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