remain many challenges to the Lego concept [8]. Advances in our understanding of the molecular genetics of natural-product biosynthesis are driving the resurgence of this field [9], yet it remains clear that we are far from understanding and harnessing the many diverse mechanisms microbes employ to generate natural-product diversity.

Bacterial natural products encompass a dizzying array of complex chemical structures. They were once the primary source of chemical diversity screened by the pharmaceutical industry; however, competing technologies led to the implementation of alternative discovery platforms [10]. Despite falling out of favor, natural products or their derivatives continue to account for a large percentage of drugs used in the clinic today [11]. While they have proven their worth as biologically active chemical entities, the ecological functions of most of these compounds remain unknown.

Trends

Genome sequencing is providing unprecedented opportunities to explore the diversity and distributions of natural-product biosynthetic gene clusters (BGCs) among bacteria.

Genomic surveys reveal extensive BGC diversity among closely related strains, suggesting a strategy to maximize the population-level secondary metabolome while minimizing the number of gene clusters carried by any one strain.

Studies of BGC evolutionary history in relation to that of the bacteria in which they reside are providing unique opportunities to explore the mechanisms by which chemical diversity is created in Nature.

Acquiring a BGC via horizontal gene transfer (HGT) provides an immediate opportunity to test the effects of the encoded small molecules on fitness.

[1]Center for Marine Biotechnology and Biomedicine, Center for Microbiome Innovation, Scripps Institution of Oceanography, University of California San Diego, San Diego, California, USA

*Correspondence: pjensen@ucsd.edu (P.R. Jensen).

Box 1. Natural-Product Biosynthetic Logic

Great strides have been made in our understanding of the molecular genetics of natural-product biosynthesis. A number of excellent reviews describe these advances for a variety of biosynthetic paradigms [71–75]. Canonical examples include type I modular polyketide synthases (PKSs). These large multifunctional enzymes act successively in an assembly-line fashion, with individual modules responsible for the addition of acyl units onto a growing polyketide chain [73]. The remarkable structural diversity observed among type I PKS-derived polyketides arises largely from the combinatorial use of these modules, the variable use of three reductive domains, and a few simple building blocks such as acetate [72]. Nonribosomal peptide synthetases function in a similar assembly-line fashion [76], except with amino acids as the extender units. These types of linear system are relatively amenable to bioinformatic interpretation compared to iteratively acting enzymes such as type II PKSs, which can not simply be read module by module [77]. Despite major advances in this field, there seem to be exceptions to every biosynthetic rule established [78], and new types of biochemistries are continually being discovered [79], making it clear that we remain far from understanding the many diverse mechanisms of natural-product biosynthesis.

Glossary

Biosynthetic gene cluster (BGC): a group of colocalized genes that encode the production of a natural product or group of related natural products.

Horizontal gene transfer (HGT): a collection of processes used by many microorganisms to transfer genes horizontally.

Minimum Information about a Biosynthetic Gene cluster (MIBiG): a standardized framework for the annotation of BGCs and their associated metadata.

The genes that encode the production of bacterial natural products are generally clustered into contiguous stretches of DNA [12,13] and, at least among the actinomycetes, tend to be located in variable regions of the chromosome known as genomic islands [14,15]. Natural-product BGCs frequently include not only the core biosynthetic and tailoring enzymes but also genes associated with the regulation of biosynthesis and resistance to the small-molecule products they encode [16,17]. These genetic packages are highly evolved for horizontal gene transfer (HGT), as evidenced both by their clustering, frequent linkage with mobile genetic elements, and detection on plasmids [18]. BGCs resemble 'selfish operons' [19], with exchange events being favored among closely related strains due to mechanism of DNA mismatch repair and maintenance. Less common exchanges among more disparate taxa increase the host range of a cluster, which can subsequently become host adapted over time via modifications in GC content and codon usage. Thus, BGCs can be studied in the context of mobile genetic elements, and as such considered independent evolutionary entities relative to the hosts in which they occur [20]. Once acquired, they provide immediate opportunities for a new host to test the effects of the small molecules they encode on fitness (Figure 1, Key Figure). The frequency and extent of these exchange events remain unknown but provide a unique perspective on gene flow among organisms that undergo clonal reproduction and opportunities to assess acquired functional traits as the drivers of bacterial diversification.

Genomic Insights

The genome sequence of the model organism *Streptomyces coelicolor* strain A3(2) provided a remarkable revelation. Despite more than 50 years of study [21], it contained 18 natural-product BGCs for which the products had yet to be discovered [14]. This genetic potential spawned the concept of natural-product genome mining [22,23], which takes a genome-first approach to natural product discovery. Genome sequencing played a key role in the revitalization of natural products research [24], which was largely abandoned by the pharmaceutical industry, and created a rush to incorporate 'omic' sciences into discovery pipelines [25,26]. With increasing access to genome sequence data have come opportunities to ask broader questions about secondary metabolism, such as a recent global analysis of 1154 diverse bacterial genomes [27]. This study identified over 33 000 putative BGCs, the vast majority of which are uncharacterized. This remarkable level of diversity clearly attests to the importance of natural-product biosynthesis as a defining feature of bacterial metabolism.

