

<https://doi.org/10.1038/s44183-024-00088-0>

Conducting marine genetic research for whom? Mapping knowledge flows from science to patents

Paul Dunshirn¹ ✉ & Erik Zhivkopoulos²

Marine genetic resources hold great value for biotechnological innovation and sustainability-oriented research. However, studies indicate that intellectual property rights pertaining to these resources are concentrated in a handful of companies and countries, triggering contentious international debates. This paper highlights an overlooked dimension: the flow of scientific knowledge in the provision and use of marine genetic research. We identified 23,417 scientific studies that use marine genetic sequences, traced patent references to these studies, and revealed provider and user countries. We found that molecular biological research, oftentimes on bacterial species, attracted the most patenting interest. Some countries owned relatively more patents than they provided in science, with the U.S. benefiting the most from available research. Science by developing countries' researchers saw limited uptake in patents. These findings aim to inform international biodiversity regimes, such as the UN Treaty on Biodiversity Beyond National Jurisdiction (BBNJ), on how to more equitably distribute marine genetic resources' economic, socio-cultural, and ecological benefits.

Having evolved over 3.7 billion years, marine biodiversity constitutes a highly rich yet understudied pool of genetic diversity¹. Conceptualized as 'marine genetic resources', genetic materials from the ocean and the sequence data derived thereof have attracted increased economic interest as a source for biotechnological innovation in recent decades². For instance, scientists have turned to the deep sea to search for new antibiotics in the face of emerging multidrug resistance³. Microorganisms of the genus *Pseudomonas* were found to degrade plastics, which has led scientists to consider these properties in bioremediation strategies⁴. The market value for marine biotechnology has been projected to reach \$6.4 billion by 2025⁵. Beyond commercial value, genes and genetic diversity provide essential ecosystem services⁶. This includes ecological benefits, as the genetic variability of a species can play a crucial role in its adaptive capacity and resilience to environmental pressures^{7,8}, such as ocean warming^{9,10}, acidification⁹, or decreases in salinity levels¹¹. As most attributes of organisms are encoded within genes, marine genetic resources also bear value to humanity and to Indigenous and local communities, more specifically as a source of nutrition¹² and an important element in socio-cultural practices^{13,14}.

In light of their multiple benefits, marine genetic resources and related biotechnological applications are considered an important pillar of the 'blue economy', which aims to promote environmentally sustainable and equity-focused economic activities in the ocean¹⁵. Due to the richness of marine biodiversity in many places of the Global South, marine genetic resources

bear particularly great potential among various sectors of the blue economy in supporting more equitable use of ocean resources¹⁶. However, achieving benefits from marine genetic resources requires economic, scientific and technical capacities to collect and use the necessary materials and data¹⁷. At present, these capacities are highly unequally distributed at the global level^{18–23}. Capacities to conduct marine biological research, measured in terms of publication output, are dominated by developed countries^{22–24}. Developing countries have increased their capacities over the past years, but they are still comparatively low^{23,24}. This can also be observed in the number of vessels nationally available for scientific explorations²⁵. Global imbalances in capacity shape the distribution of intellectual property on marine genetic resources as well, as patent applications are highly concentrated in only a handful of countries and companies^{18–20}. New policies and practices of resource management will be needed for marine genetic resources to reach their full potential.

Capacity-building and benefit-sharing for developing countries have thus been central issues for biodiversity governance in recent years. Following international frameworks on plant and viral genetic resources under the Food and Agricultural Organization of the United Nations (FAO) and the World Health Organization (WHO), the Nagoya Protocol²⁶ to the Convention on Biological Diversity (CBD) was the first step to regulate marine genetic resources²⁷. Having entered into force in 2014, this protocol aims to prevent cases of inequitable extraction of terrestrial and marine

¹Research Platform Governance of Digital Practices, University of Vienna, Vienna, Austria. ²Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden. ✉e-mail: paul.dunshirn@univie.ac.at

genetic resources by obliging users to enter bilateral agreements with those countries in which they collect materials. In the context of a recent decision to develop access and benefit-sharing measures for genetic sequence data (also called ‘digital sequence information’—DSI)²⁸, Parties to the CBD are now considering how to extend or transform these existing arrangements²⁹. These discussions have carried over to negotiations on a legally binding treaty on the conservation and sustainable use of Biodiversity Beyond National Jurisdiction (BBNJ), which was adopted in June 2023. States contentiously negotiated for over five years on how benefits arising from the use of marine genetic resources should be shared multilaterally. While developing countries aimed for a treaty that draws on the so-called ‘common heritage of humankind principle’ with substantive provisions to alter inequities in access and abilities to conduct research on marine genetic resources, developed countries aimed for provisions more in line with the status quo, namely the ‘freedom of the High Seas’^{30–32}. Parties eventually agreed on a compromise that includes the sharing of scientific data, international research opportunities, and the regular redistribution of monetary benefits arising from the use of marine genetic resources to capacity-building and conservation projects through a purposed fund³³. However, no concrete provisions affecting intellectual property regimes were introduced³⁴ and heavy bureaucratic burdens on scientific researchers will likely be avoided³⁵. Many details of the capacity-building and benefit-sharing schemes under the BBNJ treaty and the CBD still need to be negotiated moving forward. In the meantime, there is an acute need for research on how such international frameworks may effectively increase the equitability and inclusivity of marine genetic research and related innovation practices.

The literature on genetic resource governance provides insights into mechanisms that, on the one hand, contribute to inequities in the global distribution of scientific and technical capacities and, on the other hand, could be leveraged to create more equity. A frequently highlighted point is the importance of infrastructure in enabling marine genetic research, such as sequencing equipment and other scientific tools³⁶, research vessels²⁵, funding³⁷, and data-sharing^{38,39}. Development of infrastructure can be supported by the above-mentioned biodiversity treaties and their mechanisms for capacity-building^{23,40} and benefit-sharing^{25,41–43}, as well as through other public or private investments. Another line of research focuses on social relations between scientists, highlighting how international collaboration choices can re-produce global imbalances as they favor countries that are already within the scientific elite²¹. Without concrete interventions in existing social networks, there are little opportunities for new relationships and practices to emerge that would strengthen the capacities of developing countries and their scientists^{21,44,45}. A third line of argumentation points to the importance of creating interactions between academic research and corporate activities to strengthen blue economies, including in developing countries^{17,46–48}. A few case studies^{49,50} provide insights into how such interactions may develop at different stages of typical marine bioprospecting pipelines⁵¹. Described cases include a bacteria isolated from the deep sea worm *Alvinella Pompejana*, which was collected by French researchers and later used to develop a face cream^{49,52}, and the ‘bengamide’ molecule, which U.S. American scientists found in a sea sponge from Fiji and which inspired anticancer clinical trials by Novartis^{50,53}. However, it remains unclear how insights from these studies generalize to interactions between marine genetic research and corporate activities beyond these cases. A separate body of literature in innovation studies has researched relations between scientific and corporate activities in more generalizable terms as knowledge flows between actors^{54–60}. However, this literature has not yet informed the discussions or the research on equity related to marine genetic resources.

To address this lack of systematic research on relations between marine genetic research and corporate activities, we identified 23,417 scientific papers associated with genetic sequences from 31,578 marine species in GenBank, the biggest repository of sequence data. By tracing knowledge flows from these studies, we established connections to a set of 7170 patents that cite these papers on their front page. Finally, we linked scientific papers

to provider countries and patents to user countries through a combination of available metadata and manual curation. We focus on the nation-state level as this has been the primary level of analysis in studies of global capacity in marine genetic research and innovation activities. Even though nation-states are not the immediate producers or users of knowledge flows, they can nevertheless be understood as actors themselves as they regulate, fund and benefit from national science and innovation systems⁶¹. Furthermore, they establish the conditions for accessing and using genetic resources within their jurisdiction and are entangled with the interests of affiliated companies and scientists^{62,63}.

Our analysis first reveals which countries contributed most to the publication of marine genetic studies. Second, it estimates the relevance of these studies for either follow-up scientific publications (from here on ‘secondary science’) or patents and compares this at the national level. Third, it explores which research content was most relevant to either secondary science or patents. Fourth, it examines how science and patenting activities balanced across countries, measured in terms of aggregate publication and patent counts, as well as via estimation of knowledge flow from science provider to patenting countries.

