



# Retrospective analysis of natural products provides insights for future discovery trends

Cameron R. Pye<sup>a</sup>, Matthew J. Bertin<sup>b,c</sup>, R. Scott Lokey<sup>a</sup>, William H. Gerwick<sup>b,c,1</sup>, and Roger G. Linington<sup>d,1</sup>

<sup>a</sup>Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064; <sup>b</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; <sup>c</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093; and <sup>d</sup>Department of Chemistry, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved April 4, 2017 (received for review September 1, 2016)

Understanding of the capacity of the natural world to produce secondary metabolites is important to a broad range of fields, including drug discovery, ecology, biosynthesis, and chemical biology, among others. Both the absolute number and the rate of discovery of natural products have increased significantly in recent years. However, there is a perception and concern that the fundamental novelty of these discoveries is decreasing relative to previously known natural products. This study presents a quantitative examination of the field from the perspective of both number of compounds and compound novelty using a dataset of all published microbial and marine-derived natural products. This analysis aimed to explore a number of key questions, such as how the rate of discovery of new natural products has changed over the past decades, how the average natural product structural novelty has changed as a function of time, whether exploring novel taxonomic space affords an advantage in terms of novel compound discovery, and whether it is possible to estimate how close we are to having described all of the chemical space covered by natural products. Our analyses demonstrate that most natural products being published today bear structural similarity to previously published compounds, and that the range of scaffolds readily accessible from nature is limited. However, the analysis also shows that the field continues to discover appreciable numbers of natural products with no structural precedent. Together, these results suggest that the development of innovative discovery methods will continue to yield compounds with unique structural and biological properties.

natural products | chemical diversity | structural similarity | drug discovery | chemoinformatics

Natural products can be broadly defined as the set of small molecules derived from the environment that are not involved in primary metabolism. These compounds are mostly genetically encoded and produced by secondary metabolic pathways. Many of today's small molecule therapeutics trace their origins to natural products, estimated variably as providing or inspiring the development of between 50–70% of all agents in clinical use today (1). Whereas the natural environment is frequently identified as a rich source of unique chemical diversity for pharmaceutical lead-compound discovery (2, 3), the rediscovery of known natural product structures is an increasing challenge for the field (4–6). The concern in some quarters is that natural product diversity accessible by “top-down” approaches (e.g., bioassay- or chemical signature-guided isolation) has been largely exhausted, and that existing discovery models are no longer capable of delivering novel lead compounds (7). In this regard, it has been proposed that “bottom-up” approaches (e.g., genetic information-driven natural product isolation) have the capacity to access the unexpressed genetic potential of microorganisms, and can thus lead to a “renaissance” in the field of natural products (8). Ultimately, the veracity of these propositions will be revealed in their relative success records, and perhaps the more reasoned view is that the success of the discipline is best achieved by a diversity of approaches.

Antibiotic discovery is an example where there is particular concern over the ability of top-down natural products investigations

to yield fundamentally new classes of agents (9, 10). Almost all of the early antibiotic scaffolds were derived from natural sources, and there have been no new clinically approved natural product-based antibiotics discovered for over 30 y (1, 11). Even those that have entered the market more recently, such as daptomycin (12) and ticacumicin B1 (13, 14), have their discovery origins back in the 1980s. This lack of discovery of new natural product-based antibiotics, despite substantial effort in this area by both academia and industry, raises the question of whether all of the clinically relevant natural product-based antibiotics have already been discovered. This would of course present a quite terrifying prospect for patients and the biomedical community alike.

To provide a perspective on the issue of natural product structural diversity, we have performed a series of analyses on the structures of all microbial and marine-derived natural products published during the period 1941–2015. These analyses were designed to examine the rates of natural product discovery over time as well as the relationship between year of discovery and structural novelty. In this regard, such an analysis provides a description of the current state of natural products research and its ability to potentially yield new classes of therapeutic agents in the future.

To accomplish this objective, we assembled a dataset comprising all published microbial and marine-derived natural products from the period 1941–2015. The data for the period 1941–2011 are contained in the commercial database AntiMarin. The data for the period 2012–2015 were assembled for this study through manual curation of all published articles from a large panel of journals in the chemistry and chemical biology arena (for details of dataset construction see *Experimental Methods*, *Dataset Creation and Curation* and *Dataset S1*). Plant-derived

## Significance

Natural products research seems to be at a critical juncture in terms of its relevance to modern biological science. We have evaluated this landscape of chemical diversity to ask key questions, including the following. How has the rate of discovery of new natural products progressed over the past 70 y? Has natural product structural novelty changed as a function of time? Has the rate of novel discovery declined in recent years? Does exploring novel taxonomic space afford an advantage in terms of novel compound discovery? Is it possible to estimate how close we are to describing all of the chemical space covered by natural products? And, finally, is there still value in exploring natural products space for novel biologically active natural products?

Author contributions: C.R.P., M.J.B., R.S.L., W.H.G., and R.G.L. designed research; C.R.P., M.J.B., and R.G.L. performed research; C.R.P., M.J.B., and R.G.L. analyzed data; and C.R.P., R.S.L., W.H.G., and R.G.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 5564.