Genomics has taught us that bacteria harbor large numbers of orphan BGCs. These are defined as BGCs that have not been linked to the natural products they encode. While genome mining is a new technology, natural-product chemists have isolated and characterized tens of thousands of compounds, the vast majority of which have not been linked to their respective BGCs (Box 2). Thus, we do not know how many orphan clusters encode the production of new compounds versus known compounds whose biosynthetic machinery has yet to be identified. Given that the

Key Figure

Ecological and Evolutionary Implications of Secondary Metabolism

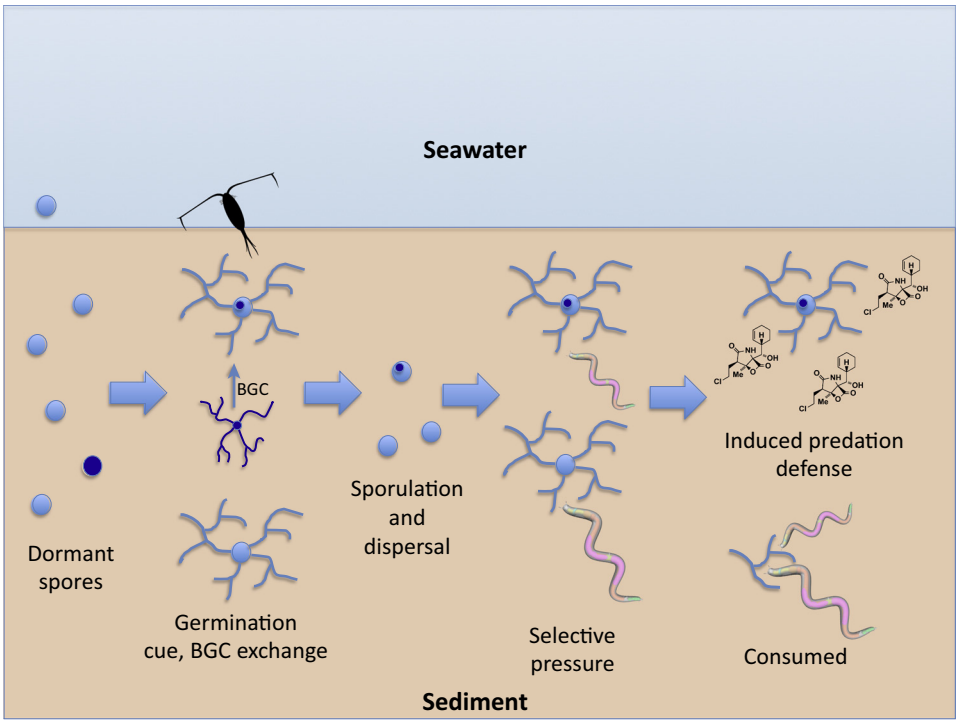

Figure 1. In this scenario, dormant spores are distributed via ocean currents and deposited in sediments. Growth is triggered by germination cues provided by a senescent zooplankton bloom. Actively growing members of the sediment community exchange biosynthetic gene clusters (BGCs) via horizontal gene transfer. These populations subsequently sporulate, and the spores are distributed to new locations where they undergo another round of growth. The arrival of predatory nematodes provides selective pressures that induce secondary metabolite production. Only the lineage that acquired an effective feeding deterrent survives. It's unknown if the spatial distributions of actively growing strains are such that social cheating can develop. If sufficiently advantageous, this lineage may outcompete conspecifics, resulting in a selective sweep that purges diversity from the remainder of the population experiencing the selective pressure. Alternatively, if the acquired trait provides access to a new niche, this population may become ecologically isolated, initiating a sympatric speciation event.

