



DISSERTATION | DOCTORAL THESIS

Titel | Title

The Ocean Genome as a Global Commons: Empirical
Foundations for Equitable Governance

verfasst von | submitted by

Paul Dunshirn BA BA MSc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Doktor der Philosophie (Dr.phil.)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 796 310 300

Dissertationsgebiet lt. Studienblatt | Field of
study as it appears on the student record
sheet:

Politikwissenschaft

Betreut von | Supervisor:

Univ.-Prof. Mag. Dr. Alice Vadrot

Univ.-Prof. i.R. Mag. Dr. Immanuel Bomze

Table of contents

List of abbreviations	3
Publication overview and contributions.....	4
Acknowledgements.....	5
Abstract.....	6
Zusammenfassung.....	7
1. Introduction.....	8
2. The social and political context of marine genetic resource governance.....	11
2.1. Marine genetic resources' value to nature, society, and biotechnology.....	11
2.2. The emergence of genetic resources as a global political issue	13
2.3. Marine genetic resources as a new frontier in High Seas governance	15
2.4. Current political issues around genetic resources and DSI.....	16
2.5. Empirical studies on the provision and use of genetic resources and DSI.....	18
3. Conceptualizing the ocean genome commons	21
3.1. Commons governance theory.....	21
3.2. Extension of the commons framework to hybrid resources	23
3.3. Conceptualizing the ocean genome commons	24
4. Research design and methodology.....	25
4.1. Analytical framework for the systematic study of marine genetic resource collection and use	25
4.2. Methodological approach for the three empirical studies	29
4.3. Research ethics.....	29
5. Guide through the research papers.....	29
6. Paper 1: Global Extractivism of Digital Sequence Information: Structural and Temporal Dynamics	31
7. Paper 2: Conducting marine genetic research for whom? Mapping knowledge flows from science to patents	51
8. Paper 3: Growing prominence of deep-sea life in marine bioprospecting.....	63

9. Discussion	75
9.1. Empirical and theoretical contributions	75
9.2. Policy implications and pathways towards improved equity	76
9.3. Limitations and future research.....	78
10. Conclusion	79
11. References.....	79

List of abbreviations

ABS – Access and Benefit-Sharing

BBNJ – Biodiversity Beyond National Jurisdiction

CBD – Convention on Biological Diversity

DSI – Digital Sequence Information

FAO – Food and Agriculture Organization

GISN – Global Influenza Surveillance Network

IAD framework – Institutional Analysis and Development framework

INSDC – International Nucleotide Sequence Database Collaboration

OBIS – Ocean Biodiversity Information System

OECD – Organization for Economic Co-operation and Development

PIP framework – Pandemic Influenza Preparedness framework

UN – United Nations

UNCLOS – United Nations Convention on the Law of the Sea

WHO – World Health Organization

Publication overview and contributions

This cumulative dissertation is based on three empirical studies. ‘Global Extractivism of Digital Sequence Information: Structural and Temporal Dynamics’ (Paper 1) was submitted to *PNAS* on the 25th of February 2025 and is currently under review. ‘Conducting marine genetic research for whom? Mapping knowledge flows from science to patents’ (Paper 2) was published in *npj Ocean Sustainability* in October 2024 (<https://www.nature.com/articles/s44183-024-00088-0>). ‘Growing prominence of deep-sea life in marine bioprospecting’ (Paper 3) was published in *Nature Sustainability* in August 2024 (<https://www.nature.com/articles/s41893-024-01392-w>).

I was in the lead and fully responsible for the conceptualization, design, supervision, data collection, data processing, analysis, literature research, writing, and critical review of Papers 1 and 2. All co-authors of Papers 1 and 2 supported me in the writing process, and Erik Zhivkoplías supported me with additional data analysis steps for both papers. Erik Zhivkoplías is the first author of Paper 3. I contributed substantively to writing and revising the manuscript, particularly in elaborating the relevance of its findings to multilateral political processes. I was also involved in Paper 3’s conceptualization, design, and data analysis stages.

Acknowledgements

This thesis is the outcome of several years' challenging work. While this process was not without a few moments of despair, it offered me a unique opportunity to get to know and learn from many incredible people around the world involved in biodiversity and ocean governance. These people enabled me to carry out my research and made me realize the immense societal, environmental, and cultural importance of marine biodiversity and ocean genes, which significantly contributed to my motivation to complete this research.

I would like to particularly thank my supervisor, Prof. Alice Vadrot, who trusted me to become part of her incredible research team, MARIPOLDATA, and the Environmental Politics Research Group at the University of Vienna. Alice's expertise, support, and openness were crucial to my personal development as a researcher and enabled me to enter the world of international ocean governance, a terrain previously unfamiliar to me. I am grateful to all MARIPOLDATA and Environmental Politics Research Group members for the feedback, shared lessons, and joint experiences throughout the four years of my PhD. I also want to thank Prof. Barbara Prainsack, Prof. Nikolaus Forgo, and the other Research Platform Governance of Digital Practices members for supporting my work and setting the scene for my inter- and transdisciplinary research. I am highly grateful to Prof. Immanuel Bomze, my co-supervisor, for his continued support, particularly during early research design phases and my experiments with (too) many different datasets.

Amongst the many inspiring encounters outside of the University of Vienna, the support of Assoc. Prof. Robert Blasiak and his team at the Stockholm Resilience Centre was particularly important in helping me find my footing in the topic of marine genetic resource governance. I am incredibly grateful for the many discussions and joint work with Erik Zhivkoplías, which allowed me to understand many interconnections between politics, bioinformatics, and environmental sciences that would have otherwise remained inaccessible to me. I also want to thank all members of the network Interdisciplinary Researchers Working on Digital Sequence Information (iDSI) and the informal 'Vienna DSI group', Daniel Kachelriess and Dr. Arne Langlet, for many interesting discussions and their support during my PhD. My gratitude also goes out to Assoc. Prof. Siva Thambisetty, Assoc. Prof. Fran Humphries, Prof. Marcel Jaspars, and Fabian Rohden for their constructive feedback on my PhD research.

Most importantly, I want to thank my family and friends whose love and support have positioned me to carry out and find joy in my research.

Abstract

This thesis examines the flow of marine genetic resources from the point of sample collection, through scientific research and the extraction of genetic sequence data, to their eventual use in patents. It presents three empirical studies that systematically explore global relationships between provider and user countries emerging from this process. Using a combination of social network, semantic, and regression analyses, the studies identify patterns across diverse data sources, including genetic sequence metadata and taxonomic, bibliometric, and patent data. These analyses offer a novel systems-level perspective on marine genetic resources, capturing how various stages of collection and use constitute an overarching resource system – here described as the ‘ocean genome commons’. The thesis evaluates this system using criteria from commons governance theory to assess its functioning and equity. While the findings indicate positive developments, including growing international scientific collaboration, they also highlight persistent global disparities in research and innovation capacities and a ‘freeriding problem’ in the provision and use of marine genetic research in patents. The thesis discusses how similar challenges have been addressed in other global commons and considers how emerging multilateral regimes – specifically the UN Treaty on Biodiversity Beyond National Jurisdiction (BBNJ Treaty) and the Convention on Biological Diversity (CBD) – can incorporate these lessons. By offering a holistic analysis of the ocean genome commons, it provides new impetus to multilateral and national policy efforts fostering equitable collection and use of marine genetic resources.

Zusammenfassung

Diese Dissertation untersucht die globale Nutzung meeresgenetischer Ressourcen, ausgehend von der Entnahme genetischer Proben über deren wissenschaftliche Beforschung und die Gewinnung genetischer Sequenzdaten, bis hin zu ihrer Verwendung in Patenten. Sie stellt drei empirische Studien vor, welche systematisch Beziehungen zwischen Anbieter- und Nutzerländern in diesem Fluss an Information und Materialien nachverfolgen. Anhand einer methodologischen Kombination aus sozialer Netzwerkanalyse, semantischer Analyse und Regressionsanalyse identifizieren die Studien Muster über verschiedene Datenquellen hinweg, insbesondere Metainformation zu genetischen Sequenzdaten sowie taxonomische, bibliometrische und patentbezogene Daten. Diese Analysen zeigen auf, wie die verschiedenen Stufen der Sammlung und Nutzung meeresgenetischer Ressourcen ein übergreifendes System bilden, das hier als „Ocean Genome Commons“ beschrieben wird. Die Dissertation bezieht sich auf Kriterien aus der „Commons Governance Theory“, um zu beurteilen, wie funktional und gerecht dieses derzeitige System ist. Die Ergebnisse deuten zum einen auf eine positiv zu bewertende Zunahme an internationaler wissenschaftlicher Zusammenarbeit hin. Zum anderen stellen sie jedoch auch anhaltende globalen Ungleichheiten in Forschungs- und Innovationskapazitäten fest, in Kombination mit Mustern von „Freeriding“ bei der Bereitstellung und Nutzung meeresgenetischer Forschung in Patenten. Anhand dieser Resultate geht die Arbeit der Frage nach, wie ähnliche Herausforderungen im Kontext anderer globaler Gemeinschaftsgüter („Commons“) angegangen wurden. Schlussendlich bespricht die Arbeit wie entstehende multilaterale Verträge - insbesondere der neuen UN-Hochseeabkommen (BBNJ-Vertrag) und das Übereinkommen über die biologische Vielfalt (CBD) - diese Erkenntnisse berücksichtigen können.