<sup>1</sup>To whom correspondence may be addressed. Email: wgerwick@ucsd.edu or rliningt@sfu.ca.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1614680114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1614680114/-DCSupplemental).

natural products were not included in this study due to lack of access to an appropriate plant natural products database. Nevertheless, examination of structural diversity for plant-derived natural products is an interesting question with its own set of unique challenges, and one worthy of further future analysis and investigation. For a recent study in this area see Kong et al. (4).

### Exploring Trends in Chemical Diversity

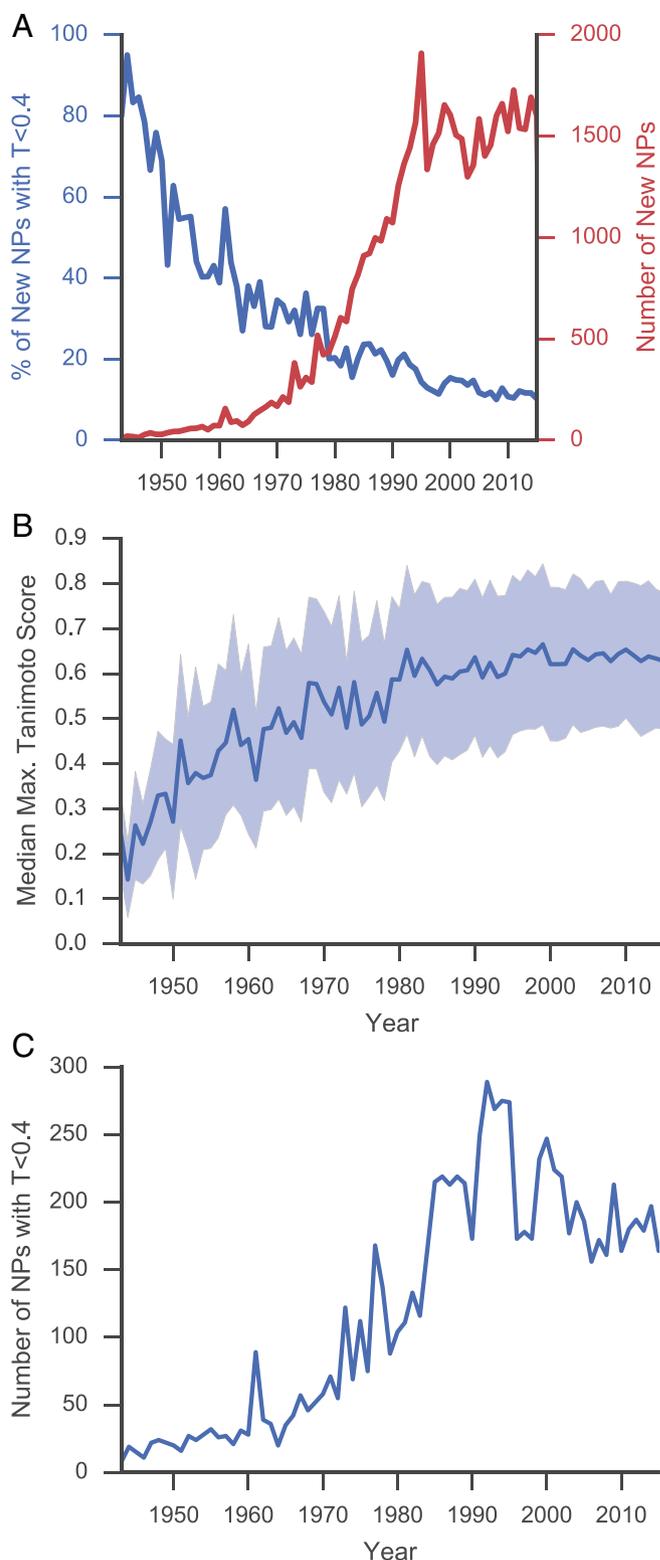
**How Has the Rate of Discovery of New Natural Products Changed As a Function of Time?** Initially we examined the rate of natural product discovery as a function of time. Fig. 1A shows that the number of compounds being published from these sources increased dramatically from relatively few compounds per year in the 1940s to an average of ~1,600 per year over the last two decades. The rate of increase in numbers of newly reported compounds was greatest from the 1970s through the mid-1990s and has remained relatively constant since then. There are likely a number of factors that contributed to the dramatic rise in natural product discoveries in the last part of the previous century. In the 1940s this was still a new field, with relatively few practitioners. Moreover, the analytical tools available were very limited, meaning that structure determination was extremely challenging and time-consuming. However, after the success of early therapeutic leads from nature, particularly the early antibiotics, the number of research groups in this area increased significantly. With the advent of better instrumentation (high-performance liquid chromatography, NMR spectroscopy, and mass spectrometry) and the invention of 2D NMR methods in the mid-1980s, the process of compound isolation and structure determination improved greatly, leading to a steady increase in the annual number of published structures from nature.

