



Natural products discovery needs improved taxonomic and geographic information

Journal:	<i>Natural Product Reports</i>
Manuscript ID	NP-PER-10-2015-000130.R2
Article Type:	Viewpoint
Date Submitted by the Author:	10-Feb-2016
Complete List of Authors:	Leal, Miguel; EAWAG: Swiss Federal Institute of Aquatic Science and Technology, Department of Fish Ecology & Evolution Hilario, Ana; Universidade de Aveiro, Departamento de Biologia & CESAM Munro, Murray; University of Canterbury, Department of Chemistry Blunt, John; University of Canterbury, Department of Chemistry Calado, Ricardo; University of Aveiro, Department of Biology & CESAM



Natural Product Reports

Viewpoint

Natural products discovery needs improved taxonomic and geographic information

Received 00th January 20xx,
Accepted 00th January 20xx

DOI:
10.1039/x0xx00000x

www.rsc.org/

Miguel C. Leal,^{a†} Ana Hilário,^b Murray H. G. Munro,^c John W. Blunt,^c and Ricardo Calado^b

Marine and terrestrial organisms yield a remarkable chemical diversity and are important sources for discovery of new chemical products. In order to maximize the bioprospecting efficiency of natural products (NP), taxonomy, geography and biodiversity are starting to be used to draw conclusions on which taxonomic groups and/or regions may be of interest for future research. However, accurate taxonomic information and sampling location of source organisms have often been overlooked. Although these issues were already reported a few decades ago and improvements have been made, such outstanding problems are still recurrent in recent peer-reviewed literature. Here, we focus on the importance of taxonomic and geographic identification of source material and illustrate how taxonomic and geographic data of source organisms continues to be poorly handled. It is our opinion that this issue needs to be discussed within the NP community with the ultimate goal of improving publication standards and guaranteeing the scientific principle of research reproducibility. Moreover, by doing so, it will be possible to take advantage of information available in the literature to develop cross-disciplinary meta-analyses that may help to advance the state of the art of NP research and future bioprospecting endeavours.

1. Introduction

Marine and terrestrial organisms have been a constant and growing source of new natural products (NP) and have already provided a number of promising leads for a plethora of applications to the pharmaceutical, nutraceutical and cosmeceutical industries. The ever growing interest for discovering new NPs has motivated researchers to develop sampling strategies to survey chemical diversity more efficiently, such as: (i) explore untapped geographical sources, (ii) explore overlooked groups of organisms, or (iii) combine these two strategies. Other, more conservative, strategies include (iv) target organisms already known to yield a large chemical diversity, (v) target geographic regions where a large number of NPs has already been discovered, and (vi) a combination of these two. In order to follow any of these strategies it is important to have accurate information on NPs previously described, particularly what species were sampled, where they were collected, and what compounds were isolated. This information is critical to guarantee one of the main principles of the scientific method – that research is reproducible. Additionally, accurate taxonomic and geographic information allows for meta-analytic statistical approaches that use biogeographic concepts and have as their ultimate

goal to reveal untapped regions and groups of organisms for a more effective discovery of new NPs.^{1,2}

Taxonomic experts have historically performed the identification of sampled organisms yielding new NPs, with voucher specimens being deposited in museums or institutional collections. In the past few decades the science of taxonomy has been continuously updated, particularly with the advent of molecular genetics. Consequently, taxonomic information has been regularly revised and updated with description of new species.³ The accuracy of the geographic location of the sampling site has also been changing at an unprecedented level due to technological developments. Simple descriptions of collection sites that solely included country or region (e.g. Hainan province, Great Australian Bight, Indonesia) were often reported in early literature. Currently, the accurate identification of sampling location notably benefits from the advent of space-based navigation methods, such as Global Positioning System (GPS). This is an affordable and user-friendly technology that can be easily implemented by researchers during field campaigns. Despite notable methodological developments to accurately address the taxonomic identification of collected specimens and geographically pinpoint sampling locations, inconsistencies are still observed in recent literature describing new NPs.

2. Challenges and recurrent problems

Accurate classification of sampled organisms is critical, although not always easy. This is particularly problematic for a number of groups of organisms that have been intensively

^a Department of Fish Ecology and Evolution, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Center for Ecology, Evolution and Biogeochemistry, Kastanienbaum, Switzerland.

^b Departamento de Biologia & CESAM, Universidade de Aveiro, Aveiro, Portugal.

^c Department of Chemistry, University of Canterbury, Christchurch, New Zealand.

† miguelcleal@gmail.com

studied, such as microorganisms, plants, sponges and cnidarians, and are usually difficult to identify to species level relying solely on standard morphological features. The NP community has acknowledged this limitation and circumvented this constraint by collaborating with taxonomic experts to accurately identify the source organisms yielding NPs and by depositing voucher specimens in museums or natural history collections. As DNA sequence analysis is increasingly encouraged as a tool to assist with taxonomic identification it is advised that genetic information from screened specimens is deposited in online open-access repositories, similar to the principle of depositing a voucher specimen in a museum.

Overall, it must be acknowledged that taxonomy is not a straightforward task for NP researchers. Taxonomic classification is not always consensual and may require expertise and infrastructures that may not be readily available, particularly those classifications requiring advanced genetic tools. Moreover, a solid taxonomic classification is also hindered by a widely recognized and growing shortage of trained taxonomists and curators worldwide. Taxonomy and systematics are evolving fields just like any other, and whenever species identification is not possible, it is important to retain good vouchers that are publicly available. However, there are still a number of publications where organisms are solely described at Family level without a single reference to voucher specimens or DNA sequence information. It is also possible to read taxonomic classifications in the experimental sections of many publications stating “unidentified bacterium” or “unidentified sponge”.