The Ocean Genome as a Global Commons: Empirical Foundations for Equitable Governance

1. Introduction

The evolution of life in the ocean dates back around 3.7 billion years, positioning marine biodiversity as a vibrant and diverse source of genetic information (Blasiak et al. 2020). Such genetic information encodes most of the attributes and behaviors of marine organisms, playing a crucial role in their ability to thrive in specific habitats or adapt to changing environmental conditions. Studying these genes thus becomes a highly valuable endeavor, enabling scientists to understand and ‘learn from’ nature to develop much-needed strategies and technologies to respond to the multiple crises nature and societies are currently facing (Richardson et al. 2023; Biermann and Kim 2020).

Besides contributing to our understanding and preparing us for the consequences of continued biodiversity and genetic diversity loss (Richardson et al. 2023), genetic materials of marine organisms constitute valuable ‘marine genetic resources’¹, which can be the source for natural product discovery or the development of biosynthetic products (Rotter et al. 2021). For instance, researchers discovered marine bacteria that degrade plastics (Wilkes and Aristilde 2017), identified candidate compounds for new antibiotics (Tortorella et al. 2018), and use marine genetic resources to develop bio-based chemicals to replace environmentally harmful products (Beygmoradi and Homaei 2017). Such research and innovation activities typically involve a complex chain of events, sometimes described as the ‘marine biodiscovery pipeline’ (Jaspars et al. 2016a). This pipeline can span across the initial collection of genetic materials, their analysis and processing in scientific laboratories, the publication of research, the sharing of materials and sequence data, and their use in follow-up scientific and commercial activities. International research collaborations, for instance in international research expeditions, and cooperation between countries in the cross-border transfer of materials and data are crucial to the functioning of genetic research and biodiscovery pipelines (Jaspars, Humphries, and Rabone 2021).

However, a range of politically contentious issues are entangled with these activities, which is why they have become subject to international negotiations and regulations in recent decades. First and foremost, scientists active in marine genetic research or marine biotechnological innovation are primarily located in developed countries, creating strong global imbalances between developed and developing countries in the ability to benefit from the wealth of knowledge and information (Blasiak et al. 2018; Tolochko

¹ Marine genetic resources are defined as ‘any material of marine plant, animal, microbial or other origin containing functional units of heredity of actual or potential value’ (United Nations General Assembly 2023). This definition is a specification of the older and broader definition of genetic resources under the CBD (Convention on Biological Diversity 1992).

and Vadrot 2021; Oldham et al. 2025). This status quo arguably runs counter to the objectives of the CBD – the UN’s central mechanism for biodiversity governance, which aims to facilitate equitable and sustainable conservation and use of biological diversity, including genetic resources and associated digital sequence information (DSI) (Klünker 2024). These imbalances also appear to conflict with the so-called ‘Common Heritage Principle’, a guiding principle under the ‘specialized’ multilateral Treaty on Plant Genetic Resources for Food and Agriculture (‘Seed Treaty’) and the newly adopted BBNJ Treaty (Gottlieb, Kachelriess, and Slobodian 2025; Halewood 2013). Even though its exact meaning has been contested (Vadrot, Langlet, and Tessnow-von Wysocki 2021), the Common Heritage Principle generally prescribes a degree of shared ownership and rights amongst all countries to benefit from genetic resources (see section 2.3). The capacity imbalances, the legal complexity, and the diversity of involved actors, including local communities, public institutions, scientists, and private companies, altogether pose challenges to the development of functional and equitable multilateral governance frameworks for marine and other types of genetic resources.

The academic and policy-oriented literature informing the governance of marine genetic resources mainly consists of case studies, legal analyses, or perspective pieces focusing on one specific set of practices involved in the collection and use of ocean genes. For instance, Robert Blasiak et al.’s much cited work on ‘corporate control of marine genetic resources’ (Blasiak et al. 2018) identified a concentration of intellectual property in a handful of companies and countries; the work of Fran Humphries, Muriel Rabone, and Marcel Jaspars explored various examples of the collection and use of ocean genes they were able to derive from existing literature and data to analyze possibilities of ‘tracing’ such activities via a multilateral system (Humphries, Rabone, and Jaspars 2021; Jaspars, Humphries, and Rabone 2021); Miguel C. Leal et al. analyzed the impact of global capacity building on developing countries’ abilities to discover marine natural products (Leal et al. 2020); and Paul Oldham’s work provided large-scale overviews of implicated science, intellectual property, and commercial landscapes (Oldham et al. 2014; 2025). While highlighting different equity issues in an instructive and empirically grounded manner, these studies provide little insight into overarching system dynamics spanning across the various stages of the marine biodiscovery pipeline. This lack of a holistic understanding limits the literature’s ability to inform practitioners and policymakers interested in working towards more equitable collection and use of marine genetic resources.

This thesis addresses this gap by providing a holistic, systematic, and empirically grounded analysis across various marine genetic resource collection and use stages. To do so, I introduce an analytical framework that, inspired by existing conceptions of genetic resource traceability (Humphries, Rabone, and Jaspars 2021), allowed me to trace global flows of genetic materials, research, and data across three key domains – countries of origin, science, and patenting. A core component of this framework is the idea to use scientific publication identifiers to establish connections across these three domains. The

three empirical papers at the core of the thesis operationalize this framework via large-scale data integration and computational analysis (including social network analysis, automated text analysis, and logistic regression) and a variety of publicly available datasets – including genetic sequence repositories [GenBank](#) and [WiLDSI](#) (Lange et al. 2021), bibliometric and patent data from [The Lens](#), and taxonomic data from [OBIS](#). This methodological approach fills key gaps in existing literature by providing a systematic and scalable way to analyze the collection and use of marine genetic resources not in terms of disparate domains but rather as a holistic resource system, here further conceptualized as the ‘ocean genome commons’.

I introduce the notion of the ocean genome commons to connect my empirical observations on marine genetic resource provision and use to commons governance theory (Ostrom 1990; Hess and Ostrom 2007). This theoretical embedding allows me to critically assess the resource system’s ability to deliver on ambitions voiced by involved stakeholders, particularly concerning equity, transparency, and the continued conduct and provisioning of scientific research. In proposing the specific notion of ocean genome commons, I am referring to the concept of the ‘ocean genome’, which is defined as ‘the ensemble of genetic material present in all marine biodiversity, including both the physical genes and the information they encode’ (Blasiak et al. 2020). This concept usefully captures the ‘hybridity’ (Wolter and Sievers-Glotzbach 2019) of marine genetic resources when treated as a governance issue, as their resource status varies substantially across contexts (from living organisms to samples, scientific articles, genetic sequence data, patents, and commercial products). The combined notion of the ocean genome commons thus allows me to critically assess the system of users, providers, governing institutions, and guiding rules around this hybrid resource, enriching my interpretation of the findings and providing directions for systematic policy interventions.

The next sections are organized as follows: Section 2 reviews the literature on the political and social context of marine genetic resource governance (2.1 – 2.4) and the empirical research on marine genetic resource provisioning and use (2.5). Section 3 describes my use of commons governance theory, firstly by summarizing its basic logic and principles (3.1), its extension to knowledge commons and hybrid resources (3.2), and by conceptualizing marine genetic resource provisioning and use as a system of interactions between rule-guided actors that constitute the ocean genome commons (3.3). Section 4 introduces my analytical framework (4.1), its operationalization following a multi-methods approach (4.2), and ethical considerations (4.3). Section 5 overviews the three empirical papers and their data sources, methods, scope, and theoretical embedding. Section 6 is Paper 1, which explores relationships between marine genetic resources’ geographic origins and researchers’ affiliation countries. Section 7 presents Paper 2 on the science-patent nexus in marine genetic research. Section 8 is the third paper providing an interdisciplinary analysis of the marine bioprospecting patent landscape. Section 9 discusses how the thesis contributes to the literature on marine genetic resources and commons

governance (9.1), the findings' policy implications (9.2), and limitations and future research needs (9.3). Section 10 concludes the thesis.

2. The social and political context of marine genetic resource governance

This literature review section contextualizes the thesis' empirical contributions in broader societal themes and political processes. It first surveys the literature on marine genetic resources' significance to nature, people, and innovation (2.1). It then turns to the historical emergence of genetic resources as a political issue in international politics (2.2) and discusses the specificity of governing the collection and use of marine genetic resources within and beyond national jurisdiction (2.3). This is followed by a more detailed discussion of the two contemporary political questions directly addressed by this thesis, namely which bureaucratic measures are needed to reach satisfactory levels of traceability and sharing of benefits derived from marine genetic resources and DSI (2.4.1), and how to maintain the openness of science while holding corporate users of such science and related data responsible (2.4.2). Finally, it surveys empirical research on the collection and use of marine genetic resources, critically reflecting on existing knowledge gaps that must be addressed to establish a holistic and systematic understanding to guide policy development (2.5).