Box 2. Linking Compounds to Their Gene Clusters

It is now possible to computationally identify thousands of BGCs from sequence data. This information can be used to make predictions about the natural-product biosynthetic potential of a sample. These predictive capacities continue to improve as more BGCs become experimentally linked to the products they encode. This is generally accomplished by time-consuming genetic manipulations such as knocking out a key gene in the pathway and observing the loss of compound production [22] or heterologous expression [80]. Once experimental links are established, the detection of a similar BGC in other strains can be used to predict the production of similar compounds. The MIBiG specification [81] provides a robust community standard for annotations and metadata on biosynthetic gene clusters and their molecular products (http://mibig.secondarymetabolites.org/index.html). This useful resource facilitates standardized BGC deposition and retrieval and the development of new analytical tools. While experimental verification and *de novo* structure elucidation remain bottlenecks, new approaches in synthetic biology aim to speed the throughput with which the products of orphan BGCs can be identified [82].

products encoded by many BGCs remain difficult to predict using bioinformatic approaches, BGC novelty must still be addressed on a case-by-case basis using labor-intensive and often challenging laboratory experiments to interrogate gene function. Surprisingly, we also know very little about how many BGCs are silent, that is, are not expressed when strains are cultivated in the laboratory, and about the cues that trigger expression in Nature. The temporal relationships between gene expression, translation into functional protein, and small-molecule assembly also remain poorly understood. Finally, limitations in the analytical techniques used to detect, isolate, and identify natural products likely account for some of the disparity between biosynthetic potential and the compounds discovered to date. While the jury remains out in terms of the rewards that may come from genome mining, this avenue of research nonetheless holds great promise and is driving important developments in synthetic biology [28] and in computational biology [29,30].

As might be expected, bacteria that are well endowed for natural-product biosynthesis tend to have large genomes and complex life cycles. These features are exemplified by the actinomycetes [31] and the myxobacteria [32]. Actinomycetes (defined here as bacteria in the Order Actinomycetales) are slow-growing, high G+C Gram-positive bacteria of which many include a spore stage in their life cycle. They are well known from soils where they produce exoenzymes that hydrolyze recalcitrant polymers such as lignocellulose and chitin [33]. Actinomycetes are also consistently isolated from marine sources, including sediments [34,35]. In collaboration with colleagues, my laboratory has been studying secondary metabolism in the marine actinomycete genus *Salinispora* [36]. An accounting of two major classes of BGCs among 75 *Salinispora* strains revealed extensive BGC diversity, with more than 50% of the gene clusters occurring in only one or two strains [37]. While it was known that BGCs are exchanged by HGT [38], these results led to the surprising observation that most of the BGCs detected had been acquired relatively recently in the evolutionary history of the genus. These recent acquisitions were interpreted as sampling events initiated at the strain level and presumably dictated by the availability of BGCs in the local gene pool. Following this logic, the complement of BGCs maintained by bacteria at a specific location would be influenced by the genetic diversity available for sampling. The detection of this type of biogeographic signal would require high rates of gene exchange relative to cell dispersal, two variables that remain difficult to quantify. Nonetheless, these concepts provide opportunities to address the biogeography of secondary metabolism, both in terms of the organisms that participate in this process [39], the environments in which they reside [40], and the effects of community composition on genome evolution.

Evolutionary Considerations

Evolution by natural selection is often considered in the context of random nucleotide change and selection. The acquisition of a BGC by HGT shares similarities with a nucleotide substitution, except that an acquisition event provides the opportunity to test the fitness effects of a small molecule encoded by a large and complex gene cluster as opposed to an amino acid change in a protein. Some acquired BGCs may not produce their intended products, that is, they may be dysfunctional in the new host, or their products may not provide a selective advantage or disadvantage, as is the case for most mutational changes in a nucleotide sequence. If the neutral theory of molecular evolution were applied to BGCs, it would suggest that most acquisition events are selectively neutral or slightly deleterious [41], the former of which has been proposed for most genes acquired by HGT [42]. This concept has been invoked for a majority of the rare, flexible genome in *Prochlorococcus* populations [43]. One explanation for the observation that most *Salinispora* BGCs were observed in only one or two strains is that the products they encode do not provide a strong selective advantage. This is in contrast to a limited number of BGCs that have been consistently observed among globally distributed populations [44]. The 'fixation' of these BGCs provides compelling evidence for strong positive selection maintaining the functions of the encoded products, which include potent antibiotics and cytotoxins.

However, one difference between a point mutation and an acquired but selectively neutral BGC is that the latter includes the cost of carrying additional DNA. Among *Salinispora* spp., there is evidence that similar pathway types can replace each other at the same chromosomal location [37] thus mitigating the cost of some acquisition events. This bears similarities to previous observations of "replacement" genomic islands, where different sets of genes occupy similar positions in otherwise clonal lineages [45]. Another distinguishing feature of the BGCs acquired by *Salinispora* spp. is their recruitment to genomic islands [15] where they are less likely to disrupt essential cellular functions and more likely to be available for subsequent loss events. We currently have little to no perspective on the rates at which BGCs are gained, lost, or replaced in bacterial genomes, the ratios of functional to non-functional BGCs, and the BGCs that are maintained due to selective processes. Increased access to genome sequences from large numbers of closely related strains will help resolve these questions. As with the maintenance of random nucleotide changes among individuals, BGC acquisition events provide a mechanism to maximize population-level diversity while limiting the number of BGCs maintained within individual genomes.