Results

Global imbalances characterize scientific capacity to sequence marine genetic resources

Figure 1 displays how many marine genetic studies researchers from economic groups and countries contributed to, either as lead or co-authors. Organisation for Economic Co-operation and Development (OECD) countries were responsible for around 77% of all papers. Brazil, Russia, India, China, and South Africa (BRICS) accounted for 16%, Group of 77 (G77) countries for 4%, and others for 2%. Researchers from the U.S., China, and Japan alone covered almost 50% of all papers. The most active G77 country was Thailand with 1%, and the biggest provider outside of the economic blocs was Taiwan with 2% of all publications (publication numbers for the top 20 countries are visible in Table S1).

To estimate the relevance of publications to follow-up scientific studies, we contrasted publication numbers with average citations by secondary science. The countries with the highest average citations by science among the top 20 were the United States, Germany, and Switzerland (Table S1). Countries like China and South Korea had many publications but relatively small average citations by science. We found that the number of publications was moderately correlated ($p < 0.05$, $r = 0.27$) with average citations by science, and not correlated with average citations by patents (Table S2). The data used for national-level analysis are available in Supplementary data file 3.

As relevant to science, as relevant to patents: comparing citation frequencies

We compared how often a country’s marine genetic research got cited by secondary science and by patents. We found a moderate positive and significant correlation ($p < 0.01$, $r = 0.36$), implying that if a country’s research is relevant for secondary science, it is also likely to be of some interest for patenting (Table S2; Fig. 2).

However, for some countries, the relevance of their research for patenting lay below the linear prediction. For instance, research that included authors from Panama (‘PA’) had many citations from secondary science, but close to no citations from patents. Countries whose research was similarly relevant to science would receive, on average, about one more patent citation per paper. This is a lot, considering that the country most relevant to patenting displayed here, Russia, only received around four citations by patents per scientific paper. Other countries with low relevance to patents were Austria (‘AT’), Czech Republic (‘CZ’), and Ireland (‘IE’). At the other end of the spectrum, countries above the trend were Russia (‘RU’), Iceland (‘IS’), and the Netherlands (‘NL’). These latter countries’ research got more cited in patents than would be expected from their relevance to secondary research.

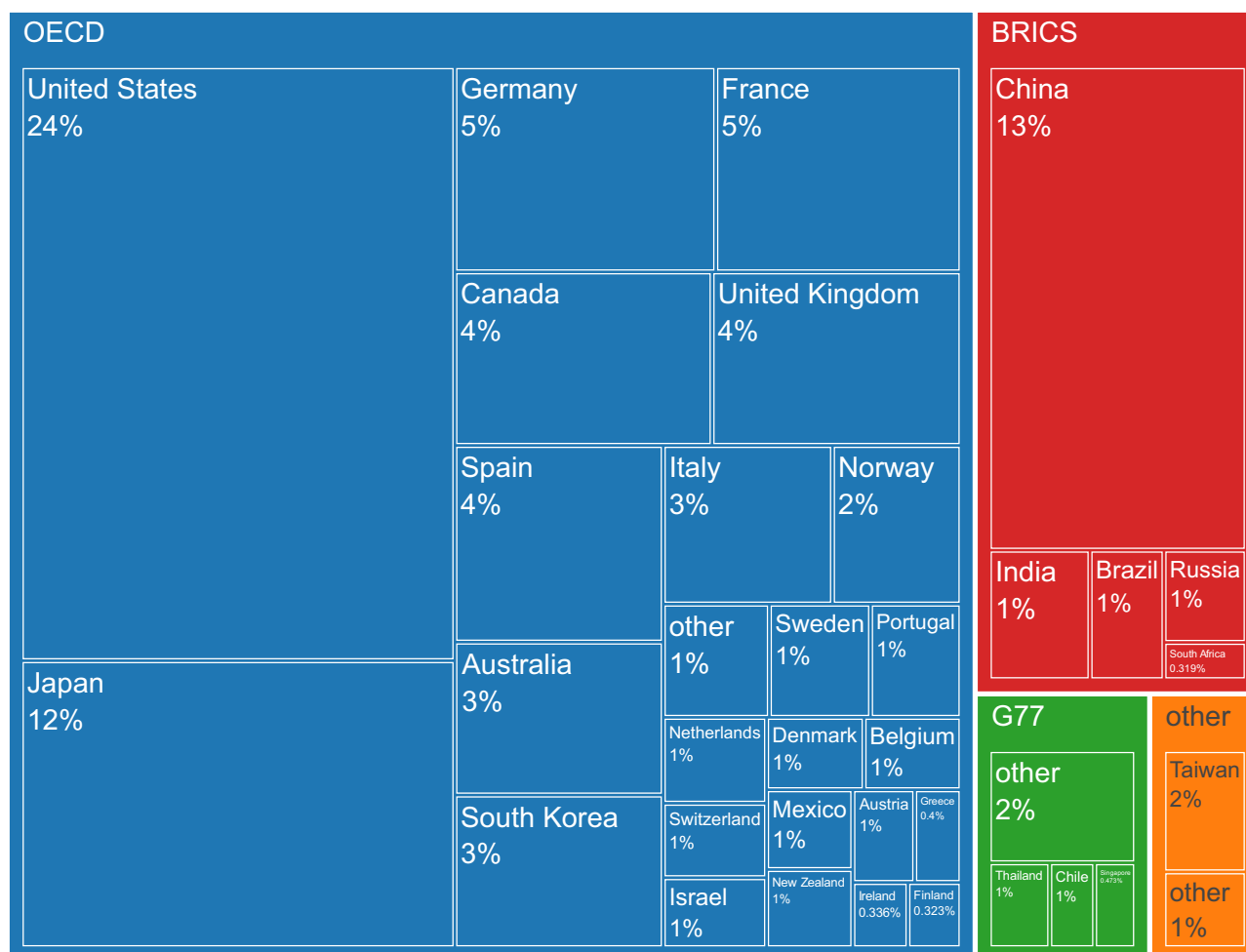


Fig. 1 | Marine genetic studies per country and economic group. Box sizes represent shares of total publications.

Semantic polarization between science and patenting

To explain why certain types of research may be more or less relevant to science or patenting, we constructed an indicator that ranks terms found in the titles of marine genetic studies based on their average citations from both scientific publications and patents (Fig. 3). The indicator weights the words by their frequency to reveal semantic leaning of commonly used words⁶⁴.

Many of the words with the highest relevance to patenting were based on one or two specific outlier publications that had received many patent references (see Fig. S1). The relevance of the term ‘fluoresc’ related to a publication titled ‘Fluorescent proteins from nonbioluminescent Anthozoa species’ by a team of Russian scientists⁶⁵, which had received over 300 patent citations. The significance of ‘human’ to patenting was based on ‘Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences’⁶⁶, published by Canadian and U.S. American researchers. ‘Desaturase’ is a type of enzyme, and the word was linked to two publications of high relevance to patenting: one from a team of researchers from the University of Hamburg and the company BASF that together had explored a marine diatom for cloning genes encoding fatty acid desaturases⁶⁷; the other from Canadian and U.S. American researchers that had also been interested in fatty acids and biosynthesis⁶⁸. Considering the words of high relevance to science, ‘genom’ and ‘evolut’ often appeared together in publications, such as the study ‘The genomic basis of adaptive evolution in threespine sticklebacks’⁶⁹, published by a multinational team.

Selection and analysis of exclusively relevant words for patenting revealed that many of them refer to specific marine organisms (such as ‘rhodothermus’, ‘tricornutum’, ‘tertiolecta’—indications of bacterial species names) or molecular biology terms (for instance, biomolecules like fatty

acids, serine, and desaturase). On the contrary, words that were exclusively relevant to scientific literature tend to be broader. Based on our selection criteria (see methods section), only 72 words were identified as more exclusive to patents in contrast with 1475 words more relevant to secondary science. For studies where at least one of the researchers was affiliated with a G77 country, the difference was similar: 10 and 177, respectively. The full list of publications used for this analysis is available in Supplementary data file 1 and a list of words that were most exclusively relevant to patents and science in Supplementary data file 2.

National science-patent balances

Next, we contrasted countries’ number of marine genetic studies to the number of granted patents citing those studies (Fig. 4a). The assignment of patent holders to countries was based on headquarters location (for companies) or institutional location (for universities or public institutions). Publication and patent counts for the top 20 countries can be found in Table S1 and for the top 20 institutions in Tables S3 and S4.