2.1. Marine genetic resources' value to nature, society, and biotechnology

Most characteristics of biological life are encoded in the genomes of organisms, whether it is the ability of crops to resist environmental pressures, the properties of viruses to attack other organisms, or an organism's physical appearance. The ability to use and benefit from such natural genetic properties has played a central role in the emergence of human societies, as can be seen in the use of wild plants for early agricultural systems (Raustiala and Victor 2004) or traditional medicine (Vierros and Aalbersberg 2010).

The scientific discovery of the gene as a biological organizing principle in the 20th century triggered more purposeful exploration of genetics, positioning it as central to contemporary biological, health, and environmental research (Keller 2002). Since then, 'bioprospecting' practices have emerged and inspired many highly valuable biotechnologies. Bioprospecting describes the systemic search and analysis of genes responsible for an organism's essential processes, such as intra- and interspecies communication, bacterial defense, and antifouling (Rotter et al. 2021; Puglisi et al. 2014), with an interest in re-using the gained knowledge to develop products for human application, including medicines, cosmetics, or foods.

The Human Genome Project (1990-2003), which created a draft sequence of the entire human genome, not only revolutionized our understanding of genetics but also paved the way towards the establishment of large-scale open-access international genetic sequence databases. This development was partly

driven by the organizational need of involved scientists, who operated in a global network of laboratories, to regularly exchange and publicly share data (Maxson Jones, Ankeny, and Cook-Deegan 2018). Nowadays, most journals require scientists to publish their data in the so-called International Nucleotide Sequence Database Collaboration ([INSDC](#)), a network of the three largest databases that regularly mirror their data. This information ecosystem has become the backbone of much of genetic research and biotechnological innovation, as it enables large-scale comparisons of functional patterns related to known genetic properties and processes (Contreras and Knoppers 2018). In this ‘data-driven’ age of biological research (Strasser 2012), the value of genetic resources no longer exists merely in physical materials but also in the globally shared information derived from them.

The interest in specifically marine genetic resources arose only in recent decades. Marine biodiversity is nowadays seen as an up-and-coming source for bioprospecting due to its comparatively long evolutionary history, its genetic richness, and the fact that terrestrial genetic diversity has already been further explored (Eren and Delmont 2024; Chen et al. 2024; Blasiak et al. 2020). Genes found in extreme environments, such as deep sea and ecosystems around hydrothermal vents, attract particular interest due to the special conditions under which such ‘extremophile’ species manage to thrive (Zhang et al. 2024). Antibiotics, which have predominantly relied on terrestrial genetic sources, are a good example of this trend: as antibiotic resistance to established compounds increases, researchers increasingly search for novel antibiotic sources in the ocean, including the deep sea (Tortorella et al. 2018). Overall, the market value for marine biotechnology was estimated to reach 6.4 billion US dollars by 2025 (Hurst et al. 2016).

Marine genetic resources are not only of value to science, innovation, and bioprospecting activities, but also to ecosystem and planetary health, as genes and genetic diversity underpin critical ecosystem services (Stange, Barrett, and Hendry 2021). Genetic variability within species is fundamental to adaptive capacity and resilience in response to environmental stressors such as ocean warming, acidification, and reductions in salinity (Webster et al. 2017). Furthermore, genetic research and technologies have revolutionized the ability to monitor biological life, particularly in ecosystem monitoring, for instance, by identifying and monitoring invasive species via environmental DNA (Ardura et al. 2015).

Finally, marine genetic resources play essential roles in socio-cultural traditions and practices, including those of Indigenous Peoples and Local Communities (IPLCs). For instance, they are crucial to traditional environmental management practices, ecological knowledge, and medicine (Vierros and Aalbersberg 2010; Drew 2005). IPLCs have played central roles in the exploration and cultivation of marine genetic resources, which in many cases inspired scientific research or biotechnological

innovation (Bhatia and Chugh 2015) – a link that often gets neglected or misrepresented (Golan et al. 2022).

2.2. The emergence of genetic resources as a global political issue

The political concerns over genetic resources date back to the United Nations Educational, Scientific and Cultural Organization's (UNESCO) International Biosphere Conference in 1968 and the Stockholm Conference on the Human environment in 1972, which formulated first recommendations to Parties and the United Nations Food and Agricultural Organization (FAO) to take measures to protect and preserve samples of increasingly threatened ecosystems and species (Sirakaya 2022). Combined with concerns over global food security, this led the FAO to build a worldwide network of gene banks to store crop genetic materials, conserve their genetic diversity, and invest in the development of new crop varieties and agricultural biotechnologies. The first voluntary international instrument to regulate access to genetic resources stored in these gene banks was adopted by the FAO in 1983. This process led to the adoption of the so-called Seed Treaty, which entered into force in 2004, ensuring global access to the genetic resources of 64 specific crop types (Sirakaya 2022).

In parallel to the discussions at the FAO, developing countries pushed for measures that would extend sovereign rights over biological resources under the United Nations Environmental Program (UNEP). This development was driven by long-standing concerns over 'bio-piracy', which occurs when external scientists or companies (often from developed countries) extract genetic materials from developing countries and Indigenous communities without consent or appropriate recognition of rights (Wynberg 2023). A well-documented historical example of biopiracy is the case of a British smuggler who transported thousands of rubber tree seeds out of Brazil, which, in turn, allowed the British Empire to produce rubber at an industrial scale in South-East Asia (Gollin 2008; Jackson 2008). Furthermore, developing countries perceived FAO's conceptualization of genetic resources as Common Heritage of Humankind as further contributing to such inequities if applied to other genetic resources under their jurisdiction (Sirakaya 2022).

UNEP negotiations on a new biodiversity Treaty started as Ad Hoc Working Groups in 1988 and concluded with the adoption of the CBD at the Rio Conference in 1992 and its entry into force one year later. The CBD granted sovereign rights to nation states over their biological resources and defined a mechanism for bilateral access and benefit-sharing on genetic resources, including marine genetic resources within their jurisdiction (Kamau and Winter 2013; Sirakaya 2022). This bilateral system constituted a departure from the Seed Treaty's Common Heritage system towards stronger control over domestic resources (Raustiala and Victor 2004). The Nagoya Protocol to the CBD, which entered into force in 2014, imposed more concrete legal obligations in line with this bilateral system. This CBD mechanism relies on an implicit balancing act in which countries grant freedoms to research on 'their'

biodiversity under the condition that part of the created benefits will be shared with the host country (Broggiato et al. 2018).

Since then, the World Health Organization (WHO) introduced the Pandemic Influenza Preparedness (PIP) framework in 2011. This framework was built upon WHO's Global Influenza Surveillance Network (GISN), which has been operating since 1952 to ensure rapid global collection and circulation of influenza virus information. Responding to unsatisfactory access to vaccines and tensions over intellectual property rights under GISN, the PIP framework introduced an additional access and benefit-sharing mechanism to balance the sharing of influenza virus information with regulation on the sharing of intellectual property and vaccines (Rabitz 2017). WHO Parties are currently negotiating a new 'Pathogen Access and Benefit-Sharing System' (PABS) to cover a broader range of pathogens (Walckiers, Frison, and Aubry 2024), which is expected to be formulated as an annex to the WHO Pandemic Agreement, adopted in 2025.

Together with the access and benefit-sharing mechanism on marine genetic resources under the recently adopted BBNJ Treaty (see next section), these frameworks constitute a regime complex with potentially overlapping provisions and scopes (Raustiala and Victor 2004; Morin 2014). Article 4 of the Nagoya Protocol enables Parties to recognize other regimes as specialized instruments, creating the possibility for diverging regulation on specific genetic resources under their scope (Lawson, Rourke, and Humphries 2020). However, no regimes have been formally recognized as such by the CBD until today, leaving the exact relationship between these regimes subject to further deliberation. Governments have introduced their own domestic rules in the meantime, such as the European Union regulation 511/2014 that treats the PIP framework and the Seed Treaty as specialized instruments, exempting users from the Nagoya Protocol in applicable cases (Kachelriess et al. 2025).

All of these fora are currently deliberating on how to extend their frameworks to cover not only physical genetic materials but also genetic sequence data, also known as DSI, derived from such materials (Aubry et al. 2022; Lawson, Rourke, and Humphries 2020). When the BBNJ Treaty was adopted in 2023, it was the first legally binding instrument to explicitly include both material and DSI in its legal scope. The CBD's 2024 decision on a multilateral mechanism on DSI (Convention on Biological Diversity 2024) constituted another step in extending genetic resource governance to the digital realm. The discussions on DSI are shaped by familiar issues, such as the concerns over bio-piracy (Bond and Scott 2020), but also by new challenges, such as the increased use of artificial intelligence and the associated lack of data traceability (Sett et al. 2024).