Salinispora spp. average 14–18 polyketide and nonribosomal peptide BGCs per strain, while a total of 124 distinct BGCs were detected among 75 different strains with even more predicted with continued sampling [37]. This potentially vast array of molecules produced at the population level increases the likelihood of an effective, secondary-metabolite-mediated response, by at least a few individuals, to new selective pressures and thus provides an ecological rationale for the extensive BGC diversity observed among *Salinispora* spp. [37]. This interpretation shares parallels with the concept that potent biomolecular activity is a rare trait that requires strains to maintain large natural product inventories [46]. However in this case, the inventories are maintained within an ecologically cohesive group, or species, as opposed to an individual genome, and the compounds may be rarely needed as opposed to rarely active. This type of plasticity would be most effective when selective pressures are ephemeral and varied, as might be expected when spores become distributed over large geographic areas. Diversifying the population-level secondary metabolome also provides an effective strategy to avoid the development of resistance, as in the application of combinatorial drug therapy, and the subsequent need to enter into a coevolutionary arms race [47,48]. Following this logic, populations that evolve the most effective sampling strategies would gain a selective advantage. While HGT comes at a cost [20], the extensive BGC diversity observed among *Salinispora* strains suggests that the benefits of testing new traits outweigh the costs of sampling from the community gene pool. This concept is clearly not limited to secondary metabolism and has been proposed as a more general mechanism for bacteria to survive and effectively exploit their environment [45].

In addition to acquiring new traits, the acquisition of a BGC also provides the opportunity to improve upon an existing trait. If an acquisition event yields a product that does a better job at a specific function than an existing compound, that strain can outcompete other members of the population, and the associated genome can theoretically sweep through the population, purging the diversity observed outside of this more competitive individual, in a process known as periodic selection [49]. Fred Cohan at Wesleyan University has championed the selective sweep concept, and a recent time-series metagenomic study provides the first support for genome-wide selective sweeps among environmental bacteria [50]. We see evidence for selective sweeps in *Salinispora arenicola*, which maintains more fixed BGCs and considerably less diversity than *Salinispora pacifica* despite being the more deeply rooted and broadly distributed lineage [51]. There is also evidence in *Vibrio* spp. that sweeps can be localized to specific genes as opposed to bacterial genomes [52], thus suggesting that different models of evolution may apply to different groups of bacteria. Alternatively, if the products of an acquired BGC introduce a new functional trait that allows the producer to expand into a new niche, the resulting populations can become ecologically distinct and free to diverge independently of the ancestral population [49].

Linking the acquisition of BGCs to diversification events supports the concept that secondary metabolites represent key ecotype-defining traits that function as drivers of bacterial diversification [15].

How Many BGCs Are There?

At any one point in time, the biosphere holds a defined number of BGCs. We do not know this number or the rate at which new clusters are being created or lost, although it has been proposed that they are among life's most diverse and rapidly evolving genetic elements [53]. We know that mutation, recombination, gene gain, loss and duplication, and the successive merger of smaller subclusters all contribute to the evolution of BGC diversity [7,8,53]. Evolutionary mechanisms for the creation and maintenance of natural-product chemical diversity have also been presented [54]. While laboratories around the globe employ techniques in synthetic biology to alter existing BGCs as an approach to generate new chemical diversity [55,56], environmental bacteria continue to perform similar experiments in a more natural setting and at an unknown pace, guided by changing selective pressures and opportunities for genetic exchange. Bacteria have clearly been successful in these efforts, as evidenced by the extraordinary diversity of compounds discovered to date. However, the rates at which new chemical diversity is created in Nature may not be sufficient to meet the needs of traditional natural-product discovery efforts once the current inventory of compounds has been exhausted. While advances in synthetic biology may provide some relief to this problem [56], much can also be gained from learning more about the mechanisms bacteria employ to generate new chemical diversity. Exploiting these mechanisms, and developing methods to enhance the rates at which they occur, present tantalizing opportunities to incorporate experimental evolution into natural-product discovery efforts.

While we know little about how frequently bacteria engineer the biosynthesis of new natural products, there are many examples in which changes in BGCs have been linked to changes in the small molecules they encode [53,57]. These processes are most easily observed when comparing sets of related BGCs with the secondary metabolites they encode. The *sal* cluster [58], which encodes the production of compounds in the salinosporamide series [59,60], provides an interesting example. The version of this BGC observed in *Salinispora tropica* yields salinosporamide A and related compounds, while the version observed in *S. pacifica* yields salinosporamide K [61], which lacks the chloroethyl side chain. It could be inferred that the ancestral pathway was acquired prior to the *S. tropica–S. pacific* diversification event and subsequently evolved independently in these two lineages [62]. Salinosporamide A is an extremely potent inhibitor of 20S proteasome function and has been observed in all *S. tropica* strains studied to date, suggesting that the production of this compound is under strong positive selection and fundamentally important to this species. The *S. pacifica* version of the *sal* BGC is missing the genes required for the biosynthesis of the chloroethyl side chain in salinosporamide A and yields a product that is considerably less active [61]. Correspondingly, the *S. pacifica* version of the *sal* BGC is sporadically distributed within this species, suggesting relatively weak selection for its maintenance. These observations provide a rare glimpse into the evolution of structural diversity and how biological activity may drive the distribution of individual BGCs. The ecological functions of these compounds remain to be determined.