Figure 4a shows a strong and significant positive correlation ($p < 0.001$, $r = 0.84$) between the number of published studies and patents (Table S2). The more a country publishes the more likely it is to own patents that rely on marine genetic research. At the same time, the varying distances of each country to the linear prediction line show that some countries contributed relatively more in terms of science than they patented, and, vice versa, others patented more than they provided. Countries like Portugal (‘PT’), Mexico (‘MX’), and Italy (‘IT’) published a lot while not patenting a lot, while the United States (‘US’), the Netherlands (‘NL’), Denmark (‘DK’), and Switzerland (‘CH’) patented a lot compared to how much they published. The

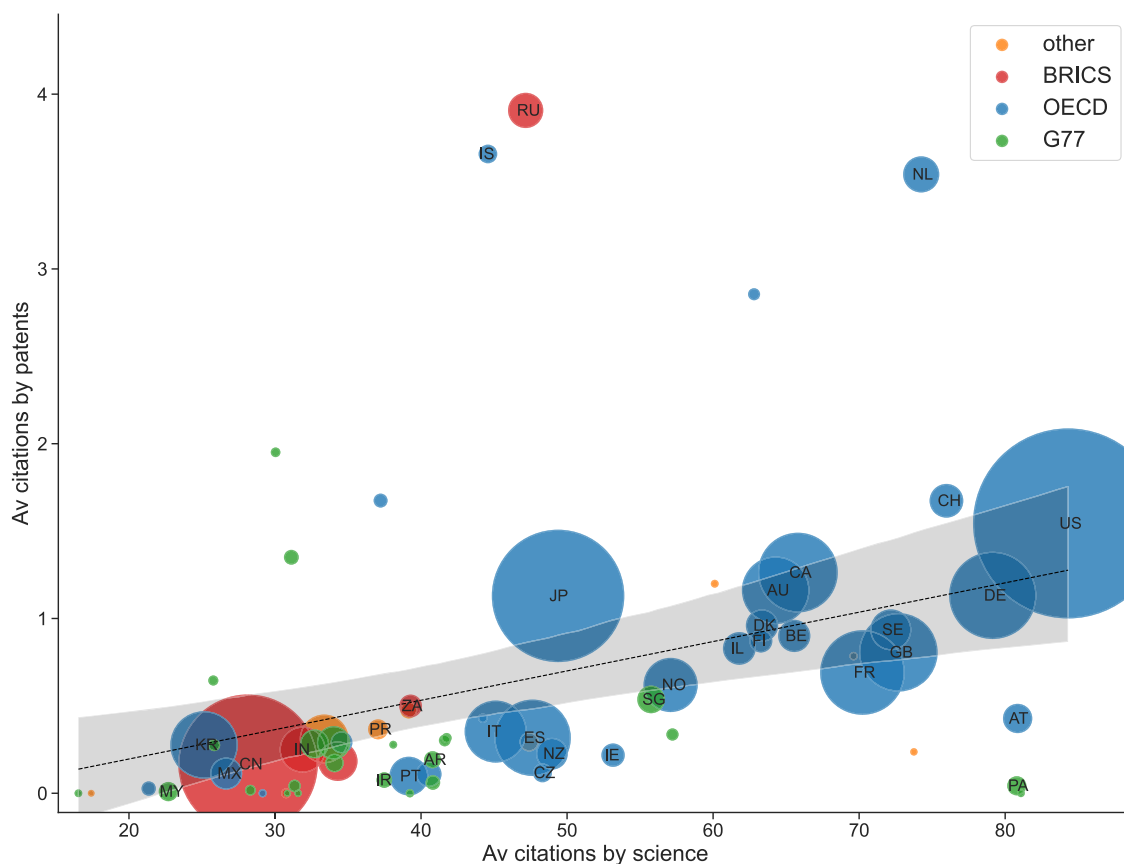


Fig. 2 | Relationship between average citations by secondary science and average citations by patents per country. Figure is based on countries with more than 5 publications ($n = 70$). Node size displays the number of publications and color displays economic group. 95% confidence intervals are displayed. Countries: US United States of America, AT Austria, PA Panama, DE Germany, CH Switzerland, NL Netherlands, GB United Kingdom of Great Britain and Northern Ireland, SE

Sweden, FR France, CA Canada, BE Belgium, AU Australia, DK Denmark, FI Finland, IL Israel, NO Norway, SG Singapore, IE Ireland, JP Japan, NZ New Zealand, RU Russian Federation, CZ Czechia, ES Spain, IS Iceland, IT Italy, AR Argentina, ZA South Africa, PT Portugal, IR Iran, PR Puerto Rico, IN India, CN China, MX Mexico, KR South Korea, MY Malaysia.

BRICS countries China ('CN') and Brazil ('BR') also fell far below the predicted number of granted patents, and countries like Thailand ('TH') or Greece ('GR') did not own any patents citing marine genetic studies. G77 countries were generally underrepresented in this plot, but a few exceptions included Singapore ('SG'), Chile ('CL'), and Saudi Arabia ('SA').

Finally, we illustrate the knowledge flow between science provider and user countries, measured via patent references to science (Fig. 4b). All top 10 users and 9 out of 10 biggest providers belonged to the OECD block, whereas the share of other coalitions was very modest. Among the two biggest providers of knowledge, Japan provided 2.8 times as much scientific knowledge to other countries as it used from them for patenting and the United States only contributed 0.6 times the amount it used from other countries (Table S1).

Discussion

This study established a relational perspective on the provision and use of marine genetic research by tracing how it becomes relevant to future scientific and corporate activities and by identifying actors involved in the process. We found research that employed molecular biological terms and bacterial species names to have attracted the most references from patents, unlike research on other scientifically relevant species. Nationally, our findings reveal the dominance of marine genetic research in OECD countries, while the G77 was largely absent. Yet, important differences existed even among OECD countries. Being the largest provider of scientific knowledge and owner of patents, the U.S. used more knowledge from other countries' science than it provided to them. These findings have important implications for international capacity-building and benefit-sharing.

Our analysis of national capacities to conduct marine genetic research resembles findings of previous studies on marine biological research capacities^{21–24}. Alongside publication counts, we considered how frequently a country's research received citations from either secondary science or patents and found a significant correlation between these indicators. Thus, if a country's research is considered valuable by other researchers, it would likely also be valuable for patenting. However, a few outliers like Russia and Iceland existed. A closer look at those two countries revealed that the deviation may have been caused by the work of two scientific groups that later became founders of highly successful biotechnological startups, Evrogen in Russia (<https://evrogen.com/about/about.shtml>) and Prokazyne in Iceland (<https://prokazyne.com/about-us/>). Other explanations may be found in research content (such as applied versus basic research), language⁷⁰, types of sequencing techniques (e.g., whole-genome sequencing papers may receive more scientific citations than from patents), or international collaboration patterns.

Our semantic analysis of marine genetic studies' titles revealed words polarized in their relevance either for patenting or for secondary science. While some terms of high scientific relevance, such as genomic ('genom') and evolutionary ('evolut'), related to basic research contexts, others may indicate potentials for future natural product discovery. This included terms like 'mykiss' (The rainbow trout *O. mykiss*), 'dinoflagel' (*Dinoflagellates*, a group of marine plankton), 'diatom' (large taxonomic group of microalgae), and 'ciona' (a genus of sea squirts), which referred to a specific taxonomic group of marine organisms. These latter terms were highly used by scientific articles that studied molecular adaptation or contributed with full genome sequence data yet did not receive much attention from patents. As genome