2.3. Marine genetic resources as a new frontier in High Seas governance

The High Seas have been a key site of struggle between global powers for centuries. In the 16th and 17th centuries, Europeans fought over control of trade routes, navigation, and resources in these vastly expanding parts of the ocean under no national jurisdiction. Contrasting ideas of how the High Seas should be governed have populated international debates since that time. Hugo Grotius' 'Mare Liberum' (Grotius 1914) is often mentioned as one of the most influential publications from that time, proposing the 'Freedom of the High Seas', according to which no nation should have exclusive rights over the High Seas itself, nor over navigation or trade on them (Vieira 2003). This principle was partly formulated to support Dutch traders' interests in Southeast Asia and challenged Portuguese sovereignty claims in these regions (Vieira 2003).

While the Freedoms of the High Seas have since become formalized as guiding principles of modern Ocean governance under the United Nations Convention on the Law of the Sea (UNCLOS), they have not remained unchallenged. It has been pointed out that these principles were initially introduced to serve the interests of European colonial states (Eslava, Fakhri, and Nesiah 2017; Enyew 2022). Some argue that, while the Freedoms formally established equal rights amongst all states on Areas Beyond National Jurisdiction (the legal term designating the High Seas and the Seabed), they have in fact introduced a 'first come – first serve' regime that primarily benefits powerful states, which are free to exploit resources (such as fish stocks or minerals) at the disadvantage of states with less technological or scientific capacity (Enyew 2022).

This skepticism towards the Freedoms of the High Seas has shaped developing countries' positions in the third UNCLOS negotiations (1973-1982) as they aimed to neutralize them by pushing for the Common Heritage Principle to guide processes under UNCLOS. This principle was initially proposed by Maltese diplomat Arvid Pardo in 1967 to distribute revenues from mineral resources found beyond national jurisdiction to counteract global equity imbalances (Vadrot, Langlet, and Tessnow-von Wysocki 2021; Hannigan 2015). After over two decades of contentious negotiation, states finally agreed to include the Common Heritage Principle under UNCLOS Article 136 (Wolfrum 2023; Tuerk 2017).

As already discussed above, the tensions between the Freedom of the High Seas and the Common Heritage Principle have since extended to the topic of this thesis, namely, marine genetic resources and their use in research, biotechnology, and nature conservation. Even though the collection and use of marine genetic resources within national jurisdiction had already been regulated under the CBD since 1993, no such framework had existed for Areas Beyond National Jurisdiction. Considering the high scientific, ecological, and economic potential of marine genetic resources (Blasiak et al. 2020) and the regulatory uncertainty arising from the gap in regulation on Areas Beyond National Jurisdiction (Tiller et al. 2019), countries decided to integrate this issue into broader negotiations on a new UNCLOS

implementing agreement on the conservation and sustainable use of marine biodiversity beyond national jurisdiction, nowadays called the BBNJ Treaty. When the BBNJ negotiations formally started in 2018, marine genetic resources constituted one of the Treaty's four package elements, alongside area-based management tools, environmental impact assessments, and capacity-building. During negotiations, developed countries, and particularly those with advanced capacities in genetic research and biotechnology, wanted to see access and use of marine genetic resources governed in terms of the Freedoms of the High Seas (Vadrot, Langlet, and Tessnow-von Wysocki 2021; Marciniak 2020). In turn, developing countries argued that the Common Heritage Principle should be extended from mineral resources to such genetic resources.

This underlying divergence of views on legal principles also structured the BBNJ negotiations on the Treaty's more technical details. Advocates of the Common Heritage Principle tended to support obligatory forms of monetary benefit-sharing to developing countries as well as stricter types of traceability, sometimes described as 'track-and-trace' approaches (Humphries, Rabone, and Jaspars 2021; Langlet and Dunshirn 2023). On the other hand, proponents of the Freedom of the High Seas favored more light-weight transparency measures and emphasized the significance of non-monetary benefit-sharing contributions (Brogiato et al. 2018).

A compromise was eventually found that enabled the BBNJ Treaty to be adopted in June 2023. It included both the Freedoms of the High Seas and the Common Heritage Principle (Humphries, Berry, and Gottlieb 2025). At the technical level, the compromise combined monetary benefit-sharing with a 'smart' traceability system based on a novel 'BBNJ batch identifier' linked to genetic resources and associated data originating from Areas Beyond National Jurisdiction (Oldham and Thambisetty 2023). Many questions on the traceability and benefit-sharing system still need to be resolved moving forward, including how monetary contributions for Parties will be assessed. While Parties are preparing for the Treaty's entry into force, which will happen one hundred days after sixty ratifications are reached, there is a pressing need for a knowledge foundation supporting deliberations on the unclarified issues.

2.4. Current political issues around genetic resources and DSI

Multilateral versus bilateral access and benefit-sharing

Discussions on the further development of the genetic resource regime complex are shaped by frustrations among scientific and corporate users with the CBD's Nagoya Protocol (Prathapan et al. 2018). The Protocol's bilateral logic created a complex landscape of diverging national regulations that users need to abide by when accessing or using genetic resources from different provider countries (Salem and Kaltenpoth 2023). Evidence on whether introduced measures have led to the desired benefit-sharing and capacity-building is lacking (Scholz et al. 2021), as bilateral agreements often remain

undisclosed to the public (Muñoz-García, Lago, and Scholz 2024; Robinson and Wright 2023). However, it should be mentioned that the question of whether the Protocol ‘works’ has turned into a political issue itself, as several developing countries keep defending its legacy and effectiveness. For instance, Brazil has invested substantially in developing and improving its national access and benefit-sharing system to reduce bureaucracy while maintaining control over its ‘genetic heritage’ – Brazil’s legal concept for both domestic physical genetic resources and DSI (Relly 2024).

In this context, stakeholders have started exploring new conceptualizations of access and benefit-sharing to address existing problems. Both the BBNJ Treaty and the CBD’s recent decision on DSI (Convention on Biological Diversity 2024) introduced novel multilateral systems that do not rely on bilateral agreements but instead on a centralized logic that guides traceability, access, and benefit-sharing between Parties. A multilateral solution was, to some extent, logical under the BBNJ Treaty, which by definition does not regulate genetic resources originating from a particular country. CBD negotiations on a new mechanism for DSI were more controversial. Some Parties advocated for a strictly multilateral system that, according to them, could even replace existing bilateral systems. Others have advocated for a hybrid system, which would allow countries to decide on their own whether they regulate DSI under their national access and benefit-sharing regulation or under the novel multilateral system. The CBD’s 2024 decision has not resolved these tensions, which means that further discussions and potential co-existence of multilateral and bilateral access and benefit-sharing on DSI will occur moving forward.

Discussions on the relationship between traceability and benefit-sharing

Another component of current debates is the extent to which traceability of genetic resources and DSI is needed to establish equitable benefit-sharing. Discussions on track-and-trace approaches versus light-weight transparency were a controversial subject in the BBNJ negotiations, as described in the previous section. The negotiations on DSI under the CBD have mirrored this divide. The CBD’s initial COP15 decision to establish a multilateral mechanism on DSI included a somewhat controversial clause stating that ‘tracking and tracing of all DSI is not practical’ (Convention on Biological Diversity 2022). Some interpreted this essentially as recognition that track-and-trace is impracticable, while others interpreted it as stating that it may be impracticable for *all* but not for *some* DSI (Oldham, Chiarolla, and Thambisetty 2023).

During later stages of the BBNJ negotiations, the idea of an alternative ‘decoupled’ approach to access and benefit-sharing emerged. Decoupling is the idea to distribute benefits according to certain indicators that would not necessarily be informed by singular events of access or utilization of DSI (or genetic resources) (Langlet and Dunshirn 2023; Scholz et al. 2023). This idea led to more differentiated discussions on the extent of decoupling, where the option to ‘fully decouple’ was contrasted to a ‘light

decoupling’, which would maintain certain connections between access, use, and benefit-sharing calculations, albeit in terms of aggregate indicators rather than track-and-trace of individual sequences (Oldham, Chiarolla, and Thambisetty 2023; Langlet and Dunshirn 2023). BBNJ Treaty’s article 14 ultimately installed a ‘two-phase’ benefit-sharing system in which developed countries will initially contribute 50% of their assessed annual contributions to the Agreement, with various options for the transformation of the mechanism to be considered at future BBNJ COP meetings, including various decoupled options (Brogiato et al. 2024).