Public Goods, Social Cheaters, and Spore Formers

Once secreted, microbial natural products become public goods and as such can provide benefit to individuals other than the direct producer. These types of cooperative behavior are challenging to interpret from an evolutionary perspective as they create opportunities for social cheating, where nonproducers benefit from compounds without paying the cost of production [63]. Natural-product social cheating has been observed among bacteria that produce quorum-sensing compounds [64] and iron-chelating siderophores [65]. One evolutionary theory used to

explain natural products as public goods despite the potential for cheating is kin selection [66], where traits are favored because of their beneficial effects on the fitness of relatives. Kin selection requires a high degree of genetic relatedness and has been suggested to work best when there is limited dispersal, thus increasing the chances that neighbors will be close relatives [63]. Kin selection is facilitated by kin discrimination [67], which, in the case of natural products, would include compounds whose functions benefit only closely related strains. This could be realized by a siderophore that can be taken up only by close relatives or an antibiotic to which only close relatives are resistant. This later concept could help account for the observation that environmental bacteria are organized into socially cohesive units in which antagonism occurs between, rather than within, ecologically defined populations [68]. There is evidence that genes associated with natural-product biosynthesis are under diversifying selection [69], which may facilitate kin discrimination by leading to the production of diverse and highly specific compounds. This may help explain why some bacteria maintain so many BGCs and why many BGCs yield suites of related compounds.

In the case of spore-forming actinomycetes, a local population arising from an individual spore will have high levels of genetic relatedness, thus facilitating cooperative behaviors such the production of public goods. However, it becomes difficult to reconcile the development of social cheating in the context of spore-forming bacteria without understanding spore distributions and the spatiotemporal dynamics of mycelium development following germination and growth, variables for which little information is currently available. In terms of time scales, germination events may be insufficient for cheaters to arise within populations that originate from a single or group of clonal spores. If cheaters were to develop, the spores they subsequently produce would be at a disadvantage if the selective pressures addressed by the public goods remain in place yet germination occurs in locations where the public goods are not available. Thus it becomes critical to understand if germination events are spatially localized or result in