It is less clear why the number of published compounds has continued to increase or remain steady over the past two decades, even as the pharmaceutical industry has largely exited the natural products arena (15, 16). This continued productivity might reflect the increasing globalization of natural products research, with most countries that possess significant academic research infrastructure also supporting vibrant natural products research programs. For example, there has been a considerable increase in productivity in natural products research from China, Korea, Brazil, and India during this period (17). From a global perspective, the numbers indicate that there is still a healthy focus on natural products research. Moreover, the robust productivity of these efforts demonstrates that the natural world continues to provide large numbers of new and bioactive molecules year on year.

### Has the Rate of Novel Natural Product Discovery Changed in Recent Years?

Although the raw number of new molecules published per year provides some information about the productivity of the natural products research community, it provides no information about the structural novelty of the compounds being reported. To explore the structural relationships within this set of natural products, we calculated Tanimoto similarity scores between all molecule pairs (for a discussion of chemical similarity scoring methods see *Experimental Methods, Tanimoto and Tversky Scoring*) (18). We then separated the compounds into bins based on year of discovery and determined the highest Tanimoto score between each molecule in a given bin and all of the molecules published in previous years. This analysis provides a measure of the structural novelty of a given compound at the time at which it was discovered. Taken together, these results provide one metric for evaluating both trends in structural novelty (based on median values) and diversity (based on median average deviation).

As shown in Fig. 1B, it is clear that in parallel with the steady increase in number of compounds reported per year, median maximum Tanimoto scores also increased rapidly from the 1950s to 1970s (Fig. 1B, blue line). The rate of increase tapered during the 1980s and 1990s to reach a plateau at ~0.65 by the mid-1990s, a value where it remains today. A cursory review of these



**Fig. 1.** Examining structural diversity. (A) Number of compounds published per year and rate of novel compound isolation as a percentage of total natural product isolation. (B) Median maximum Tanimoto scores as a function of time. Median average deviation shown as shaded blue region. (C) Absolute number of low similarity compounds ( $T < 0.4$ ) per year. NP, natural product.

data might suggest that the field of natural products is no longer discovering novel chemical entities, and that natural products chemical space has been largely described. However, contained

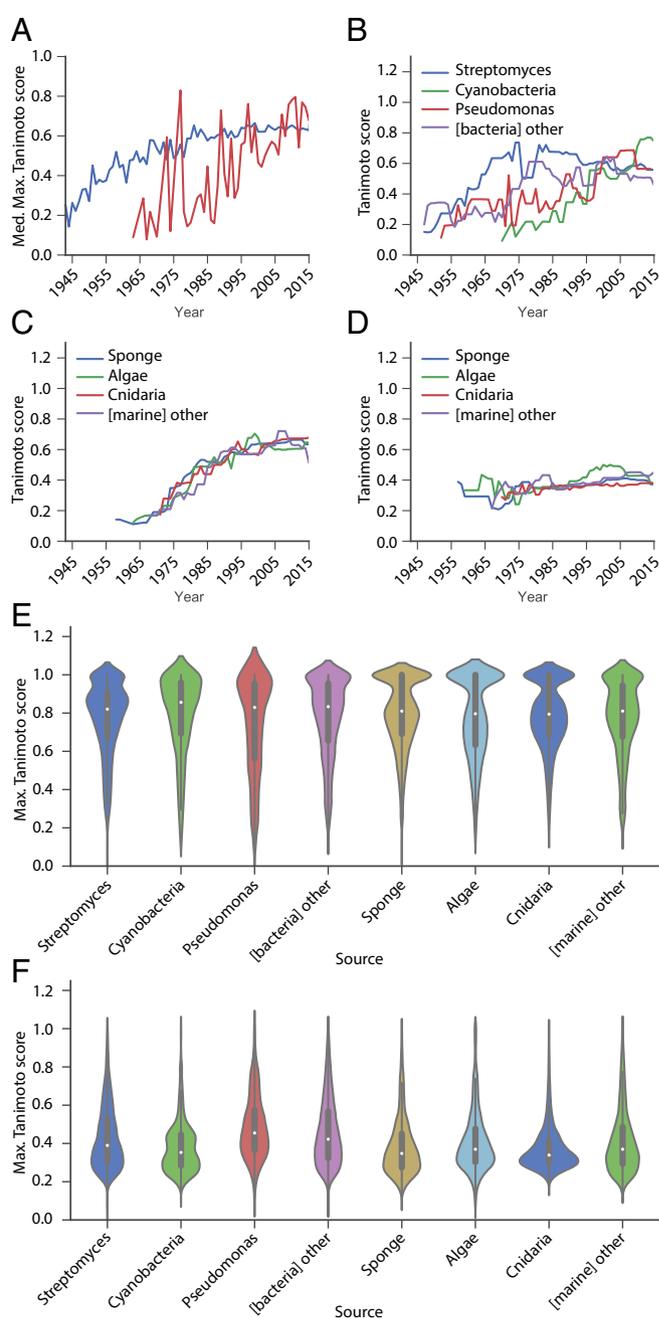
within this huge compendium of structures are many examples of fundamentally unique molecules, often with unprecedented structural and/or functional attributes. Therefore, in addition to considering the median Tanimoto score distribution, it is also important to evaluate the distribution of molecules with low similarity scores ( $T < 0.4$ ) (Fig. 1C). This analysis is interesting because it reveals that the number of novel compounds increased in parallel to the increase in number of compounds reported through the mid-1990s, followed by a steady or slightly decreasing rate of novel compound discovery in recent years. It is impressive and significant that the absolute number of molecules with low similarities remains high over this most recent period, despite the ever-increasing bar set by the addition of thousands of new structures to the dataset each year.

