LETTER

REPLY TO SKINNIDER AND MAGARVEY: Rates of novel natural product discovery remain high

Cameron R. Pye[a], Matthew J. Bertin[b,c], R. Scott Lokey[a], William H. Gerwick[b,c,1], and Roger G. Linington[d,1]

It is encouraging that our recent article examining trends in discovery rates and structural diversity for natural products (NP) (1) is generating discussion in this fascinating area (2). However, we wish to correct several misconceptions presented in the comments from Skinnider and Magarvey (3).

Skinnider and Magarvey's (3) letter incorrectly summarizes the key conclusion of our work. The letter states that "[t]heir analysis suggests that the pace of structurally unique NP discovery is decreasing." Our study makes precisely the opposite conclusion: "A cursory review of these data might suggest that the field of natural products is no longer discovering novel chemical entities. . .[However,] it is also important to evaluate the distribution of molecules with low similarity scores. . . 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..." (1).

Skinnider and Magarvey (3) raise two main concerns about our (1) analyses. First, they suggest that the observed trends in similarity are due solely to increasing sample size, and do not inform questions related to NP diversity. In fact, as noted by Burke and coworkers (2), it is precisely this rise in values that allows us to conclude that the NP chemical space discovered to date is bounded; if available NP chemical space were vastly greater than what has been observed, this curve (see figure 1B in ref. 1) would contain almost uniformly low values. The analysis by Skinnider and Magarvey (3), replacing a subset of NPs with compounds from the ZINC database (4), yields much lower similarity values, demonstrating that: (*i*) NPs are not like ZINC compounds and (*ii*) if one removes some of the NPs, then the ability to describe the boundary of NP space goes down. This finding is further supported by our analysis of source subclasses (figure 2D in ref. 1), which demonstrates that compounds from one source subclass bear low structural similarities to all other marine compounds, regardless of how many compounds are added to the dataset. These plots counter the suggestion that library size alone is responsible for the observed trends.

Second, Skinnider and Magarvey (3) question our (1) conclusion that more compounds published in recent years are derivatives of known scaffolds than was true in previous decades. To support their position, Skinnider and Magarvey (3) compare actual similarity trends against a randomized dataset. For practical reasons this approach is fundamentally flawed because it ignores the fact that many NP manuscripts report multiple family members in a single article. Taking 2015 as an example, our dataset contains 1,576 compounds derived from 484 papers. Of these, 49% of compounds possess at least one other compound in the same article with a Tanimoto score > 0.9. Our original analysis excludes intra-article comparisons because it only compares compounds to those found in previous years. By randomizing compounds across bins, Skinnider and Magarvey (3) have introduced a large number of derivative relationships between family members, precluding direct comparison between the actual and randomized datasets. Therefore, we must conclude that the analyses presented in Skinnider and Magarvey's (3) letter do not accurately reflect the current situation for NP research. Instead, to echo our original conclusion, "the future for natural products is very bright indeed" (1).

1 Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Linington RG (2017) Retrospective analysis of natural products provides insights for future discovery trends. *Proc Natl Acad Sci USA* 114:5601–5606.

2 Palazzolo AME, Simons CLW, Burke MD (2017) The natural productome. *Proc Natl Acad Sci USA* 114:5564–5566.

3 Skinnider MA, Magarvey NA (2017) Statistical reanalysis of natural products reveals increasing chemical diversity. *Proc Natl Acad Sci USA* 114:E6271–E6272.

4 Irwin JJ, Shoichet BK (2005) ZINC—A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182.

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

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., W.H.G., and R.G.L. analyzed data; and C.R.P., M.J.B., R.S.L., W.H.G., and R.G.L. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: wgerwick@ucsd.edu or rliningt@sfu.ca.