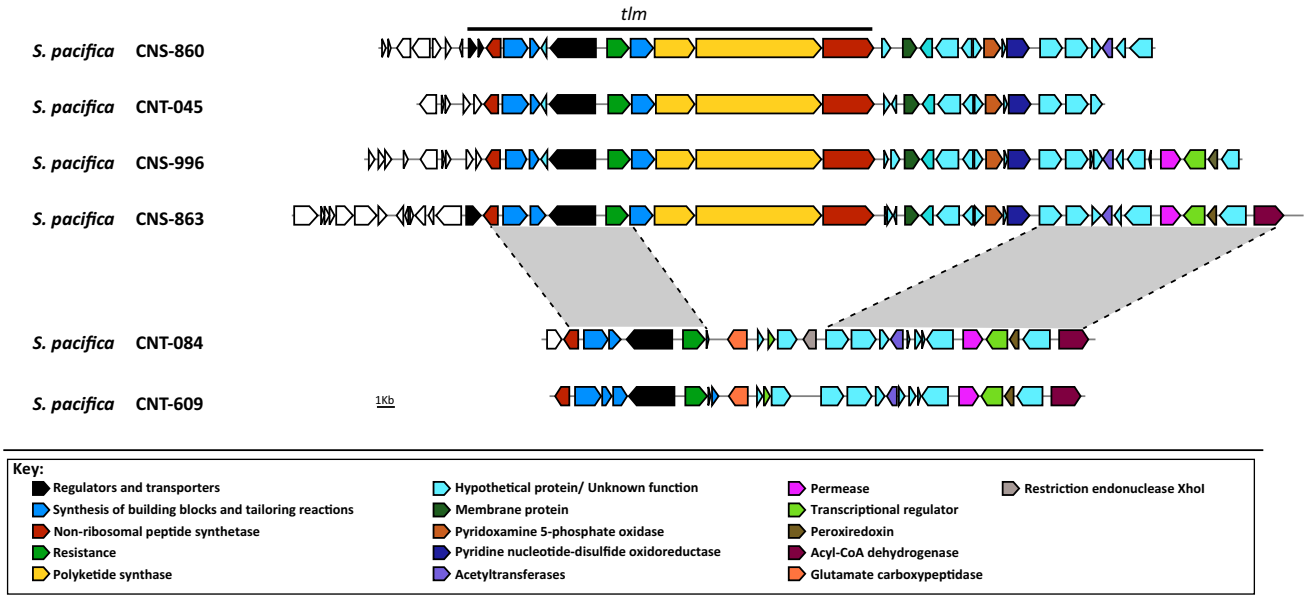

Figure 2. Evidence of Social Cheating in *Salinispora pacifica*. The biosynthetic gene cluster (BGC) encoding the production of the antibiotic thiolactomycin has been identified (*tlm*). This BGC was detected in four strains of *S. pacifica*. Two additional strains contain portions of the pathway but lack key polyketide synthase (red) and nonribosomal peptide synthetase (yellow) genes required for compound production. These strains retain the *fabB/H* gene (green) that confers resistance to the antibiotic effects of this compound. Reprinted with permission from [70], copyright 2015 American Chemical Society.

interactions among diverse populations. Without these interactions, it would be difficult for cheater populations to be maintained through cycles of spore production, dispersal, and growth.

There is some evidence for social cheating among *Salinispora* strains. This was observed in the *tlm* pathway, which encodes the production of a rare class of thiotetronic acid antibiotics [70]. These compounds inhibit bacterial fatty-acid synthesis, and the associated BGC was discovered using a target-directed genome-mining approach in which candidate resistance genes were identified within orphan BGCs as an *a priori* method to predict the molecular targets of the small molecules they encode. The *tlm* BGC contained what appeared to be a duplicated copy of a gene involved in fatty-acid synthesis (*fabB/H*), suggesting that it may be associated with resistance. The product of the BGC was subsequently identified as thiolactomycin and the resistance-conferring ability of the *fabB/H* paralog confirmed [70]. While the *tlm* BGC was only observed in 4 of 86 *Salinispora* strains, all of these originated from independent marine sediment samples collected from Fiji. In addition to these four strains, two additional strains, also isolated from Fiji, maintained the resistance gene but had key biosynthetic genes missing from the BGC (Figure 2). These two strains are not capable of producing thiolactomycin yet they maintain the genetic basis for resistance, making them strong candidates for social cheaters. An exhaustive, bioinformatic analysis of natural-product social cheating remains to be performed on the available *Salinispora* genome sequences.

Concluding Remarks

Omic sciences are driving natural-products research in exciting new directions. Genome sequence data, coupled with an improved understanding of the molecular genetics of natural-product biosynthesis, and online tools to facilitate computational analyses, are providing unprecedented opportunities to explore the diversity and distributions of natural-product biosynthetic gene clusters in Nature. While most of the BGCs observed in bacterial genome sequences are orphan, it remains unclear how many of these will yield new compounds (see Outstanding Questions). Synthetic biology has taken center stage in terms of discovering the products of orphan BGCs; however, we remain surprisingly unaware of the natural cues that trigger expression in the native hosts. A better understanding of the regulatory cues and ecological functions of natural products will certainly facilitate future discovery efforts. By sequencing large numbers of closely related strains, it has become clear that certain groups of bacteria maintain extensive population-level BGC diversity and that most of this diversity is rarely observed among strains. This observation suggests that most BGCs are not under strong positive selection and that maintaining an expansive and dynamic secondary metabolome maximizes opportunities for population-level responses to episodic environmental pressures. While much remains to be learned about the mechanisms by which chemical diversity is created, natural-product BGCs provide opportunities to explore secondary metabolism as it relates to the ecology and evolution of bacteria in addition to exploiting their products for useful purposes.

Outstanding Questions

Do most orphan BGCs encode new compounds or compounds that are already known but have yet to be linked to their biosynthetic machinery?

Are most orphan BGCs silent or do we simply miss the associated compounds due to limitations associated with extraction and analytical techniques?

Are we missing environmental or regulatory signals needed for BGC expression when bacteria are grown in the laboratory?

At what rate are BGCs created and lost?

How often do new chemical scaffolds evolve?

Can experimental evolution be used to exploit the intrinsic capacities of bacteria to create new chemical diversity?

Are natural products functional traits that define bacterial ecotypes?

Can natural-product BGC acquisition events drive speciation?