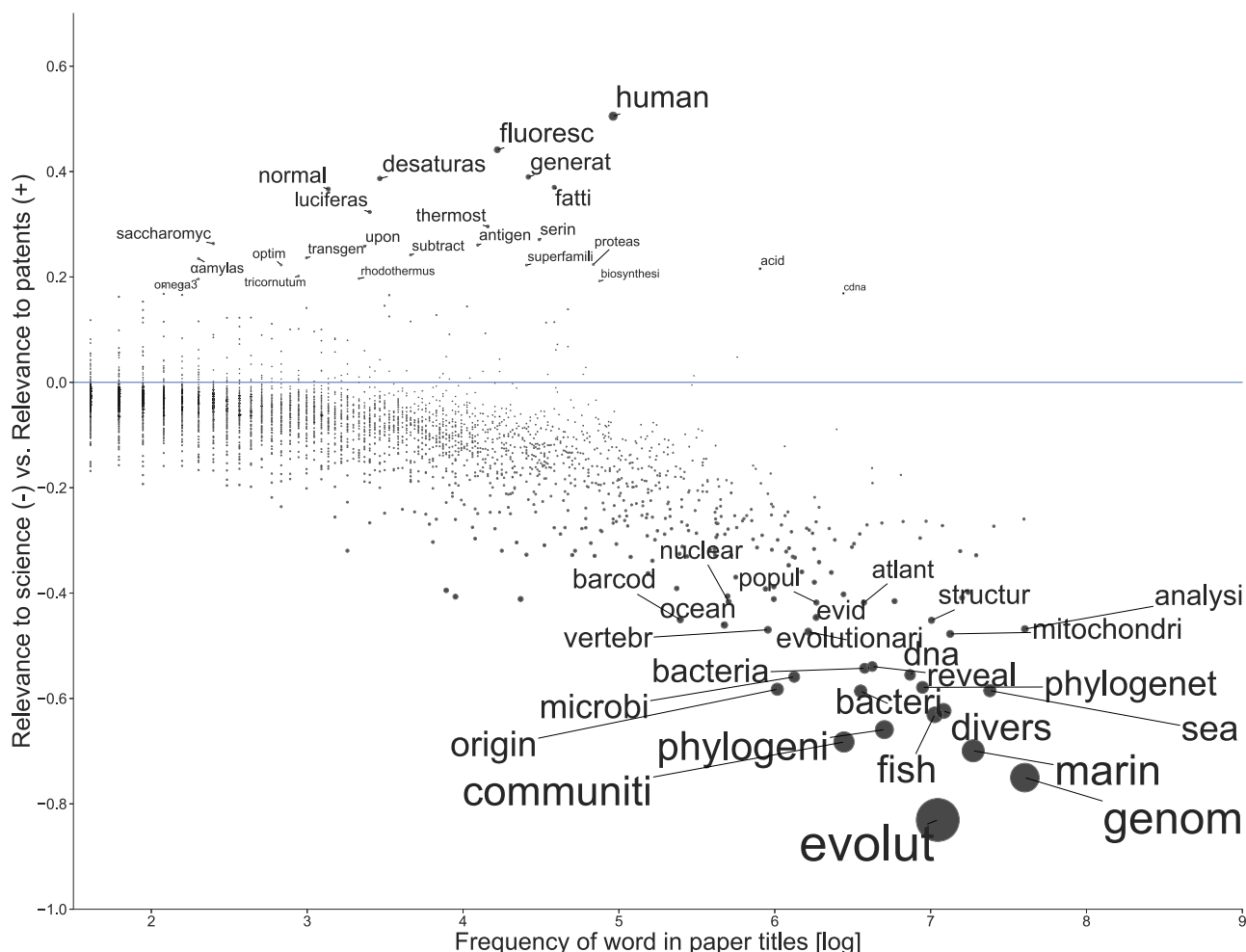


Fig. 3 | Relevance to science versus relevance to patents for words in paper titles. Node size corresponds to how polarized a word was according to the relevance indicator. The indicator ranges from -1 , indicating exclusive relevance to science, to

$+1$, indicating exclusive relevance to patents; 0 indicates that a word was equally relevant to both science and patents. Words above the 0.995 quantile and below the 0.005 quantile were labeled to highlight most polarized words.

mining has become a time and cost-efficient method for predicting novel biological enzymes⁷¹, these organisms may become relevant to patenting efforts. Words that were mostly relevant to patenting related most to bacterial organisms, such as ‘*rhodothermus*’ and ‘*tricornutum*’, and molecular biology terms and applications. For instance, the term ‘*fluoresc*’ points to research on fluorescent proteins, which have been a major source of marine biotechnological innovation¹.

The method of categorizing marine genetic research based on its relevance to different applications holds significance to international capacity-building initiatives and the broader design of research and innovation systems. By moving beyond abstract and unrefined concepts of research capacity in policy-making, we can identify concrete research content that is likely to lead to particular research or innovation and invest in local collaboration networks around corresponding scientific fields^{21,57}. This method could, for instance, be a useful guide to identify environmental and capacity-building projects that should be funded via the purposed multilateral fund under Article 52 of the BBNJ treaty⁷². Subsequent research may further refine these findings and employ a more differentiated scheme of application than the simple distinction between patents and secondary science employed here. For example, it might be reasonable to single out scientific research or technologies that demonstrate a high potential in achieving the Sustainable Development Goals, such as microbial⁷³ or microalgal biotechnology⁷⁴, and explore which semantic content they are related to. While we are currently not

aware of any research that evaluates applications of marine genetic resources in terms of sustainability, planned future research should provide concrete ideas in this regard.

After examining citation flows from science to the actual patent documents to understand how much marine genetic research each country ‘contributed’ and ‘used’, in the sense of referencing it as prior art in patent applications. We found that scientific capacity corresponded to more intellectual property on the basis of this research. The link between national science and patenting activities might, to some extent, be related to the central role of universities in patenting and commercializing research findings^{58,75,76}. Another plausible explanation is knowledge flows from research to corporate activities that are not directly channeled via university researchers yet still geographically bounded⁵⁷. Looking more concretely at the role of countries in the provider-user relationship, countries such as the U.S., the Netherlands, or Switzerland owned considerably more patents than expected based on the number of published studies, and, vice versa, other countries owned fewer than expected (Fig. 4a). Overall, OECD countries were by far the biggest producers of marine genetic research and owned the most patents referencing this science. The predominant role of OECD countries was even stronger when looking at knowledge flows between provider and user countries, in which the U.S. and Japan acted as the biggest knowledge providers (Fig. 4b). However, when disregarding inner-country knowledge flows, we discovered that the U.S., Australia, Great Britain, and others provided less knowledge to other countries than