Overall, this analysis indicates that the discovery rate of new molecular architectures among natural products has increased since the origins of this field and has remained at a significant rate despite the ever-increasing number of published natural products (Fig. 1C). However, it should also be noted that an increasing number of the total reported natural products do have structural precedent in the literature, and thus constitute derivative structures. Overall, structurally unique compounds represent a decreasing percentage of the total number of compounds isolated from natural sources (Fig. 1A). Therefore, if structural novelty is an important and valued component of natural products research, a central question for the field becomes how do we prioritize the discovery of these unique molecules from within this large pool of natural products with known structural scaffolds.

**Source Diversity vs. Structural Diversity: Does Exploring Novel Taxonomic Space Afford an Advantage in Terms of Novel Compound Discovery?** It has long been a tenet of natural products discovery research that examination of unexplored and unusual source organisms, or those from unique environments, provides opportunities for finding novel natural products. Recent examples of such habitats include caves (19), hydrothermal vents (20), Arctic (17) and Antarctic waters (21), plant endophytes (22, 23), and vertebrate (24–26) and invertebrate (27) microbiota. To examine the relationship between organism type and chemical diversity we subdivided the dataset into subgroups within two major designations (bacterial and marine).

As an example of the impact of studying a unique type of source organism on structural novelty, we first examined the compounds in the cyanobacterial subgroup. Cyanobacteria were selected for this analysis because (i) they are morphologically distinct from other organisms and not easy to misassign in terms of fundamental taxonomic classification, (ii) they have been shown to possess the genes required to produce many of the classes of compounds associated with this phylum (28, 29), limiting the risk that these molecules are actually produced by endosymbionts from other phyla, (iii) there are a sufficient number of research groups studying cyanobacteria that the results are not likely to be biased by research strategies of an individual research group (30, 31), and (iv) the investigation of cyanobacterial metabolites significantly postdates the exploration of other sources of natural products, providing an ideal model for examining the impact of exploring new biological space on chemical novelty.

In the first plot (Fig. 2A), temporal variation in median Tanimoto scores for all compounds (blue line) were compared with the median Tanimoto scores between only the cyanobacterial metabolites (red line). As expected, the trend of the cyanobacterial compound data shows that study of a new source organism initially yielded compounds with little similarity to one another. Over time these median Tanimoto values gradually increase, suggesting that the easily accessible chemical diversity from cyanobacteria was described during this period (~1980–2000). Ultimately these values have come to match those observed



**Fig. 2.** Examining source diversity. (A) Plot of median maximum Tanimoto score by year for the full dataset (blue) and the intrasubgroup values for the cyanobacterial subgroup (red). (B) Plot of intrasubgroup median maximum Tanimoto scores by year for bacterial subgroups. (C) Plot of intrasubgroup median maximum Tanimoto scores by year for marine subgroups. (D) Plot of extrasubgroup median maximum Tanimoto scores by year for marine subgroups. (E) Violin plots for intrasubgroup median maximum Tanimoto scores for bacterial and marine subgroups. (F) Violin plots for extrasubgroup median maximum Tanimoto scores for bacterial and marine subgroups. Med., median.

for the broader set of natural products (blue line), indicating that the study of cyanobacterial chemistry has now reached maturity.

To examine the prevalence of this phenomenon in natural products discovery we plotted the temporal progression of in-class similarity scores for a selection of subgroups from the dataset. Looking at the subgroups for the bacterial designation (*Streptomyces*, *Pseudomonas*, Cyanobacteria, and other; Fig. 2B) a number of interesting observations emerge. First, as expected,