Political tensions around the science-business nexus

A more specific issue underlying these debates is the relationship between science and commercial activities. This relationship has been a recurrent theme in the emergence of genetic research and its governance. One of the main motivations for scientists and funders to engage in the already mentioned Human Genome Project was an economic one, namely that a thorough understanding of the human genome would enable a range of follow-up innovation processes (Maxson Jones, Ankeny, and Cook-Deegan 2018). At the same time, the leaders of the Human Genome Project adopted an explicit policy, the Bermuda Data Principles, for sequences to be published daily as prior art to prevent subsequent ‘broad’ patenting of the human genome (Cook-Degan & Heaney 2010; Eisenberg 2000). Up until today, many genetic researchers perceive the possibility that their findings flow into patents or inspire valuable products not only as economically beneficial but also as a relatively central motivation behind their research activities (Jaspars et al. 2016a), often while promoting open science and data in parallel.

The complicated relationship between public and private interests has shaped genetic resource governance before and after the Human Genome Project, as I already discussed above for the Seed Treaty and the PIP framework. It also shaped the BBNJ negotiations, which were still in progress when the COVID-19 pandemic broke out in 2020. Particularly at the start of the pandemic, developing countries struggled to get access to much-needed vaccines partly due to pharmaceutical companies’ insistence on intellectual property rights (Thambisetty et al. 2021). Developing countries explicitly referred to these global inequities when arguing for the Common Heritage Principle in relation to marine genetic resources (Thambisetty 2022). However, intellectual property rights remained largely excluded from the scope of the final Treaty, leaving it uncertain whether and, if so, how its provisions may address political tensions concerning the science-patent nexus.

2.5. Empirical studies on the provision and use of genetic resources and DSI

Different literature traditions have empirically explored the scientific and corporate activities addressed by the political debates on marine genetic resources. I distinguish between:

- ‘Landscape studies’ - research that identifies actors either providing or using genetic resources or DSI and assessing their rate of activity in a specific domain, such as marine genetic resource patenting;
- ‘Case studies’, which provide highly detailed descriptions of how different domains are connected in practice, such as how scientific research flows into subsequent patenting activities;
- ‘Relational studies’, which address the same connections between domains but on a larger scale.

These different traditions provide valuable insights into aspects of marine genetic resource provision and use but also entail important gaps that this thesis addresses.

2.5.1. Landscape studies

Several studies provide systematic empirical insights into the scientific landscape of marine biology. Across the board, these studies point to strong global imbalances in the distribution of research capacities, which can, for instance, be observed in the number of vessels nationally available for scientific explorations (Sierra-Correa et al. 2020). Imbalances are also visible in international collaboration networks, in which researchers from developed countries maintain predominant positions (Tolochko and Vadrot 2021). Publication output continues to be driven by developed countries (Leal et al. 2020; Lara-Lopez et al. 2020). While developing countries’ capacities have increased over the past years, they are still comparatively low (Leal et al. 2020; Lara-Lopez et al. 2020). Similar trends in publication output were shown to exist specifically for research involving marine genetic resources in policy-oriented reports (Oldham et al. 2014; 2025), but such analysis has not yet made it into peer-reviewed scientific literature, which Paper 2 accomplishes.

Regarding corporate activities involving marine genetic resources, most research focuses on assessing patent data. Blasiak et al. (2018) have built on previous research by Arnaud-Haond et al. (2011) to show that 47% of all sequences included in marine genetic resource patents belong to one single German company (BASF), and that 98% of all actors that have patented marine sequences are associated with only ten countries worldwide. While providing important empirical and systematic insights into patenting activities, the findings of Blasiak et al. (2018) have been challenged by some academic peers and by corporate stakeholders involved in the BBNJ negotiations. Their critique questioned whether sequence or document counts constitute reasonable metrics to assess the concentration of patenting activities and whether the framing of the findings in terms of ‘corporate control’ of marine genetic resources is politically appropriate (own observations). Paper 3 addresses these concerns and provides an improved update of the study from 2018, largely confirming its original empirical findings.

Little research has been conducted on corporate activities involving marine genetic resources beyond patents. This is likely due to the absence of readily available methods or data to conduct such research beyond individual case studies (Humphries, Rabone, and Jaspars 2021). A recent report (Oldham et al. 2025) introduced an innovative web scraping methodology to identify commercial products.

For both the scientific and corporate landscape studies, it remains difficult to differentiate between marine genetic resources of Areas Beyond National Jurisdiction and those collected from within national jurisdiction. This has made it difficult to assess the economic potential of marine genetic resources of Areas Beyond National Jurisdiction (Oldham et al. 2014; 2025), which has been a recurring point of uncertainty in the BBNJ negotiations. Paper 3 introduces a methodological approach that addresses this question by identifying species exclusively present in Areas Beyond National Jurisdiction.

2.5.2. Case studies

A few case studies provide insights into the process that ties the collection, sequencing, scientific publishing, patenting, and commercialization of marine genetic resources together (Humphries, Rabone, and Jaspars 2021; Jaspars, Humphries, and Rabone 2021; Jaspars et al. 2016b; Broggiato et al. 2018). Described cases include a bacterium isolated from a deep-sea worm, which French researchers collected and later used to develop a face cream (Humphries, Rabone, and Jaspars 2021; Cambon-Bonavita et al. 2002), and the ‘bengamide’ molecule, which U.S. American scientists found in a sea sponge from Fiji and which inspired anticancer clinical trials by the company Novartis (Jaspars, Humphries, and Rabone 2021; White, Tenney, and Crews 2017). The case studies provide important insights into how particular marine biodiscovery pipelines operate. They indicate that publicly funded scientists can play a central role in marine bioprospecting (Blasiak et al. 2019), particularly during earlier stages of product discovery, namely in the collection and initial analysis of genetic materials, the publication of findings, and the sharing of genetic sequence data. However, it remains unclear how representative this is of all activities involved in the global collection and use of ocean genes. Central to the aim of this thesis is to explore the generalizability of these case studies.

A comprehensive review of case studies involving non-marine genetic resources is beyond the scope of this thesis. However, it should be mentioned that many of the activities involved in the marine bioprospecting pipeline resemble biotechnological innovation practices more broadly. Particularly when it comes to the use of DSI, practitioners often do not distinguish between data of marine and terrestrial origins when conducting biotechnological research (Raposo et al. 2023).

2.5.3. User-provider relations

Studies falling into the ‘user-provider relations’ tradition have addressed some of the limitations inherent in case study research by systematically studying patterns in the provisioning and use of genetic materials and data. These studies have explored the geographic origins of various types of DSI and found that most sequences are extracted from the same regions where the involved scientists are based (Rohden et al. 2020; Scholz et al. 2021). Similar research explored global user-provider balances specifically in the search for marine natural products and found an increase in developing countries’ capacities over the past 50 years (Leal et al. 2020).

There is a lack of systematic research on the politically contentious nexus between genetic research, patenting, and commercialization. As discussed above, literature exists on the genetic resource patent landscapes (Oldham, Hall, and Forero 2013; Blasiak et al. 2018). However, the extent to which patent applicants draw on publicly available science and how such intellectual property translates into commercial products remains a black box. Economic and bibliometric research on science-patent interlinkages in other contexts indicates that corporations often rely on knowledge from publicly available basic research for their internal research and development processes (Arora, Belenzon, and Pataconi 2018; Cross et al. 2021; Marx and Fuegi 2020). Paper 2 draws on this external literature to address the specific knowledge gap on the science-patent nexus in the context of marine genetic resources.

3. Conceptualizing the ocean genome commons

This section introduces the theoretical framework underlying my research approach and analysis. I first survey the original themes that inspired Elinor Ostrom and colleagues in developing the commons governance theory (Ostrom 1990) (3.1). I then discuss more recent extensions of the framework to knowledge commons (Hess and Ostrom 2007; Frischmann, Madison, and Strandburg 2014) and hybrid commons (Wolter and Sievers-Glotzbach 2019) (3.2), before introducing the notion of the ocean genome commons (3.3).

3.1. Commons governance theory

Garrett Hardin’s influential work on the ‘tragedy of the commons’ (Hardin 1968) posited that rational actors are bound to exploit commonly shared resources to an extent that diminishes the common good (e.g., through over-grazing). The commons governance literature introduced a critique of this notion, showing for a multitude of cases that cooperation between actors can lead to equitable and sustainable outcomes (Ostrom 1990). Based on various case studies, cross-case analysis, and experimental studies, the literature highlights the value of self-regulation and stewardship in accomplishing such desirable

outcomes, alongside a range of other factors that contribute to successful resource management (Ostrom 1990).

Central to most commons governance research is a two-dimensional resource classification system, distinguishing levels of subtractability and exclusion. Within this classification, the commons literature originally focused on community-managed ‘common-pool resources’. For instance, coordination amongst community members to share grazing grounds would be considered as common-pool resource management. Common-pool resources are characterized by a low level of excludability (i.e., it is hard to exclude resource users) and a high level of subtractability (i.e., one person’s use diminishes the resource’s value for others) (Ostrom 1990). These resources are typically shaped by collective action problems related to overuse or freeriding, where users reap benefits from a shared resource without sufficiently contributing to it (Hess and Ostrom 2007). By studying such collective action problems across different resource systems, researchers can abstract generalizable governance principles likely to lead to certain outcomes.