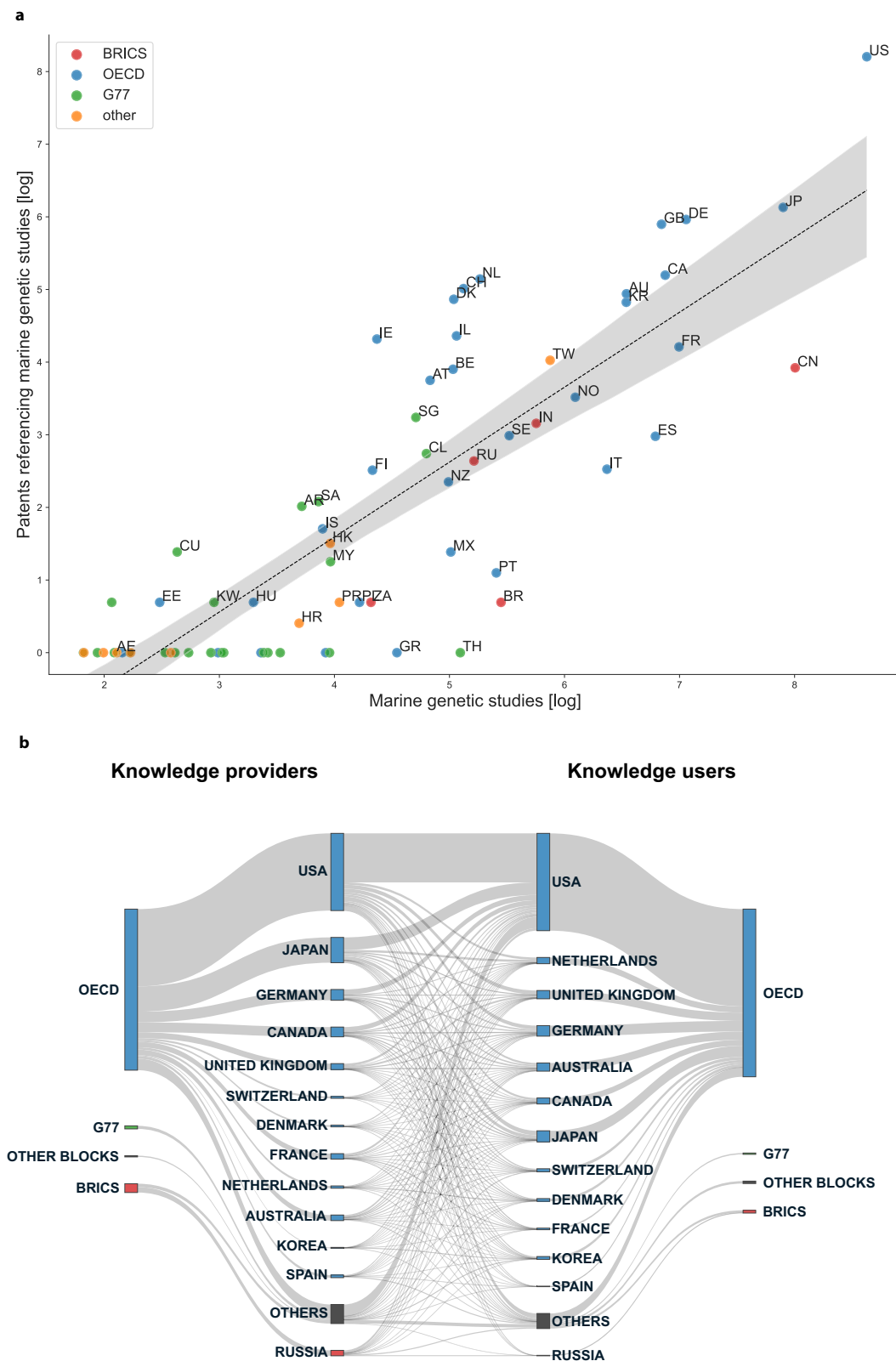


Fig. 4 | National science-patent relations. **a** Comparing counts of marine genetic studies and patents for countries with more than 5 publications ($N = 70$). Counts were transformed using $\log(x + 1)$ and $\log(y + 1)$ to avoid hiding countries without patents. A 95% confidence interval is displayed. Countries: US United States of America, CN China, JP Japan, DE Germany, FR France, CA Canada, GB United Kingdom of Great Britain and Northern Ireland, ES Spain, KR South Korea, AU Australia, IT Italy, NO Norway, TW Taiwan, IN India, SE Sweden, BR Brazil, PT

Portugal, NL Netherlands, CH Switzerland, RU Russian Federation, TH Thailand, MX Mexico, BE Belgium, IL Israel, NZ New Zealand, DK Denmark, AT Austria, CL Chile, SG Singapore, GR Greece, IE Ireland, FI Finland, ZA South Africa, PL Poland, PR Puerto Rico, MY Malaysia, HK Hong Kong, IS Iceland, SA Saudi Arabia, AR Argentina, HR Croatia, HU Hungary, KW Kuwait, CU Cuba, EE Estonia, AE United Arab Emirates. **b** Knowledge flows from provider countries to user countries grouped by economic blocks. Top 10 countries in each category are labeled.

they used, while Japan, Russia, France, and Spain provided distinctively more than they used (Table S1). Despite their high dependence on marine ecosystems⁷⁷, G77 countries were nearly absent in marine genetic research that contributed to technological innovations, which urges the need for broader and more inclusive scientific collaborations.

Considering the status of genetic research as a public good that, at least in terms of the genetic sequence data, is openly accessible and usable to anyone^{78,79}, these divergences raise the question of whether we are observing a freeriding problem: some countries ‘use’ a lot of research for their innovations without contributing much to the availability of that knowledge. The freeriding, in this case, does not necessarily occur via the subtractability of the scientific knowledge resource due to overuse, as referencing a study in a patent does not directly diminish another user’s ability to do the same. However, in mere economic terms, patents can lead to monetary benefits for their owners, and science creates non-monetary value in the form of public knowledge, arguably constituting a provider-user relationship. Such a perspective extends research on genetic resource user-provider balances, which has focused on relations between countries of origin (i.e., countries where genetic materials were collected) and scientific user countries^{80,81}. In contrast, our research revealed user-provider relations in the later stages of the marine genetic resource pipeline, namely in the flow of knowledge from science to patenting. It might be important to consider this relationship in benefit-sharing arrangements under the CBD and the BBNJ treaty. One of the options for assessing monetary contributions under the BBNJ treaty is to rely on ‘indicators measuring the aggregate level of activities by a Party’, as stated in paragraph 14.8., not necessarily requiring individual tracing of activities^{29,32,82}. Accounting for science-patent balances, as we did here, could be a way to calculate contributions: countries that patent a lot, but contribute little research would have to pay according to how much above the linear trend they lie. The introduction of a ‘BBNJ standardized batch identifier’ (Article 12 of the BBNJ treaty) will likely allow future calculations to specify whether used resources are of BBNJ origin. Similar discussions on benefit-sharing modalities are currently taking place in response to the recent CBD decision to develop a multilateral benefit-sharing system on DSI²⁸.

In conclusion, our study hopes to provide new impetus to discussions about how to make marine genetic research and related patenting activities more equitable. While we found scientific capacities to be generally concentrated in developed countries, our perspective on science-patent relations revealed that, even among the OECD, countries may act either as net providers or net users of marine genetic research. In order to build an equitable scientific knowledge commons around marine genetic resources, it may thus be important for international frameworks, such as the BBNJ treaty or the CBD, to consider how to recognize scientists in their crucial role as providers of knowledge, while putting responsibility on those patenting actors who benefit more from such research than they provide. Benefit-sharing schemes may be designed accordingly and, in connection with capacity-building measures, may allow for gradually integrating a wider community of users, making marine genetic resources an integral part of sustainable and equitable ocean economies.

Methods

Identifying marine genetic studies

It is not a straightforward task to define marine genetic resources. While certain organisms live exclusively in the sea, there are many cases of organisms that traverse marine, freshwater, and terrestrial ecosystems. This is often the case for bacteria but may also include animals like sea birds. Various biological databases may be used to identify marine species, including the World Register of Marine Species (WORMS), the Ocean Biodiversity Information System (OBIS), and the Global Biodiversity Information Facility (GBIF)²². For the purpose of identifying marine species that receive patenting interest, OBIS is of limited value as it underrepresents potential marine bacteria that attract a lot of interest²². We thus used a combination of 130,973 marine species identified via OBIS and a manually curated list of additional 1270 patented marine species from a recent study on marine bioprospecting²⁰, which followed a methodology described in an

earlier study¹⁹. The hereby selected 132,243 marine species constituted the starting point for the curation of the scientific publication and patent datasets for our study. We did not distinguish between species from areas beyond national jurisdiction and within national jurisdiction—our study addresses both legal contexts.

We identified relevant marine genetic studies by querying the species names in GenBank for corresponding genetic sequence data via Python scripts. GenBank is the main U.S. American public genetic sequence database. As it is part of the International Nucleotide Sequence Data Collaboration (INSDC), it mirrors the two other largest databases, the European Molecular Biology Laboratory (EMBL) and the DNA Data Bank of Japan (DDBJ), daily, constituting by far the biggest pool of genetic sequence data available. We chose to work with sequence data rather than samples (which are ‘closer’ to the notion of genetic material⁸³), as GenBank’s Bio Sample data are often not linked to readily identifiable publications⁸¹ and thus not as useful to our approach. Within the broad term genetic sequence data, we chose nucleotide sequences (DNA and RNA) as they lie at the core of the definition of digital sequence data in relevant policy debates⁸⁴. Previous analyses on genetic sequence data user-provider relations in the CBD context took the same methodological choice^{80,81}. Additionally, other forms of genetic sequence data, like proteins, are usually derived from nucleotide sequences, so nucleotide sequences are a useful starting point for analysis⁸¹. When searching GenBank, we identified 31,578 species that had at least one sequence entry. Overall, we found 36,849,850 sequence entries, out of which 10,294,029 were linked to a total of 23,438 PubMed identifiers. Due to the high volume of data, searches were performed for multiple weeks between 23rd of February and 8th of May 2023.

To get standardized and curated bibliometric data on those publications, we queried the PubMed identifiers via the scholarly API of the database The Lens (www.lens.org). 15 out of 23,438 PubMed identifiers turned out to be non-existent after querying on The Lens, and 5 publications did not list any authors. As such, the analysis proceeded with 23,417 publications.

Mapping scientific publication counts and citation frequencies to countries

We mapped scientific publications onto countries based on The Lens’ country classification, which was usually the location of the university. In some cases, no country classifications were available, but institutional names were, for which we tried re-constructing them using a set of manually defined rules. For instance, we applied a rule that any given author affiliation that entails the strings ‘univ’ and ‘harvard’ would be labeled as U.S. American, or if it entails ‘tokyo’, it would be labeled as Japanese. We also searched for country names entailed in the affiliation strings (e.g., ‘NRC-Biotechnology Research Institute, 6100 Royalmount Ave., Montreal, Que., Canada’) to label the cases accordingly. We tested each rule and adjusted them to prevent false positives. For cases where country labels were still missing after applying these rules, we looked them up and added them manually for all institutions that registered more than 5 publications. Remaining unlabeled affiliations that were either unclear (e.g., ‘Marine Science Institute’) or with less than 5 publications were not considered for any analysis displaying country-level information. Out of the 23,417, we connected 23,231 publications to at least one affiliation country via this method.