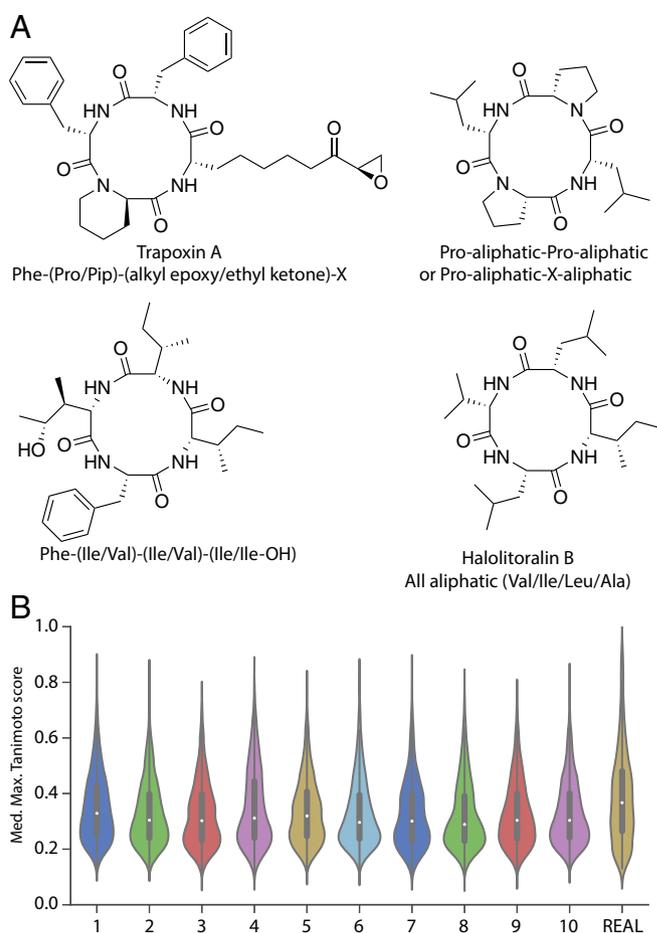
all plots have upward trends with values in later years in the region 0.6, suggesting that there is a limited period in which unique chemistry can be easily found within a given source type. This phenomenon is even more pronounced among the subgroups from the marine environment. Fig. 2C shows that compound similarities follow very similar trends, regardless of source subgroup, demonstrating that we have a mature understanding of the predominant chemistries likely to be encountered from these sources. By contrast, the data for bacterial sources (Fig. 2B) do not follow such strong trends. In the case of *Streptomyces* the number of compounds is similar to those from the Porifera subgroup (6,547 vs. 7,263) and the taxonomic classification is narrower (genus vs. phylum). However, the median maximum Tanimoto scores are more variable, with values in the past 30 y following a decreasing trend. There are many factors that could contribute to this trend including application of new discovery strategies, higher chemical diversity, and differences in program objectives or discovery models among researchers in the two subdisciplines. However, what is apparent from these results is that at present the average molecule reported from bacteria of the genus *Streptomyces* is less similar to other *Streptomyces* compounds than is the case with compounds within the subgroup Porifera.

Interestingly, when compounds in a given marine subgroup were compared against compounds from all other subgroups in the marine set the median Tanimoto scores remained low, regardless of year (Fig. 2D). This supports the long-held understanding in the marine natural products community that the chemistries derived from a given organism type are often fundamentally different from the chemistries encountered from all other organisms in that environment. These results suggest that novel sources of natural products have been, and remain, an important and productive source of novel chemical diversity, albeit with a limited period of expected novel compound discovery. The counterpoint to this conclusion is that, in the absence of significant innovation in discovery approaches, there are diminishing returns in terms of the discovery of fundamentally new chemical diversity from continued investigations of the same classes of organisms.

Finally, to examine the absolute chemical diversity within each subgroup we calculated the maximum Tanimoto score for each compound compared with either other compounds within that subgroup (Fig. 2E), or to other members of the source type (bacterial or marine; Fig. 2F), irrespective of year of discovery. Fig. 2E illustrates that most subgroups contain moderate to large numbers of compounds with high similarities to one or more other compounds in that class. This is particularly apparent for Porifera, algae, and Cnidaria, all of which have large “hammerhead” distributions (Fig. 2E). However, compared with the other subgroups within that source type (e.g., Cnidaria vs. all other marine compounds; Fig. 2F) the distribution of maximum Tanimoto scores is centered around very low values, confirming the previous observation that the chemistries from these source organisms are typically not found elsewhere in the marine world.

**Evaluating the Chemical Space Occupied by Natural Products: How Much of the “Natural Product-Like” Chemical Space Is Actually Occupied by Natural Products?** Recently, there have been a number of efforts to describe natural product chemical space and to use this space in various ways including as a boundary for designing natural product-like synthetic screening libraries (32, 33). For a given biosynthetic class of natural products, a very large number of theoretical molecules can be created from primary building blocks, such as amino acids, sugars, acetate and propionate, mevalonate, and so on. Using these diverse and often chiral components, natural product libraries should therefore exceed synthetic libraries in terms of structural diversity of chemical scaffolds.

One way to explore this hypothesis is to examine the chemical diversity within classes of compounds that are easily and accurately characterized by boundary conditions. To explore this idea we examined the chemical diversity of all currently published cyclic tetrapeptides from our dataset. Cyclic tetrapeptides were chosen because these are relatively easy to identify from the dataset, and the building blocks are well-defined. There are numerous ways to estimate the theoretical number of molecules that are possible within this class; thus, to simplify the analysis, we considered only the 20 proteinogenic amino acids as possible building blocks. Although this oversimplifies the analysis (e.g., nonproteinogenic and D-amino acids are excluded), it provides at least a minimum theoretical limit for possible structural diversity. With four positions in the peptide and 20 aa building blocks there are 40,110 possible molecules that can be produced. This is less than the  $20^4$  possible compounds one might expect, because of the rotational symmetry of some conceivable products. However, to our surprise, examination of the natural product database developed in this study revealed just 65 cyclic tetrapeptides, the majority of which fall into just four structural classes (Fig. 3A and Fig. S1). The largest class, containing 27 members, incorporates a residue rarely encountered elsewhere in nature containing an alkyl chain terminating in either an epoxy ketone or an ethyl ketone; trapoxin A (34) and apicidin A (35) are well-known