The commons governance literature typically relies on the so-called Institutional Analysis and Development (IAD) framework, which distinguishes three main variable clusters:

- ‘Resource system characteristics’, which include: ‘biophysical characteristics’, describing resource size, location, boundaries, and abundance as determined by its nature and the available technologies to cultivate or harvest it (Hess and Ostrom 2007); ‘attributes of the community’, which indicate community organization, the roles of different members, and shared goals or objectives; ‘rules in use’, which are formal or informal rules that are known and, at least to a certain degree, enforced by community members or via implicated institutions (Wolter and Sievers-Glotzbach 2019);
- ‘Action arenas’ – concrete situations in which actors engage with one another as mediated through the resource system and its components, thereby reproducing the resource system’s characteristics (Hess and Ostrom 2007). Behaviors observed in such action arenas can be indicative of so-called ‘patterns of interaction’, which are practices that emerge from repeated actions (Hess and Ostrom 2007);
- ‘Outcomes’, which are produced by patterned interactions, such as the enclosure of scientific knowledge behind journal paywalls. ‘Evaluative criteria’ allow us to judge whether these outcomes are positive or negative in terms of maintaining or building a functional commons (Hess and Ostrom 2007).

The core evaluative criteria of the original commons literature are efficiency and equity (Ostrom 1990; Keohane and Ostrom 1995). Efficiency describes, in essence, the extent to which resources invested in a commons are used or governed in a way to maximize returns for users (Ostrom 2009). Equity is often

differentiated into ‘fiscal equivalence’, describing whether those who benefit from a resource have also appropriately contributed to it, and ‘distributional equity’, which considers whether resources are shared in a manner that counteracts capacity or other imbalances amongst users (Ostrom 2009). Various other evaluative criteria have been employed in the commons literature, such as sustainability, accessibility of knowledge, or reciprocity (Hess and Ostrom 2007).

3.2. Extension of the commons framework to hybrid resources

Researchers have extended the commons framework to knowledge resources in recent years. Some knowledge resources may also constitute common-pool resources, but others may be better described as ‘public goods’, another category of the mentioned resource classification system (Hess and Ostrom 2007; Frischmann, Madison, and Strandburg 2014). The difference between common-pools and public goods lies in their level of subtractability: public goods, such as the Internet, do not necessarily diminish in value through use. As such, other types of collective action problems relating to resource underuse can be observed for knowledge commons. Resource underuse problems may occur, for instance, if investments into a digitally shared library do not pay off due to low user engagement or if an information ecosystem gets polluted with useless information (Hess and Ostrom 2007). Nevertheless, themes of overuse or knowledge enclosure were also identified for knowledge commons, for instance concerning the effects of paywalled journals on science accessibility (Kranich in the knowledge commons book) or issues related to patent law (Adelman 2005; Heller 2013).

Several researchers have pointed out that a clear distinction between public goods and common-pool resources seems difficult when applying this theory to genetic resources. For instance, Wolter and Sievers-Glotzbach (2019) describe how a collectively organized apple breeding initiative in Germany incorporates both common-pool and public good logics. The initiative clearly delineates a community of users and defines property and use rights over seeds and their scions, implying a common-pool resource logic. At the same time, the community considers its knowledge, for instance on breeding techniques, as a public good that it openly shares with external farmers through publications or personal contacts. Wolter and Sievers-Glotzbach (2019) have thus characterized this specific arrangement as a ‘hybrid commons’. Other literature on seed genetic resources has equally embraced the fact that governance institutions need to account for the resource’s economic, informational, and cultural value, constituting commons that combine both material and knowledge components (Sievers-Glotzbach et al. 2020; Girard 2018; Frison 2018; Halewood 2013).

Hybridity has also been observed by several scholars working on marine genetic resources. As Siva Thambisetty pointed out, some actors in the BBNJ negotiations framed marine genetic resources as a public good, arguing that little to no value gets subtracted through the collection of genetic samples and subsequent use (Thambisetty 2023). However, the same ocean genes may transform into a common-

pool resource once actors patent applications inscribed in the genes, diminishing the value of the resource to others through the enforcement of intellectual property rights (Thambisetty 2023). Blasiak et al. (2020) proposed the concept of the ‘ocean genome’ to capture such shifting conceptions of marine genetic resources and associated governance issues.

3.3. Conceptualizing the ocean genome commons

Combining the commons governance literature and the concept of the ocean genome, I now introduce the notion of the ‘ocean genome commons’. Similarly to the case of the German apple breeding initiative (Wolter and Sievers-Glotzbach 2019), the ocean genome commons is best characterized as a hybrid commons. It contains physical components, such as living marine organisms (eukaryotes, viruses, bacteria, and archaea), samples extracted from them, research vessels, laboratories, gene banks, and databases. The ocean genome commons’ digital components are, for instance, DSI extracted from marine organisms, research papers, and patent documents. The thesis explores several of these components across its three empirical papers, studying the provisioning of genetic materials for DSI-related publications (Paper 1), the number of marine genetic research articles cited in patent documents (Paper 2), or the number of marine species whose sequences get cited in patents (Paper 3).

The ocean genome commons involves a variety of members, which may be distinguished horizontally between domains, such as academia versus corporate domain, or vertically between world regions, countries, academic institutions, or companies, as well as individual scientists or inventors. Due to the thesis’ interest in contributing to policy development under multilateral agreements, the analytical focus lies on nation states, which mediate the activities of various sub-national actors that participate in the commons. While this is not a very typical level of analysis in the commons literature, previous work identified similarities between commons and institutionalist international relations theory as both consider institutions as key to cooperative interactions between actors (Keohane and Ostrom 1995). As discussed in more detail in Paper 2, nation states play crucial roles in regulating and incentivizing the collection and use of genetic resources, for instance, dispersing research funding or granting access to researchers, and designing rules governing the ocean genome via multilateral agreements. As such, I understand nation states as central actors within this resource system.

The ocean genome commons’ constitutional rules are shaped by the CBD and the BBNJ Treaty, including their national implementing regulations. Furthermore, social practices constitute important rules-in-use that guide actors’ behaviors and day-to-day decision-making on the ground (Dedeurwaerdere 2010). These practices are influenced by social relations and discourses. For instance, Pálsson and Prainsack (2011) highlighted altruistic motivations and discourses guiding geneticists and biologists in making their data public. As highlighted throughout this thesis, the ocean genome commons’ relevant rules in use are not limited to the sharing or use of specific genetic information or

materials, but also include the reporting of metadata to establish traceability, transparency, and trust amongst community members (Golan et al. 2022; Blasiak et al. 2019).

A widely recognized goal across genetic resource Treaties and involved stakeholders is equity, which, following the legal logic of the CBD's Nagoya Protocol, is often understood to be achieved as soon as providers and users have entered a bilateral agreement (Klünker 2024). This is a very specific understanding of equity, focusing on communicative equity between user and provider countries (Klünker 2024), thereby neglecting other types of equity as elaborated in the commons literature (Frison 2018). The thesis adopts a wider conceptualization of equity, assessing outcomes in terms of fiscal and distributional equity, but also in reference to reciprocity (Oldham 2009; Mauss 1956) (Paper 1), the provisioning of open access to scientific knowledge and data (Paper 2), and sustainability (Paper 3) (see Table 1).

4. Research design and methodology

This section describes the analytical framework that guided the research for this thesis (4.1), the employed research methods and datasets (4.2), as well as ethical considerations (4.3). It focuses on the overarching logic of the research approach – detailed discussions of methodological choices concerning individual studies can be found directly in the papers.

4.1. Analytical framework for the systematic study of marine genetic resource collection and use

My primary research aim was to establish a systematic and holistic understanding of various stages of marine genetic resource collection and use by studying involved actors, their interactions, and associated outcomes. I followed an abductive logic (Timmermans and Tavory 2012) to identify relevant units of analysis and data sources. On the one hand, I relied on established conceptions of the marine biodiversity pipeline and of the flow of materials and data therein (Humphries, Rabone, and Jaspars 2021; Rohden et al. 2020; Blasiak et al. 2018). On the other hand, the scarcity of generalizable research on involved practices and actors required me to combine this with my own exploration and experimentation with different data sources and methods.

Humphries et al. (2021) was instructive in this process, providing a detailed analysis of how marine genetic resources and data could practically be traced across various stages from the departure of the research vessels to the potential commercialization of a product. The study entails an overview of marine bioprospecting stages and considered public data availability at each stage. A CBD-commissioned study on the traceability of DSI (Rohden et al. 2020) provided similarly valuable insights into the possibility of tracing connections between different stages of genetic resource collection and use. This study showed that a vast amount of sequence metadata is stored in the INSDC information

ecosystem, albeit at varying levels of coverage (Rohden et al. 2020). This metadata provides considerable traceability across initial stages of genetic resource collection, sample and genetic sequence data deposition, and publication of research papers (see Fig. 1). However, the available literature provides little clarity on traceability across subsequent marine bioprospecting stages where the publicly available materials, data, and research flow into corporate activities.