While some studies on genetic resources limited their observations either to first authors⁸⁰ or first and last authors²³ to assess country affiliations, we followed a study that considered all authors⁸⁵. This made sense as we were interested in all providers of scientific knowledge, even if not listed as lead authors. Including non-leading authors may matter in scenarios of North-South collaboration, such as joint research cruises led by a country of the Global North or when genetic materials were sourced by researchers from the Global South but analyzed by researchers from the Global North. In any case, we compared the distribution of national publication counts if constructed either based on first, last, or all authors and found that they are all strongly correlated across the authorship types. The same was the case for

citation frequencies. This means that the choice is unlikely to have a big impact on country-level trends. However, working with all authors instead of the first or last authors results in a much larger set of countries (149 instead of 104, which would be the case for first-author countries), though none of these added countries registered more than 9 publications.

Working with all authors requires consideration of how much each authorship ‘adds’ to the national affiliation counts. We adopted a fractional scoring approach⁸⁵ for this purpose. If a publication had one author from France and one from China, 0.5 would be added to the publication counts of both countries. If it had two authors from France and one from China, the same scores would apply. If a country had five authors from the U.S., it would be counted as one publication. This way of counting prevented overweighting publications that included authors from a very large number of countries, which was not unlikely.

To identify instances of knowledge flow from marine genetic studies to subsequent research and innovation activities, we retrieved citations to the studies from both patents and secondary research publications via The Lens’ scholarly API and mapped them onto the countries (single-country and multi-country publications were treated according to the same logic as described above).

For the correlation table (Table S2) and Figs. 2 and 4a, we further reduced our sample to 70 countries with more than 5 publications to prevent outliers from skewing overall trends. Using median values instead of means was not a better solution, because there was too little variation in the citation rates from patents.

Semantic analysis

The semantic analysis was based on a set of all words present in the titles of the marine genetic studies. The words were processed as lower cases. Using the Python package NLTK’s stop word list for English, we removed words with little semantic content, like ‘for’ or ‘that’. We also stemmed the words using NLTK’s snowball stemmer, which reduced the heterogeneity of the words, e.g., by joining ‘genome’ and ‘genomic’ to ‘genom’. This resulted in a set of 29,000 uniquely occurring terms. After some testing, we also removed words that occurred in less than five or more than 2500 publications, as it would have been hard to interpret citation rates for such infrequent or overly frequent words. This reduced the corpus to 5181 terms.

To construct the indicator that classifies the processed terms in their relevance to science as compared to relevance for patents, we iterated through each term and calculated how often a publication entailing the term got cited on average. We excluded words higher than the 0.995 quantile of most frequently cited words for both scores to prevent outliers, which further reduced the list of terms to 5154. We weighted the average scores by multiplying values with the square root of the terms’ frequency, in order to foreground terms that were polarized and occurred frequently⁸⁴. We then normalized the score for average citations by science to correspond to the same scale as average citations by patents, as there were generally way more citations from science than from patents and calculated the difference between the corresponding values of the two variables. This difference was scaled and transformed into an orientation value within the range of −1 (exclusively relevant to science) to 1 (exclusively relevant to patents). For all words with orientation values lower than zero (more relevant to science), we selected words lower than the 0.3 quantile and for those with orientation values greater than zero (more relevant to patents) those higher than the 0.7 quantile as the most exclusive words to science and patents respectively. We analyzed the publications linked to the selected words in more detail to interpret why they have achieved relevance for either of the domains. We then repeated the same selection for scientific papers that were co-authored by at least one researcher from a G77 country.

Tracing patents that cite marine genetic studies and mapping patent owners to countries

To construct national science-patent balances, we used The Lens’ patent API to collect data on the patents that had cited the identified publications. We found that 3247 out of the overall 23,417 publications were referenced

by a total of 7170 granted patents. We also found 5243 patent applications but did not analyze these further, as it was unclear whether they actually resulted in ownership status.

Patent citations to science are a common indicator in economics to study the sources and nature of innovation in firms^{54–56,86} or to study science-patent linkages in countries²⁴. Researchers have evaluated this method through survey research⁸⁶, concluding that it is useful for studying knowledge flows from publicly available science while missing out on relevant other sources of privately communicated scientific knowledge. Patent citations to science achieve a degree of reliability as national patent laws usually require some indication of the relevant prior art, which includes previously published science. While applicants in the U.S. are required to disclose prior art themselves, references to science in non-U.S. jurisdictions are usually added by patent examiners⁵⁶. Even though patent practices differ across jurisdictions, we decided not to distinguish between them, as either type indicates knowledge flow from science to patents.

Assigning patents to countries is more complicated than it is for scientific publications. Corporate entities tend to be more dynamic than scientific institutions, as companies may merge with other companies, or they may split off sections from a parent company, yet maintaining subsidiary status⁸⁷. Additionally, companies can be multinational and run multiple subsidiary branches in various countries⁸⁷. This is visible in patent data, which usually shows the owner’s addresses: Not infrequently does a patent link to addresses across several countries for one and the same company with multiple subsidiaries listed. For our purposes, it was most reasonable to identify the country of a corporate actor’s headquarters, rather than considering multiple countries per company.

We thus applied the following rules for linking each corporate entity to one country: First, we identified the parent company or university of a given corporate entity detected as a patent owner. We did this via desk research. Where we detected subsidiaries, we merged them into their parent entity. Second, we checked which country affiliations The Lens returned for the merged entities. In cases where all country affiliations matched, we considered these as a proxy for the headquarters location. If they did not match, we manually looked up the location of the entity’s headquarters. We repeated this process for each entity that we found to own more than five patents. For entities below this number, we also applied the second set of rules for country addresses but, due to time constraints, without first checking whether they needed to be merged with another entity. We excluded individuals, uninterpretable entities, and entities where we could not find sufficient information. This resulted in a reduced set of 7150 granted patents, out of which 6434 were connected to at least one country. We compared the hereby assembled national patent counts with scientific publication counts for Fig. 4a.

To measure knowledge provision and use patterns for Fig. 4b, we configured our data in a manner that tied provider countries and patenting countries together via each scientific publication cited by patents of the respective country (Supplementary data file 4). The fractional counting approach needed to be considered in the calculation. If a publication has been published by exclusively Japan-based researchers and cited by five exclusively Japanese-owned patents and one exclusively Swizz-owned patent, it would result in a Japan-Japan knowledge flow of 5 and a Japan-Switzerland flow of 1. If a publication was published by Dutch and German researchers and cited by four exclusively U.S. American-owned patents and four mixed U.S. American- and South Korean-owned patents, this would result in the following flows: NL-US (3); DE-US (3); NL-KR (1); DE-KR (1). The measured flow total thus corresponds to the number of granted patents per case. We repeated this counting procedure for each cited publication and aggregated the provider-user flows. Finally, we constructed external provision-use balances for each country by dividing the amount of used knowledge flow from other countries by the amount of knowledge provided to other countries. Self-loops (e.g., U.S. American research that got cited by U.S. American companies) were thus not considered in this equation. The resulting balances are visible for the top 20 countries in Table S1 and in Supplementary data file 3.

Data availability

All data needed to evaluate findings are present in the paper and/or the Supplementary files. Data on identified genetic sequences and marine species can be shared upon request. Analysis scripts are available at https://github.com/pduns/Conducting_marine_genetic_research_for_whom.