**Fig. 3.** Theoretical vs. actual structural diversity. (A) Examples of the four major classes of cyclic tetrapeptides found in nature. (B) Violin plots indicating the distribution of Tanimoto scores between all members of 65 randomly selected theoretical cyclic peptides (10 trials, lanes 1–10) and between all 65 cyclic tetrapeptides from our natural product dataset (lane REAL). Ala, alanine; Ile, isoleucine; Leu, leucine; Med., median; Val, valine.

members of this family. Compounds in this class have been isolated by 20 different research groups over a 23-y span, suggesting that this cluster is large because this structural motif is relatively widespread in the environment, rather than having been the subject of intense investigation by a small number of specialized research teams.

The precise reasons for this disconnection between theoretical and observed chemical diversities are unknown but are likely due to the combined factors of physical organic chemistry (i.e., some of the conceivable molecules are not structurally plausible or are difficult to form) and chemical ecology (i.e., molecules produced in nature must confer a competitive advantage to the producing organism). Although not explored further in this paper, this trend that only a few representatives in a given structure class are produced in nature is a general phenomenon that holds true for many of the compound classes published to date. Therefore, we conclude that although the theoretical chemical space offered by natural products is very high, the number of actual compounds discovered within a given chemical class is quite low, despite all of the different screening and isolation approaches that have been used by the field in the preceding 70 y.

Furthermore, it is a general observation that compounds that are produced within a class often share many structural features, indicating that the compounds produced are not a random subset of the possible molecules, but rather possess moderate diversification within a large number of highly refined structural constraints. To explore this phenomenon within the cyclic tetrapeptide example we calculated the Tanimoto scores between all members of a set of 65 randomly selected cyclic tetrapeptides from among the 40,110 possible unique structures. Comparing the results from 10 independent trials of this experiment with the distribution from the naturally occurring molecules (Fig. 3B) we observed much higher frequencies of structural relatedness within the naturally occurring compounds than within any of the 10 randomly selected compound sets. This result supports the idea that, at least in the case of cyclic tetrapeptides, structural diversity in nature is centered on a select set of key scaffolds, rather than being randomly distributed throughout the available chemical space. An inescapable conclusion of this observation is that selection pressures at work in the natural world significantly limit the scope of structural diversity created in most compound classes.

**How Are the Known Natural Product Structures Distributed Within Chemical Space?** We next turned our attention to the question of how the 52,395 unique molecules in our dataset compared with one another in terms of structural relatedness. To allow molecules that are substructures of larger scaffolds to achieve high similarity scores we used a different distance metric for clustering molecular structures. Tversky scores ( $\alpha = 0.1$ ,  $\beta = 0.9$ ) were calculated for each pair of molecules in each direction ( $i,j$  and  $j,i$ ) and averaged to afford a single similarity score (36). Edges between nodes in the network represent averaged Tversky scores greater than or equal to 0.8. Finally, structure graphics were overlaid using the ChemViz Cytoscape plugin. In this way, we created a network diagram that describes the structural relatedness of all 52,395 compounds within the database (Fig. S2).

This analysis revealed that the dataset contained 6,414 clusters composed of two or more compounds, which together accounted for 40,229 compounds. In addition, the network contained a further 12,166 compounds that did not belong to any cluster. Therefore, 76.8% of the chemical space occupied by published natural products from these sources is described by fewer than 6,500 scaffolds, indicating that these known natural product scaffolds occupy a relatively narrow percentage of the total available natural product-like chemical space. This observation suggests that either the selective pressures from nature on the molecular evolution of natural products are convergent for particular compound classes or that we have not yet developed the technologies required to access natural compounds from the vast majority of the chemical space they are predicted to occupy.

**How Close Are We to Having Described All of the Chemical Space Covered by Natural Products?** This key question is extremely difficult to answer definitively without further improvements in a number of areas, such as (i) enhancing the ability to accurately predict structural classes from genome sequence data, (ii) refining the capacity to find biosynthetic genes from genomic DNA, and (iii) developing the facility to derive complete whole-genome sequencing data from complex environmental samples. Nevertheless, it is clear from the temporal trends in similarity values discussed above that a significant number of molecules published in recent years bear close structural similarity to established scaffolds.

However, over 23% of the natural products in the database have low structural similarity to all other compounds and therefore appear as singletons in the network diagram. Examples of compounds with low similarity scores are shown in Fig. S3. These compounds, all of which possessed maximum Tanimoto scores of 0.4 or less at the time of discovery, encompass a broad array of biosynthetic origins, structural complexities, source organisms, and discovery methods, suggesting that there remains significant opportunity for novel compound discovery across a range of sources and scientific approaches. It is clear that there are many unique and structurally novel compounds in the natural products world, and the rate of discovery of these unusual structures is not changing significantly, despite the presence of a large and growing canon of known natural product structures (Fig. 1C). Although there seems to be much left to discover that is truly novel, significant innovation will be required to access these novel compounds in an efficient manner and retain the impressive historical rate of novel compound discovery from the natural world.