Takifugu rubripes mRNA for IgM heavy chain constant region, partial cds, membrane-bound IgM

GenBank: AB125610.1

[FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS AB125610 293 bp mRNA linear VRT 26-JUL-2016
 DEFINITION Takifugu rubripes mRNA for IgM heavy chain constant region, partial cds, membrane-bound IgM.
 ACCESSION AB125610
 VERSION AB125610.1
 KEYWORDS .
 SOURCE Takifugu rubripes (Fugu rubripes)
 ORGANISM [Takifugu rubripes](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Neoteleostei; Acanthomorpha; Eupercaria; Tetraodontiformes; Tetraodontoidea; Tetraodontidae; Takifugu.
 REFERENCE 1
 AUTHORS Saha,N.R., Suetake,H. and Suzuki,Y.
 TITLE Analysis and characterization of the expression of the secretory and membrane forms of IgM heavy chains in the pufferfish, Takifugu rubripes
 JOURNAL Mol. Immunol. 42 (1), 113-124 (2005)
 PUBMED [15488950](#)
 REFERENCE 2 (bases 1 to 293)
 AUTHORS Saha,N.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2003) Nil Ratan Saha, The University of Tokyo, Aquatic Bioscience; 2971-4, Maisaka, Shizuoka 431-0211, Japan (E-mail:ratu20@yahoo.com, Tel:81-53-592-2821, Fax:81-53-592-2822)
 FEATURES
 source
 1..293
 /organism="Takifugu rubripes"
 /mol_type="mRNA"
 /db_xref="taxon:31033"
 /clone_lib="Fugu spleen cDNA"
 /dev_stage="Adult"
 /geo_loc_name="Japan"
 /note="synonym: Fugu rubripes"
 CDS
 <1..147
 /note="membrane-bound IgM"
 /codon_start=1
 /product="IgM heavy chain constant region"
 /protein_id="BAD26620.1"
 /translation="VNSDFLIETNSSAVHAEEDNMGDTAITFIFLFLITLLFSIATTA FKIK"
 ORIGIN
 1 gtaaattcag atttcctgat tgaacaaac agcagtgtg tgcgtcgga agaagacaac
 61 atgggggaca cagccatcac cttcatctt ctcttctca ttacactcct gttcagcatt
 121 gccaccaccg cttcaagat taaatgaaaa gaaaaaagga aaatctcacc caaattatat
 181 ttgtcaagt ttgtcttaaa ttcattgtac gtgagtaatc acaactgtcc gtacgatgtc
 241 attcgtgtc tagttataaa ataatgttt taaatgaaaa aaaaaaaaaa aaa

Fig. 1: Screenshot of an exemplary genetic sequence entry on GenBank (<https://www.ncbi.nlm.nih.gov/nuccore/AB125610>), displaying information on taxonomy, country of sample origin ('geo_loc_name'), the scientific publication linked to the sequence, and involved scientists and their affiliations.

My experimentation with datasets and methods thus focused on finding a method to trace connections between scientific and corporate activities. I first explored the possibility of tracing mentions of genetic sequence unique identifiers (i.e., ‘accession numbers’) in patent texts, an idea mentioned but not further pursued in a report on technical aspects of DSI traceability (Oldham 2020). By following up on this idea, I discovered that patent applicants often mention accession numbers when reporting outputs of sequence similarity estimations or related bioinformatic procedures. This leaves it unclear whether a particular sequence or the genetic properties of underlying genetic resources contributed to an invention in a substantial manner. Distinguishing types of accession number mentions via automated text analysis seemed too complicated at that time, which is why I gave up on this approach. Large language models, which were not available to the same extent as they are now, could aid in future efforts along those lines.

The CBD-commissioned study (Rohden et al. 2020) mentioned another option to trace connections between science and patents, which was to link patent genetic sequence identifiers, which patent offices report back to INSDC, back to original INSDC accession numbers using sequence similarity estimations. However, as a political scientist, this bioinformatic technique seemed too far out of the scope of my methodological capacities.

As such, I turned to a third option, which involved tracing patent citations to research publications that are listed in genetic sequence metadata on INSDC. These publications are sometimes referred to as ‘primary DSI publications’ (Scholz et al. 2021). As I discovered, patent-to-science citations are commonly analyzed in economic and bibliometric research on knowledge flow networks (Jaffe and de Rassenfosse 2017), where their analytical value has been tested and confirmed (Roach and Cohen 2013). Integrating this approach into research on marine genetic resources and DSI thus appeared like an interesting methodological innovation that could enhance traceability and facilitate systematic research.

Combined with direct feedback from many of the authors of the above discussed publications, this process led me to formulate an analytical framework that establishes three essential domains for the systematic study of the ocean genome commons, namely ‘countries of origin’ (i.e., countries of sample origin), ‘countries of science’ (affiliations of researchers that published DSI primary publications), and ‘countries of patenting’ (headquarters of entities applying or owning related patents). Importantly, the connections between these domains can be studied through knowledge flows implicit in scientific publications (see Fig. 2).

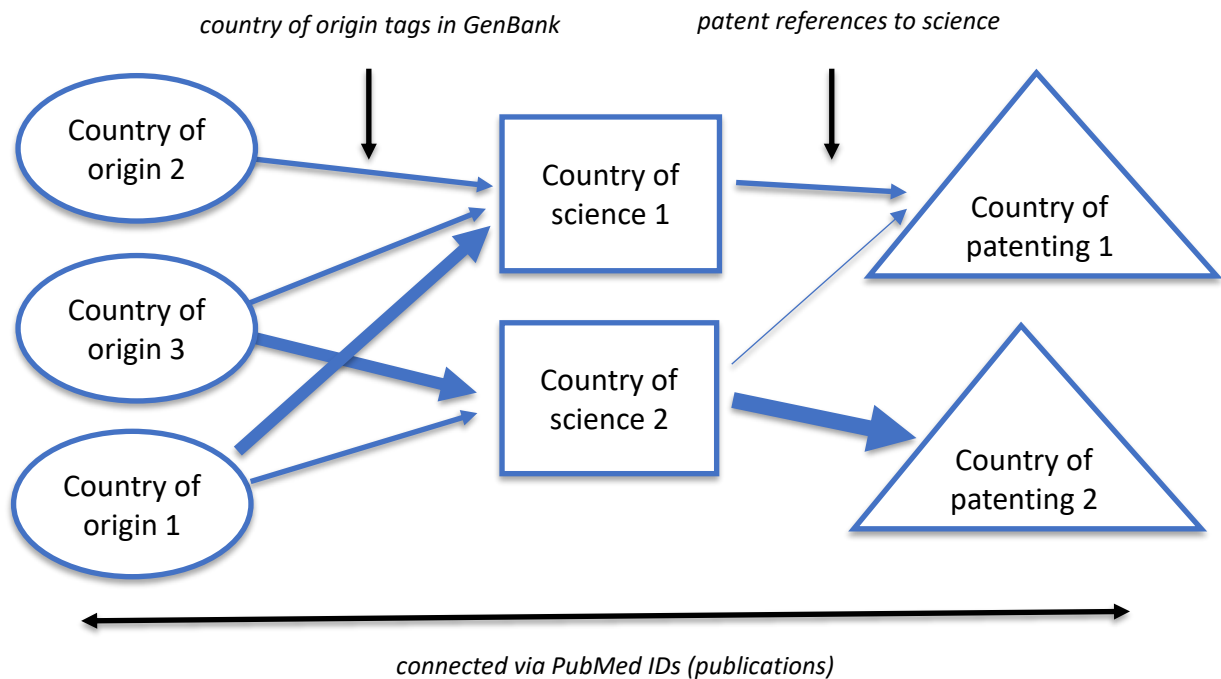


Fig. 2: Analytical framework for the systematic study of marine genetic resource collection and use across countries of origin, science, and patenting. Edges represent knowledge flows, including their direction and weight.

As visible in Fig. 2, I opted to trace knowledge flows from start to end – from geographic origins, to science, to corporate activities, even though it could have also made sense to apply the reverse approach, from endpoints back to origins (Humphries, Rabone, and Jaspars 2021). However, this would have made tracing linkages from publications back to the genetic sequence databases more difficult at that time – a step that could nowadays be done using the WiLDSI database (Lange et al. 2021).

This framework can be understood as an adaptation of the IAD, focusing on key resource system characteristics, which could subsequently be linked to action arenas, patterned interactions, and outcomes. Even though other relevant domains of the ocean genome commons are not considered in my proposed framework, most importantly the commercialization of products, I argue that it still captures most of the resource system’s essential components and enables the systematic study of their interrelations through the connecting element of scientific publications – something that to my knowledge had not been done before.

4.2. Methodological approach for the three empirical studies

All three papers relied on the public genetic sequence databases [GenBank](#) and [WiLDSI](#) and on taxonomic data from [OBIS](#). Papers 1 and 2 drew substantively on the specialized bibliometric and patent database [The Lens](#).