Received: 8 August 2023; Accepted: 20 September 2024;

Published online: 08 October 2024

References

- Blasiak, R. et al. The ocean genome and future prospects for conservation and equity. *Nat. Sustain* **3**, 588–596 (2020).
- Martins, A., Vieira, H., Gaspar, H. & Santos, S. Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Mar. Drugs* **12**, 1066–1101 (2014).
- Tortorella, E. et al. Antibiotics from deep-sea microorganisms: current discoveries and perspectives. *Mar. Drugs* **16**, 355 (2018).
- Wilkes, R. A. & Aristilde, L. Degradation and metabolism of synthetic plastics and associated products by *Pseudomonas* sp.: capabilities and challenges. *J. Appl. Microbiol.* **123**, 582–593 (2017).
- Hurst, D., Børresen, T., Almesjö, L., Raedemaeker, F. & Bergseth, S. *Marine Biotechnology Strategic Research and Innovation Roadmap—Insights to the Future Direction of European Marine Biotechnology* (Marine Biotechnology ERA-NET, 2016).
- Stange, M., Barrett, R. D. H. & Hendry, A. P. The importance of genomic variation for biodiversity, ecosystems and people. *Nat. Rev. Genet.* **22**, 89–105 (2021).
- Schindler, D. E. et al. Population diversity and the portfolio effect in an exploited species. *Nature* **465**, 609–612 (2010).
- Webster, M. S. et al. Who should pick the winners of climate change? *Trends Ecol. Evol.* **32**, 167–173 (2017).
- Collins, S. Many possible worlds: expanding the ecological scenarios in experimental evolution. *Evol. Biol.* **38**, 3–14 (2011).
- Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R. & Voolstra, C. R. Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nat. Commun.* **8**, 14213 (2017).
- Sjöqvist, C. O. & Kremp, A. Genetic diversity affects ecological performance and stress response of marine diatom populations. *ISME J.* **10**, 2755–2766 (2016).
- Johnson, B. M., Kemp, B. M. & Thorgaard, G. H. Increased mitochondrial DNA diversity in ancient Columbia River basin Chinook salmon *Oncorhynchus tshawytscha*. *PLoS ONE* **13**, e0190059 (2018).
- Vierros, M. K. et al. Considering Indigenous Peoples and local communities in governance of the global ocean commons. *Mar. Policy* **119**, 104039 (2020).
- Wynberg, R. Biopiracy: crying wolf or a lever for equity and conservation? *Res. Policy* **52**, 104674 (2023).
- Bennett, N. J. et al. Towards a sustainable and equitable blue economy. *Nat. Sustain.* **2**, 991–993 (2019).
- Cisneros-Montemayor, A. M. et al. Enabling conditions for an equitable and sustainable blue economy. *Nature* **591**, 396–401 (2021).
- Blasiak, R. et al. Making marine biotechnology work for people and nature. *Nat. Ecol. Evol.* **7**, 482–485 (2023).
- Arnaud-Haond, S., Arrieta, J. M. & Duarte, C. M. Marine biodiversity and gene patents. *Science* **331**, 1521–1522 (2011).
- Blasiak, R., Jouffray, J.-B., Wabnitz, C. C. C., Sundström, E. & Österblom, H. Corporate control and global governance of marine genetic resources. *Sci. Adv.* **4**, eaar5237 (2018).
- Zhivkopoulos, E., Jouffray, J.-B., Dunshirn, P., Pranindita, A. & Blasiak, R. Growing prominence of deep-sea life in marine bioprospecting. *Nat. Sustain.* **7**, 1027–1037 (2024).
- Tolochko, P. & Vadrot, A. B. M. The usual suspects? Distribution of collaboration capital in marine biodiversity research. *Mar. Policy* **124**, 104318 (2021).
- Oldham, P. et al. *Valuing the Deep: Marine Genetic Resources in Areas Beyond National Jurisdiction*. <https://bookdown.org/poldham/valuingthedeep/> (2014).
- Leal, M. C. et al. Fifty years of capacity building in the search for new marine natural products. *Proc. Natl Acad. Sci. USA* **117**, 24165–24172 (2020).
- Lara-Lopez, A., Valdés, L., de Pinho, R. & Enevoldsen, H. Analysis of ocean science production and impact in *Global Ocean Science Report 2020: Charting Capacity for Ocean Sustainability*. <https://doi.org/10.18356/9789216040048> (United Nations, 2020).
- Sierra-Correa, P. C. et al. Research capacity and infrastructure in *Global Ocean Science Report 2020: Charting Capacity for Ocean Sustainability*. <https://doi.org/10.18356/9789216040048> (United Nations, 2020).
- Convention on Biological Diversity. *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from Their Utilization to the Convention on Biological Diversity (Text and Annex)* (Secretariat of the Convention on Biological Diversity, 2011).
- Rabitz, F. *The Global Governance of Genetic Resources: Institutional Change and Structural Constraints* (Routledge, 2017).
- Convention on Biological Diversity. *COP Decision 15/19 on Digital Sequence Information on Genetic Resources* (Secretariat of the Convention on Biological Diversity, 2022).
- Scholz, A. H. et al. Multilateral benefit-sharing from digital sequence information will support both science and biodiversity conservation. *Nat. Commun.* **13**, 1086 (2022).
- Vadrot, A. B. M. Multilateralism as a ‘site’ of struggle over environmental knowledge: the North-South divide. *Crit. Policy Stud.* **14**, 233–245 (2020).
- Marciniak, K. The legal status of marine genetic resources in the context of BBNJ negotiations: diverse legal regimes and related problems in *New Knowledge and Changing Circumstances in the Law of the Sea* (ed. Heidar, T.) 40–64 (Brill, 2020).
- Oldham, P., Chiarolla, C. & Thambisetty, S. *Digital Sequence Information in the UN High Seas Treaty: Insights from the Global Biodiversity Framework-related Decisions*. SSRN Scholarly Paper. <https://doi.org/10.2139/ssrn.4343130> (2023).
- Deasy, K. What we know about the new High Seas Treaty. *NPJ Ocean Sustain.* **2**, 1–3 (2023).
- Mendenhall, E., Tiller, R. & Nyman, E. The ship has reached the shore: the final session of the ‘Biodiversity Beyond National Jurisdiction’ negotiations. *Mar. Policy* **155**, 105686 (2023).
- Scholz, A. H., Humphries, F., Vanagt, T. & Jaspars, M. *A New Dawn for Global Benefit-sharing: Capitalizing on the Global Biodiversity Framework for Marine Genetic Resources from Areas Beyond National Jurisdiction* (ICUN, 2023). https://www.iucn.org/sites/default/files/2023-02/bbnj_icg5bis_policy_brief_global_benefit_sharing_1.pdf.
- Capotondi, A. et al. Observational needs supporting marine ecosystems modeling and forecasting: from the global ocean to regional and coastal systems. *Front. Mar. Sci.* **6**, 623 (2019).
- Jolly, C. et al. *Funding for Ocean Science*. <https://www.un-ilibrary.org/content/books/9789216040048c008> (UN, 2021).
- Muller-Karger, F. E. et al. Marine Life 2030: building global knowledge of marine life for local action in the Ocean Decade. *ICES J. Mar. Sci.* **80**, 355–357 (2023).
- Rabone, M. et al. Access to marine genetic resources (MGR): raising awareness of best-practice through a new agreement for biodiversity beyond National Jurisdiction (BBNJ). *Front. Mar. Sci.* **6**, 520 (2019).
- Harden-Davies, H. et al. First to finish, what comes next? Putting capacity building and the transfer of marine technology under the BBNJ Agreement into practice. *NPJ Ocean Sustain.* **3**, 1–5 (2024).
- Aubry, S. et al. Bringing access and benefit sharing into the digital age. *Plants, People, Planet* **4**, 5–12 (2022).