Accurate geographical descriptions of sampling sites are equally important. In contrast to taxonomy, geographic locations are not constantly being revised nor updated, and do not require expert knowledge in order to be accurately described. As electronic tools equipped with GPS technology are increasingly common and affordable, it is our opinion that accurate sampling locations with coordinate information should be made available in all publications reporting new NPs. However, this is not possible for publications using emergent technologies to screen specimens preserved in museum collections.

We noticed numerous inaccurate descriptions of sampling sites during the assemblage of the biogeographic component of the *MarinLit* database.⁴ Although it is not our goal to report publications illustrating poor geographic descriptions, it is important to provide some examples. For instance, several publications report marine molecules isolated from specimens collected in the Philippines or in Indonesia. These two countries have relatively large Exclusive Economic Areas (1,600,000 and 6,160,000 km², respectively), which makes it impossible to accurately assess where sampling took place. Other publications refer to NPs isolated from organisms collected in the Adriatic Sea or in the Great Australian Bight, which are both massive bodies of water. We would also like to note two other extreme examples of geographic inaccuracies. The first is the description of new NPs from microorganisms isolated from within a “floating piece of wood”. The second is

the discovery of new molecules from a well-known deep-sea mussel that was sampled in the Atlantic Ocean; however this mussel species is solely known to occur in the Pacific Ocean. Besides the description of nonspecific sampling location, it is also regrettably true that a large fraction of the literature describing new NPs does not provide any geographic detail on the source organisms yielding those compounds.

3. Lessons learned?

Despite the numerous examples of poorly described sampling sites, it is possible to observe an optimistic trend in the past decades, as the number of publications with non-existent or inaccurate geographic descriptions has been decreasing over time.⁴ It is important that this pattern is maintained in the future, so that the research community can benefit the most from past research efforts. It is our opinion that such an optimistic trend is associated with a growing awareness of the NP research community for the importance of providing accurate data beyond that of chemical structure and bioactivity. It is also important to note the critical role that must be played by peer reviewers and journal editors by critically assessing whether the best possible taxonomic and geographic information of screened organisms is being provided.

As international peer-reviewed journals play a key-role on the quality threshold of published research, we have surveyed the author guidelines of the top 10 journals commonly selected by NP researchers to publish their findings.⁵ We noticed that five of these journals (*Journal of the American Chemical Society*, *Journal of Organic Chemistry*, *Marine Drugs*, *Organic & Biomolecular Chemistry*, *Organic Letters*) have no specific guidelines regarding the taxonomic classification of biological material, nor of its sampling location. Three other journals (*Chemical and Pharmaceutical Bulletin*, *Tetrahedron*, *Tetrahedron Letters*) only indicate that DNA sequences should be deposited in online repositories. The journal *Phytochemistry* has guidelines for the taxonomy of sampled specimens, as it requires the identification of the voucher specimen, the name and address of the expert who identified the organism and encourages deposition of the sample in a herbarium. This same journal also requires the identification of the strain and culture of screened microorganisms, and encourages the submission of DNA sequences. The only journal in this top 10 that provides detailed guidelines for taxonomic identification of biological material and its collection site is the *Journal of Natural Products*. The author guidelines provided by this journal clearly state that the source of the biological material screened for new NPs must be described as “cultivated or from a natural habitat”, and that authors need to indicate the deposit site and its voucher number. Moreover, authors are also instructed to provide information on collection date and the exact sampling location “using a GPS navigation tool”. Strain designation number and culture collection of all microorganisms used experimentally must also be provided, as well as a clear identification for their scientific name.

4. Conclusions and future perspectives

The discovery of new NPs has been continuously benefiting from technological developments of screening platforms, which have made the description of molecules more accurate and precise, thereby increasing reproducibility. However, the founding element of NP research is the biological organism producing the molecule of interest. Consequently, the information provided on the taxonomy and sampling location of the source organisms should be equally accurate to that of the chemical structure and bioactivity of the NPs being described.

It is our opinion that this issue needs to be urgently raised, and with this viewpoint article we aim to promote a discussion that will ultimately inspire the scientific community to acknowledge the relevance of taxonomic and geographic data on NP research. We would also like to encourage journals publishing research describing new NPs to improve their standards concerning the taxonomic identification and geographic information on the origin of screened biological material. This is of interest to all NP chemists, as well as to the entire community associated with NP research, both upstream and downstream of NP discovery. Such accurate information will ultimately maximize the success of future screening efforts and allow meta-analysis using cross-disciplinary approaches that may ultimately lead to significant breakthroughs in NP research.

Acknowledgements

We thank two anonymous reviewers for their important comments and suggestions to improve the manuscript. This work was supported by Fundação para a Ciência e Tecnologia (UID/AMB/50017/2013).

Notes and references

- 1 M. C. Leal, J. Puga, J. Serôdio, N. C. M. Gomes and R. Calado, *PLoS One*, 2012, **7**, e30580.
- 2 M. C. Leal, M. H. G. Munro, J. W. Blunt, J. Puga, B. Jesus, R. Calado, R. Rosa and C. Madeira, *Nat. Prod. Rep.*, 2013, **30**, 1380–1390.
- 3 WoRMS Editorial Board, 2015, World Register of Marine Species (www.marinespecies.org)
- 4 MarinLit, 2015, MarinLit: A database of the marine natural products literature (<http://pubs.rsc.org/marinlit>)
- 5 J. W. Blunt, B. R. Copp, M. H. G. Munro, P. T. Northcote and M. R. Prinsep, *Nat. Prod. Rep.*, 2010, **27**, 165–237.