Acknowledgments

The author acknowledges support from the National Institutes of Health (2U19TW007401, 5R01GM085770) and the National Science Foundation (OCE-1235142). Special thanks to Ricardo Pereira and one anonymous reviewer for their constructive comments on the manuscript. Genome sequencing was conducted by the U.S. Department of Energy Joint Genome Institute and supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

1. Wietz, M. *et al.* (2013) Antagonistic interactions mediated by marine bacteria: The role of small molecules. *J. Chem. Ecol.* 39, 879–891
2. Sherman, D.H. (2005) The Lego-ization of polyketide biosynthesis. *Nat. Biotechnol.* 23, 1083–1084
3. Menzella, H.G. *et al.* (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* 23, 1171–1176
4. Medema, M.H. *et al.* (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene

clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, 339–346

5. Weber, T. *et al.* (2015) antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, 237–243
6. Weber, T. (2014) *In silico* tools for the analysis of antibiotic biosynthetic pathways. *Int. J. Med. Microbiol.* 304, 230–235
7. Jenke-Kodama, H. and Dittmann, E. (2009) Evolution of metabolic diversity: insights from microbial polyketide synthases. *Phytochemistry* 70, 1858–1866
8. Medema, M.H. *et al.* (2014) A systematic computational analysis of biosynthetic gene cluster evolution: Lessons for engineering biosynthesis. *PLoS Comp. Biol.* 10, e1004016
9. Walsh, C.T. and Fischbach, M.A. (2010) Natural products version 2.0: connecting genes to molecules. *J. Am. Chem. Soc.* 132, 2469–2493
10. Li, J.W. and Vederas, J.C. (2009) Drug discovery and natural products: end of an era or an endless frontier? *Science* 325, 161–165
11. Newman, D.J. and Cragg, G.M. (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661
12. Stone, M. and Williams, D. (1992) On the evolution of functional secondary metabolites (natural products). *Mol. Microbiol.* 6, 29–34
13. Osbourn, A. (2010) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* 26, 449–457
14. Bentley, S.D. *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147
15. Penn, K. *et al.* (2009) Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J.* 3, 1193–1203
16. Bibb, M.J. (2005) Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* 8, 208–215
17. Nett, M. *et al.* (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* 26, 1362–1384
18. Kinashi, H. *et al.* (1987) Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. *Nature* 328, 454–456
19. Lawrence, J.G. and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143, 1843–1860
20. Rankin, D.J. *et al.* (2011) What traits are carried on mobile genetic elements, and why? *Heredity* 106, 1–10
21. Hopwood, D.A. (1999) Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. *Microbiology* 145, 2183–2202
22. Zerikly, M. and Challis, G.L. (2009) Strategies for the discovery of new natural products by genome mining. *ChemBioChem* 10, 625–633
23. Ziemert, N. *et al.* (2016) The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* Published online June 8, 2016. http://dx.doi.org/10.1039/C6NP00025H
24. Bode, H.B. and Müller, R. (2005) The impact of bacterial genomics on natural product research. *Angewandte Chemie Inter. Ed.* 44, 6828–6846
25. Kersten, R.D. *et al.* (2011) A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* 7, 794–802
26. Kersten, R.D. *et al.* (2013) Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4407–4416
27. Cimermancic, P. *et al.* (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158, 412–421
28. Hertweck, C. (2015) Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology. *Trends Biochem. Sci.* 40, 189–199
29. Medema, M.H. *et al.* (2012) Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol.* 10, 191–202
30. Bachmann, B.O. (2005) Decoding chemical structures from genomes. *Nat. Chem. Biol.* 1, 244–245
31. Ensign, J.C. (1978) Formation, properties, and germination of actinomycete spores. *Annu. Rev. Microbiol.* 32, 185–219
32. Reichenbach, H. *et al.* (1988) Myxobacteria: a source of new antibiotics. *Trends Biotechnol.* 6, 115–121
33. McCarthy, A.J. and Williams, S.T. (1992) Actinomycetes as agents of biodegradation in the environment – a review. *Gene* 115, 189–192
34. Maldonado, L.A. *et al.* (2005) Diversity of cultivable actinobacteria in geographically widespread marine sediments. *Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbiol.* 87, 11–18
35. Jensen, P.R. *et al.* (2005) Culturable marine actinomycete diversity from tropical Pacific Ocean sediments. *Environ. Microbiol.* 7, 1039–1048
36. Jensen, P.R. *et al.* (2015) The marine actinomycete genus *Salinispora*: a model organism for secondary metabolite discovery. *Nat. Prod. Rep.* 32, 738–751
37. Ziemert, N. *et al.* (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1130–E1139
38. Ginolhac, A. *et al.* (2005) Type I polyketide synthases may have evolved through horizontal gene transfer. *J. Mol. Evol.* 60, 716–725
39. Edlund, A. *et al.* (2011) Geographic distribution of secondary metabolite genes in the marine actinomycete *Salinispora arenicola*. *Appl. Environ. Microbiol.* 77, 5916–5925
40. Charlop-Powers, Z. *et al.* (2014) Chemical-biogeographic survey of secondary metabolism in soil. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3757–3762
41. Kimura, M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267, 275–276
42. Gogarten, J.P. and Townsend, J.P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687
43. Coleman, M.L. and Chisholm, S.W. (2010) Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18634–18639
44. Jensen, P.R. *et al.* (2007) Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl. Environ. Microbiol.* 73, 1146–1152
45. Rodriguez-Valera, F. *et al.* (2016) Flexible genomic islands as drivers of genome evolution. *Curr. Opin. Microbiol.* 31, 154–160
46. Firn, R.D. and Jones, C.G. (2000) The evolution of secondary metabolism – a unifying model. *Mol. Microbiol.* 37, 989–994
47. Patin, N.V. *et al.* (2015) Competitive strategies differentiate closely related species of marine actinobacteria. *ISME J.* 10, 478–490
48. Kinkel, L.L. *et al.* (2013) Sympatric inhibition and niche differentiation suggest alternative coevolutionary trajectories among streptomycetes. *ISME J.* 8, 249–256
49. Cohan, F. (2002) What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487
50. Bendall, M.L. *et al.* (2016) Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.* 10, 1589–1601
51. Freel, K.C. *et al.* (2013) Multilocus sequence typing reveals evidence of homologous recombination linked to antibiotic resistance in the genus *Salinispora*. *Appl. Environ. Microbiol.* 79, 5997–6005
52. Shapiro, B.J. *et al.* (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51
53. Fischbach, M.A. *et al.* (2008) The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4601–4608
54. Firn, R.D. and Jones, C.G. (2003) Natural products – a simple model to explain chemical diversity. *Nat. Prod. Rep.* 20, 382–391
55. Wilkinson, B. and Micklefield, J. (2007) Mining and engineering natural-product biosynthetic pathways. *Nat. Chem. Biol.* 3, 379–386
56. Kim, E. *et al.* (2015) Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* 11, 649–659
57. Ueoka, R. *et al.* (2015) Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat. Chem. Biol.* 11, 705–712