42. Kreiken, B. E. & Arts, B. J. M. Disruptive data: how access and benefit-sharing discourses structured ideas and decisions during the Convention on Biological Diversity negotiations over digital sequence information from 2016 to 2022. *Glob. Environ. Change* **87**, 102892 (2024).
43. Broggiato, A. et al. Mare Geneticum: Balancing Governance of Marine Genetic Resources in International Waters. *Int. J. Mar. Coast. Law* **33**, 3–33 (2018).
44. Faure, M. C., Munung, N. S., Ntusi, N. A. B., Pratt, B. & de Vries, J. Considering equity in global health collaborations: a qualitative study on experiences of equity. *PLoS ONE* **16**, e0258286 (2021).
45. Amon, D. J., Filander, Z., Harris, L. & Harden-Davies, H. Safe working environments are key to improving inclusion in open-ocean, deep-ocean, and high-seas science. *Mar. Policy* **137**, 104947 (2022).
46. Sigwart, J. D., Blasiak, R., Jaspars, M., Jouffray, J.-B. & Tasdemir, D. Unlocking the potential of marine biodiversity. *Nat. Prod. Rep.* **38**, 1235–1242 (2021).
47. Collins, J. E. et al. Inclusive innovation: enhancing global participation in and benefit sharing linked to the utilization of marine genetic resources from areas beyond national jurisdiction. *Mar. Policy* **109**, 103696 (2019).
48. Holland, C., McCarthy, A., Ferri, P. & Shapira, P. Innovation intermediaries at the convergence of digital technologies, sustainability, and governance: a case study of AI-enabled engineering biology. *Technovation* **129**, 102875 (2024).
49. Humphries, F., Rabone, M. & Jaspars, M. Traceability approaches for marine genetic resources under the proposed ocean (BBNJ) treaty. *Front. Mar. Sci.* **8**, 661313 (2021).
50. Jaspars, M., Humphries, F. & Rabone, M. *Tracing Options for Marine Genetic Resources from within National Jurisdictions* <https://eprints.qut.edu.au/227075/1/102475803.pdf> (2021).
51. Jaspars, M. et al. The marine biodiversity pipeline and ocean medicines of tomorrow. *J. Mar. Biol. Assoc. UK* **96**, 151–158 (2016).
52. Cambon-Bonavita, M.-A., Raguénès, G., Jean, J., Vincent, P. & Guezennec, J. A novel polymer produced by a bacterium isolated from a deep-sea hydrothermal vent polychaete annelid. *J. Appl. Microbiol.* **93**, 310–315 (2002).
53. White, K. N., Tenney, K. & Crews, P. The bengamides: a mini-review of natural sources, analogues, biological properties, biosynthetic origins, and future prospects. *J. Nat. Prod.* **80**, 740–755 (2017).
54. Jaffe, A. B. & de Rassenfosse, G. Patent citation data in social science research: overview and best practices. *J. Assoc. Inf. Sci. Technol.* **68**, 1360–1374 (2017).
55. Hammarfelt, B. Linking science to technology: the “patent paper citation” and the rise of patentometrics in the 1980s. *J. Doc.* **77**, 1413–1429 (2021).
56. Marx, M. & Fuegi, A. Reliance on science: worldwide front-page patent citations to scientific articles. *Strateg. Manag. J.* **41**, 1572–1594 (2020).
57. Belenzon, S. & Schankerman, M. Spreading the word: geography, policy, and knowledge spillovers. *Rev. Econ. Stat.* **95**, 884–903 (2013).
58. Arora, A., Belenzon, S. & Pataconi, A. The decline of science in corporate R&D. *Strateg. Manag. J.* **39**, 3–32 (2018).
59. Jaffe, A. B. & Trajtenberg, M. Flows of knowledge from universities and federal laboratories: modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proc. Natl Acad. Sci. USA* **93**, 12671–12677 (1996).
60. Jaffe, A. B., Trajtenberg, M. & Henderson, R. Geographic localization of knowledge spillovers as evidenced by patent citations. *Q. J. Econ.* **108**, 577–598 (1993).
61. Görg, C. & Brand, U. Global environmental politics and competition between nation-states: on the regulation of biological diversity. *Rev. Int. Political Econ.* **7**, 371–398 (2000).
62. Schadeberg, A., Kraan, M., Groeneveld, R., Trilling, D. & Bush, S. Science governs the future of the mesopelagic zone. *NPJ Ocean Sustain.* **2**, 1–9 (2023).
63. Vadrot, A. B. M. et al. Towards a reflexive, policy-relevant and engaged ocean science for the UN decade: a social science research agenda. *Earth Syst. Gov.* **14**, 100150 (2022).
64. Monroe, B. L., Colaresi, M. P. & Quinn, K. M. Fightin’ words: lexical feature selection and evaluation for identifying the content of political conflict. *Polit. Anal.* **16**, 372–403 (2008).
65. Matz, M. V. et al. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**, 969–973 (1999).
66. Mammalian Gene Collection (MGC) Program Team. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl Acad. Sci. USA* **99**, 16899–16903 (2002).
67. Domergue, F., Lerchl, J., Zähringer, U. & Heinz, E. Cloning and functional characterization of *Phaeodactylum tricornutum* front-end desaturases involved in eicosapentaenoic acid biosynthesis. *Eur. J. Biochem.* **269**, 4105–4113 (2002).
68. Qiu, X., Hong, H. & MacKenzie, S. L. Identification of a $\Delta 4$ fatty acid desaturase from *Thraustochytrium* sp. involved in the biosynthesis of docosahexanoic acid by heterologous expression in *Saccharomyces cerevisiae* and *Brassica juncea*. *J. Biol. Chem.* **276**, 31561–31566 (2001).
69. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
70. Amano, T. et al. The manifold costs of being a non-native English speaker in science. *PLoS Biol.* **21**, e3002184 (2023).
71. Kamble, A., Srinivasan, S. & Singh, H. In-silico bioprospecting: finding better enzymes. *Mol. Biotechnol.* **61**, 53–59 (2019).
72. United Nations. *Intergovernmental Conference on Marine Biodiversity of Areas Beyond National Jurisdiction* (UN, 2023).
73. Timmis, K. et al. The contribution of microbial biotechnology to sustainable development goals. *Microb. Biotechnol.* **10**, 984–987 (2017).
74. Sutherland, D. L. et al. How microalgal biotechnology can assist with the UN Sustainable Development Goals for natural resource management. *Curr. Res. Environ. Sustain.* **3**, 100050 (2021).
75. Blasiak, R., Jouffray, J.-B., Wabnitz, C. C. C. & Österblom, H. Scientists should disclose origin in marine gene patents. *Trends Ecol. Evol.* **34**, 392–395 (2019).
76. Cross, S. et al. Who funded the research behind the Oxford–AstraZeneca COVID-19 vaccine? *BMJ Glob. Health* **6**, e007321 (2021).
77. Selig, E. R. et al. Mapping global human dependence on marine ecosystems. *Conserv. Lett.* **12**, e12617 (2019).
78. Dedeurwaerdere, T., Melindi-Ghidi, P. & Broggiato, A. Global scientific research commons under the Nagoya Protocol: towards a collaborative economy model for the sharing of basic research assets. *Environ. Sci. Policy* **55**, 1–10 (2016).
79. Hess, C. & Ostrom, E. *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press, 2007).
80. Scholz, A. H. et al. Myth-busting the provider-user relationship for digital sequence information. *GigaScience* **10**, giab085 (2021).
81. Rohden, F., Huang, S., Dröge, G. & Scholz, A. H. *Combined Study on Digital Sequence Information in Public and Private Databases and Traceability*. <https://www.cbd.int/abs/DSI-peer/Study-Traceability-databases.pdf> (2020).
82. Langlet, A. & Dunshirn, P. *Traceability Options for Marine Genetic Resource from Areas beyond National Jurisdiction*. <https://www.highseasalliance.org/wp-content/uploads/2023/02/traceability-options-paper-1.pdf> (2023).
83. Oldham, P. *Digital Sequence Information—Technical Aspects*. <https://doi.org/10.13140/RG.2.2.15833.24163> (2020).
84. Housen, W., Sara, R. & Jaspars, M. *Digital Sequence Information on Genetic Resources: Concept, Scope and Current Use*. <https://www.>

- [cbd.int/doc/c/fef9/2f90/70f037ccc5da885dfb293e88/dsi-ahteg-2020-01-03-en.pdf](https://doi.org/10.1038/s44183-024-00088-0) (2020).
85. Bornmann, L., Wagner, C. & Leydesdorff, L. The geography of references in elite articles: which countries contribute to the archives of knowledge? *PLoS ONE* **13**, e0194805 (2018).
86. Roach, M. & Cohen, W. M. Lens or prism? Patent citations as a measure of knowledge flows from public research. *Manag. Sci.* **59**, 504–525 (2013).
87. Heemskerk, E. et al. The promise and perils of using big data in the study of corporate networks: problems, diagnostics and fixes. *Glob. Netw.* **18**, 3–32 (2018).

Acknowledgements

We thank Robert Blasiak, Fran Humphries, Marcel Jaspars, Fabian Rohden, Siva Thambisetty, Petro Tolochko, and the anonymous reviewers for their valuable input. P.D. is funded by the Research Platform Governance of Digital Practices at the University of Vienna. E.Z. is funded by FORMAS, project number 2020-01048. Open Access funding provided by University of Vienna. The funders played no role in the study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author contributions

P.D. led the study design and performed the data collection. P.D. and E.Z. wrote the manuscript text, performed data analysis, and prepared the figures.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44183-024-00088-0>.

Correspondence and requests for materials should be addressed to Paul Dunshirn.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024