#### Limitations of this Analysis/Points to Consider

It is important to recognize that novel structure, although scientifically intriguing, is not necessarily the primary driver of interest in natural products. That position arguably rests in the realm of biology. It is the biological roles of natural products that are the ultimate source of their value to human society and the environment. Moreover, it is widely known that small and subtle changes in a molecule's structure have the capacity to alter it from being "inactive" to being "exquisitely potent." Therefore, it must be recognized that there remains substantial value in known and derivative natural product structures, provided that these structures possess unique attributes from a biological perspective.

Natural products occupy chemical space that is not well represented by synthetic libraries, and thus there remains high value in natural product scaffolds for biomedical applications (37, 38). Activity of a natural product against a biomolecular target that has no other small molecule modulator is of enormous value, regardless of its relative structural novelty. In this regard, comparatively little is known about the biological activities of most natural products in either ecological or biomedical contexts because in many cases evaluation of natural product bioactivities has been limited to basic cytotoxicity, antibacterial, and antifungal whole-cell assays. A vast opportunity to discover and characterize the biological function of natural products in complex biological systems therefore remains largely unexplored. Such investigation need not be limited to new or novel compounds but rather should seek to understand their endogenous roles in nature and/or human health applications, regardless of structural novelty.

A further limitation of the current analysis is that it only incorporates published structures and does not provide any estimates as to the global capacity for natural product production. Additionally, it only includes organisms that have been studied to date. The majority of taxonomic space has had no systematic examination of its capacity for natural products production (39). Nevertheless, the result of this current analysis suggests that unstudied source organisms with unique metabolic and/or environmental

constraints should represent excellent sources of structurally novel natural products. Finally, this analysis does not take into consideration the products of “cryptic” biosynthetic gene clusters, many of which are predicted to produce novel natural products based on bioinformatic analyses (40, 41).

### Future Perspective

The results of our analysis indicate that the future for natural products is very bright indeed. From a variety of lines of evidence, including genetic analysis of the sequenced genomes of microorganisms and the trends documented herein, a large reservoir of chemical space exists in natural products. This has yet to be fully explored via traditional approaches, although accessing novel genetic resources as well as new biological prioritization methods are assisting these endeavors (42, 43). Nevertheless, it is imperative that the field aggressively innovate if we are to avoid increasing redundancy of effort and marginalization of natural products research in the areas of chemical biology and biotechnology. Given the trends observed in these data, it is reasonable to suggest that in most cases traditional natural product discovery platforms implemented on traditional source organisms will lead predominantly to the isolation of traditional, well-known chemical entities.

The goal of discovering a new lead compound to treat human disease is exceptionally complex. Not only does the agent need to be efficacious but it also must have appropriate bioavailability,

absorption, distribution, metabolism, and excretion and pharmacokinetic properties, and lack of toxicity. As a result, it is not only attractive, but absolutely necessary, to use a diversity of approaches to the discovery of such a lead structure, be it “top-down” or “bottom-up” natural products discovery, screening of synthetic libraries, medicinal chemistry, fragment-based or structure-based design, or some combination of the above. We should not become overly wedded to one approach, because robustness in the field of drug discovery derives in large part from this diversity of approaches and is therefore to be embraced.

Natural products have important roles to play beyond simply the reporting of novel chemical structures. Rather, contributions to ecology, biotechnology, and biomedicine will continue to form the backbone for natural products research programs in the coming decades. Development of new strategies and methods to better integrate natural product libraries into the modern biotechnology arena should therefore be considered a critical focus for academia, funding agencies, and industry alike. With the many innovations that are being developed in this highly multidisciplinary field, it will be exciting to see where the next 70 y of natural products research will lead.

**ACKNOWLEDGMENTS.** This work was funded by NIH Grants AT008718 (to R.S.L. and R.G.L.) and GM107550 (to W.H.G.) and Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2016-03962 (to R.G.L.).