58. Eustáquio, A.S. *et al.* (2009) Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from *S*-adenosyl- L-methionine. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12295–12300

59. Feling, R.H. *et al.* (2003) Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew Chem. Int. Ed.* 42, 355–357

60. Williams, P.G. *et al.* (2005) New cytotoxic salinosporamides from the marine actinomycete *Salinispora tropica*. *J. Org. Chem.* 70, 6196–6203

61. Eustáquio, A.S. *et al.* (2011) The discovery of salinosporamide K from the marine bacterium "*Salinispora pacifica*" by genome mining gives insight into pathway evolution. *ChemBioChem* 12, 61–64

62. Freel, K.C. *et al.* (2011) Evolution of secondary metabolite genes in three closely related marine actinomycete species. *Appl. Environ. Microbiol.* 77, 7261–7270

63. West, S.A. *et al.* (2006) Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* 4, 597–607

64. Sandoz, K.M. *et al.* (2007) Social cheating in *Pseudomonas aeruginosa* quorum sensing. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15876–15881

65. Harrison, F. *et al.* (2008) Interspecific competition and siderophore-mediated cooperation in *Pseudomonas aeruginosa*. *ISME J.* 2, 49–55

66. Diggle, S.P. *et al.* (2007) Cooperation and conflict in quorum-sensing bacterial populations. *Nature* 450, 411–414

67. Strassmann, J.E. *et al.* (2011) Kin discrimination and cooperation in microbes. *Annu. Rev. Microbiol.* 65, 349–367

68. Cordero, O.X. *et al.* (2012) Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* 337, 1228–1231

69. Smith, E.E. *et al.* (2005) Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. *J. Bacteriol.* 187, 2138–2147

70. Tang, X. *et al.* (2015) Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* 10, 2841–2849

71. Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* 7, 285–295

72. Fischbach, M.A. and Walsh, C.T. (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* 106, 3468–3496

73. Hertweck, C. (2009) The biosynthetic logic of polyketide diversity. *Angewandte Chemie (International ed)* 48, 4688–4716

74. Sieber, S.A. and Marahiel, M.A. (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* 105, 715–738

75. Helfrich, E.J. and Piel, J. (2016) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* 33, 231–316

76. Marahiel, M.A. *et al.* (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* 97, 2651–2674

77. Hertweck, C. *et al.* (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* 24, 162–190

78. Haynes, S. and Challis, G. (2007) Non-linear enzymatic logic in natural product modular mega-synthases and-synthetases. *Curr. Opin. Drug Disc. Devel.* 10, 203–218

79. Eustáquio, A.S. *et al.* (2008) Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* 4, 69–74

80. Yamanaka, K. *et al.* (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1957–1962

81. Medema, M.H. *et al.* (2015) Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* 11, 625–631

82. Shao, Z. *et al.* (2013) Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Syn. Biol.* 2, 662–669