- Newman DJ, Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* 79:629–661.
- Harvey AL, Edrada-Ebel R, Quinn RJ (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 14:111–129.
- Gwynn MN, Portnoy A, Rittenhouse SF, Payne DJ (2010) Challenges of antibacterial discovery revisited. *Ann N Y Acad Sci* 1213:5–19.
- Kong D-X, Guo M-Y, Xiao Z-H, Chen L-L, Zhang H-Y (2011) Historical variation of structural novelty in a natural product library. *Chem Biodivers* 8:1968–1977.
- Walsh CT (2015) A chemocentric view of the natural product inventory. *Nat Chem Biol* 11:620–624.
- Ju K-S, et al. (2015) Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc Natl Acad Sci USA* 112:12175–12180.
- Jensen PR, Chavarría KL, Fenical W, Moore BS, Ziemert N (2014) Challenges and triumphs to genomics-based natural product discovery. *J Ind Microbiol Biotechnol* 41:203–209.
- Bachmann BO, Van Lanen SG, Baltz RH (2014) Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *J Ind Microbiol Biotechnol* 41:175–184.
- Wright GD (2014) Something old, something new: Revisiting natural products in antibiotic drug discovery. *Can J Microbiol* 60:147–154.
- Walsh CT, Wenciewicz TA (2014) Prospects for new antibiotics: A molecule-centered perspective. *J Antibiot (Tokyo)* 67:7–22.
- Patridge E, Gareiss P, Kinch MS, Hoyer D (2016) An analysis of FDA-approved drugs: Natural products and their derivatives. *Drug Discov Today* 21:204–207.
- Debono M, et al. (1987) A21978C, a complex of new acidic peptide antibiotics: Isolation, chemistry, and mass spectral structure elucidation. *J Antibiot (Tokyo)* 40:761–777.
- Theriault RJ, et al. (1987) Tiacumicins, a novel complex of 18-membered macrolide antibiotics. I. Taxonomy, fermentation and antibacterial activity. *J Antibiot (Tokyo)* 40:567–574.
- Hochlowski JE, et al. (1987) Tiacumicins, a novel complex of 18-membered macrolides. II. Isolation and structure determination. *J Antibiot (Tokyo)* 40:575–588.
- Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* 4:206–220.
- David B, Wolfender J-L, Dias DA (2014) The pharmaceutical industry and natural products: Historical status and new trends. *Phytochem Rev* 14:299–315.
- Abbas S, et al. (2011) Advancement into the Arctic region for bioactive sponge secondary metabolites. *Mar Drugs* 9:2423–2437.
- Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50:742–754.
- Derewacz DK, et al. (2014) Structure and stereochemical determination of hypogeamins from a cave-derived Actinomycete. *J Nat Prod* 77:1759–1763.
- Andrianasolo EH, et al. (2011) Bathymodiolamides A and B, ceramide derivatives from a deep-sea hydrothermal vent invertebrate mussel, *Bathymodiolus thermophilus*. *J Nat Prod* 74:842–846.
- Diyabalanage T, Amsler CD, McClintock JB, Baker BJ (2006) Palmerolide A, a cytotoxic macrolide from the antarctic tunicate *Synoicum adareanum*. *J Am Chem Soc* 128:5630–5631.
- Kusari S, Hertweck C, Spiteller M (2012) Chemical ecology of endophytic fungi: Origins of secondary metabolites. *Chem Biol* 19:792–798.
- Schueffler A, Anke T (2014) Fungal natural products in research and development. *Nat Prod Rep* 31:1425–1448.
- Motley JL, et al. (2016) Opportunistic sampling of roadkill as an entry point to accessing natural products assembled by bacteria associated with non-anthropoidal mammalian microbiomes. *J Nat Prod* 80:598–608.
- Donia MS, et al. (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* 158:1402–1414.
- Theodore CM, et al. (2014) Genomic and metabolomic insights into the natural product biosynthetic diversity of a feral-hog-associated *Brevibacillus laterosporus* strain. *PLoS One* 9:e90124.
- Lin Z, et al. (2013) A bacterial source for mollusk pyrone polyketides. *Chem Biol* 20:73–81.
- Calteau A, et al. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics* 15:977.
- Kehr J-C, Gatte Picchi D, Dittmann E (2011) Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes. *Beilstein J Org Chem* 7:1622–1635.
- Salvador LA, et al. (2013) Potent elastase inhibitors from cyanobacteria: Structural basis and mechanisms mediating cytoprotective and anti-inflammatory effects in bronchial epithelial cells. *J Med Chem* 56:1276–1290.
- Kleigrewe K, et al. (2015) Combining mass spectrometric metabolic profiling with genomic analysis: A powerful approach for discovering natural products from Cyanobacteria. *J Nat Prod* 78:1671–1682.
- Rosén J, Gottfries J, Muresan S, Backlund A, Oprea TI (2009) Novel chemical space exploration via natural products. *J Med Chem* 52:1953–1962.
- Pascolutti M, et al. (2015) Capturing nature's diversity. *PLoS One* 10:e0120942.
- Itazaki H, et al. (1990) Isolation and structural elucidation of new cyclooctapeptides, trapoxins A and B, having detransformation activities as antitumor agents. *J Antibiot (Tokyo)* 43:1524–1532.
- Darkin-Rattray SJ, et al. (1996) Apicidin: A novel antiprotozoal agent that inhibits parasite histone deacetylase. *Proc Natl Acad Sci USA* 93:13143–13147.
- Kunimoto R, Vogt M, Bajorath J (2016) Maximum common substructure-based Tversky index: An asymmetric hybrid similarity measure. *J Comput Aided Mol Des* 30:523–531.
- López-Vallejo F, Giulianotti MA, Houghten RA, Medina-Franco JL (2012) Expanding the medicinally relevant chemical space with compound libraries. *Drug Discov Today* 17:718–726.
- Feher M, Schmidt JM (2003) Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci* 43:218–227.
- Cimermancic P, et al. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158:412–421.
- 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 USA* 111:1957–1962.
- Rutledge PJ, Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* 13:509–523.
- Kurita KL, Glassey E, Linington RG (2015) Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc Natl Acad Sci USA* 112:11999–12004.
- Park SR, et al. (2016) Discovery of cahuitymamins as biofilm inhibitors derived from a convergent biosynthetic pathway. *Nat Commun* 7:10710.