Methodologically, the papers relied on large-scale data integration and various computational procedures, which I carried out using the programming language Python. Essential components of each paper are descriptive statistics and visualizations of data distributions and bivariate relationships encountered in the constructed datasets. The thesis relied on automated text analysis of scientific publication titles to construct a variable of genetic research types (Paper 1) and to explore the relevance of research content to follow-up scientific or patenting activities (Paper 2). Paper 1 included a multivariate logistic regression analysis. Paper 3 relied on bioinformatic techniques, which my co-author Erik Zhivkoplías employed. Social network analysis shaped many components of the three papers, not necessarily in terms of applying concrete network analytical measures, but primarily for structuring data and tracing knowledge flows across the various stages of the ocean genome commons. All used datasets and code are publicly shared together with the publications, or, for a few exceptions, available upon request. The code used for Papers 1 and 2 is available at <https://github.com/pduns>.

4.3. Research ethics

The presented research did not involve any primary data collection, nor did it include personal or otherwise sensitive data – it is exclusively based on publicly accessible secondary data sources, including genetic sequence data, taxonomic data, bibliometric data, and patent data. None of the findings or discussions reveal personal information or identities. The project development phase included several reflections on research ethics with colleagues and supervisors, which did not reveal any concerns to be addressed during my research. The Research Platform Governance of Digital Practices at the University of Vienna funded my dissertation project.

5. Guide through the research papers

The data curation and analysis process led me to distill three empirical studies on various aspects of the ocean genome commons (Table 1). Paper 1 addresses the relationships between countries of origin and countries of science. The scope of this study goes beyond the marine context. It covers DSI primary publications more broadly, a choice made to connect to similar research on global user-provider relationships that got published during my PhD (Scholz et al. 2021). Paper 2 analyzes patent-to-science citations to trace the flow of marine genetic research between countries of science and patenting. Paper 3 is a collaborative paper with researchers from the Stockholm Resilience Center, following up previous

studies on the patent landscape around marine genetic resources (Arnaud-Haond, Arrieta, and Duarte 2011; Blasiak et al. 2018).

Title	Data sources	Methods	Scope	Embedding in the commons framework
Global Extractivism of Digital Sequence Information: Structural and Temporal Dynamics (Paper 1)	WiLDSI database (DSI metadata and lists of species covered by multilateral regimes); The Lens (bibliometric and semantic data); OBIS (taxonomic data)	Logistic regression; Word co-occurrence network analysis; Leiden community detection	Marine and other genetic resources; relations between countries of origin and countries of research	Trends in international collaboration in the collection and scientific use of genetic resources; reciprocity and equity
Conducting marine genetic research for whom? Mapping knowledge flows from science to patents (Paper 2)	OBIS (taxonomic data); GenBank (DSI metadata); The Lens (bibliometric, patent, and semantic data)	Quantitative descriptive data analysis; Semantic analysis of paper titles	Marine genetic resources; relations between countries of research and countries of patenting	Freeriding on the provisioning of scientific research; equity
Growing prominence of deep-sea life in marine bioprospecting (Paper 3)	GenBank's patent division; Orbis	Quantitative descriptive data analysis; taxonomic analysis; various bioinformatic techniques (e.g., BLAST)	Marine genetic resources; landscape of patent applicants and countries of patenting	Enclosure of knowledge and concentration of intellectual property rights; sustainability and equity

Table 1: Overview of data sources, methods, scope, and embedding in commons governance theory for the three papers.

6. Paper 1: Global Extractivism of Digital Sequence Information: Structural and Temporal Dynamics

Main Manuscript for

Global Extractivism of Digital Sequence Information: Structural and Temporal Dynamics

Paul Dunshirn*, Erik Zhivkoplis**, Immanuel Bomze***, Arne Langlet*, Katie Lee Riddle****, Alice Vadrot*

*Department of Political Science, University of Vienna, Vienna, Austria

**Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden

***Faculty of Mathematics and Research Network Data Science, University of Vienna, Vienna, Austria

****Te Kotahi Research Institute, University of Waikato, Hamilton, New Zealand

Corresponding author: Paul Dunshirn

Email: paul.dunshirn@univie.ac.at

Author Contributions: Paul Dunshirn led the study design and performed data visualization. Paul Dunshirn and Erik Zhivkoplis performed data collection, data processing, and analysis. All authors contributed to the writing of the manuscript text.

Competing Interest Statement: The authors declare no competing interests.

Classification: Social Sciences, Genetics

Keywords: digital sequence information, biodiversity governance, research collaboration, global extractivism, capacity building

This PDF contains:

Abstract (p. 2)

Significance statement (p. 2)

Main text (p. 3)

Introduction (p.3)

Results (p. 4)

Discussion (p. 8)

Materials and methods (p. 13)

Data and code availability (p. 16)

Acknowledgements (p.16)

Literature (p.16)

Abstract

Do developed countries disproportionately extract digital sequence information (DSI) from developing countries? Existing research suggests that scientists are generally more likely to use genetic sequences from within their own geographic and economic context. This paper provides a differentiated empirical analysis, exploring whether and where certain forms of 'DSI extractivism' exist. We conceptualized DSI extractivism as a relationship in which some countries source economically or scientifically valuable genetic information from other countries without the involvement of local scientists. Using a bibliometric dataset of 94,598 publications, we found such relationships to be most prevalent when researchers from developed countries sourced DSI from developing countries. Basic research activities and the sourcing of DSI from marine organisms, threatened species, or mega-biodiverse countries were most indicative of extractivism, pointing to persistent problems related to funding schemes and international cooperation. The more applied research types of biotechnology and public health were less affected. Finally, we observed a decline of extractivism over time, which can be interpreted in terms of the impacts of access and benefit sharing regulations since the 1990's as well as increases in developing countries' scientific capacities. Our findings provide a nuanced assessment of global information extractivism in genetic research, highlighting positive developments as well as persistent equity issues to be considered under relevant UN fora, such as the United Nations Convention on Biological Diversity (CBD).

Significance Statement

The paper offers a detailed analysis of non-cooperative relationships between DSI provider and user countries, a phenomenon which we call 'DSI extractivism'. Contrasting previous research that has questioned the prevalence of global imbalances in this context, it provides evidence that up to 40% of all international DSI research is extractive. However, it also shows that DSI extractivism has gradually decreased over time, giving rise to more collaborative forms of research. These insights are of high relevance to the emerging global regulatory framework on the collection and use of DSI under various multilateral UN fora, as well as to public institutions and scientific communities seeking to establish forms of international collaboration that have long-term capacity building effects.

Main Text

Introduction

Digital sequence information (DSI), or genetic sequence data, is essential to genetic research, biotechnology, and biodiversity or disease monitoring. Considered as a prime example of a global information commons (1), DSI is globally collected and made openly accessible through the International Nucleotide Sequence Data Collaboration (INSDC), a collaboration between the world's largest genetic sequence databases. Regulatory intervention into this important information ecosystem is often met with suspicion, particularly by scientific and corporate DSI users who consider it as imperative that scientific openness and freedoms are maintained (2, 3).

However, representatives of developing countries as well as Indigenous peoples and local communities have criticized the current state of this information commons for the lack of clarity on geographic DSI origins, associated sovereign rights, and benefit sharing from the use of DSI. Extending the long-standing political debates over global access to and usage of 'physical' genetic resources, this 'digital version' in form of DSI has thus appeared as a central negotiation issue under the Convention on Biological Diversity (CBD), the World Health Organization (WHO), the UN Food and Agricultural Organization (FAO), the UN Convention on the Law of the Sea (UNCLOS), and the World Intellectual Property Organization (WIPO). Across these fora, Parties are currently struggling to develop and implement a new legal framework that, on the one hand, ensures sustained access to DSI and, on the other, creates equitable conditions for providers and users of different world regions.

The concerns over access and equity regarding DSI need to be seen in the context of colonial and post-colonial histories of resource extraction from the Global South to the Global North (4), and a fundamental normative shift from a common heritage system to the legalization of property rights (5). Especially since the early 1990s, negotiations in the CBD and the FAO underscored the value of genetic resources for industrialized countries and the desire of developing countries to control access to their plant genetic resources (6, 7). While addressing these inequities in the realm of physical genetic resources has been one of the main aims of the CBD since its entry into force in 1993 and its Nagoya Protocol, which entered into force in 2014, similar concerns shape current discussions about DSI. Some observers argue that the lack of regulation on DSI constitutes a loophole that encourages new forms of 'biopiracy', allowing researchers and commercial users to bypass existing regulation on access and benefit sharing (8–10), or 'helicopter-research', describing the non-involvement of local researchers when sampling in developing countries (11–13). Furthermore, there are concerns over lacking recognition of the Indigenous origins of DSI and associated rights (14).

To better understand the relationship between DSI origins and usages, the CBD commissioned several studies in 2018 to provide a clearer picture of the global DSI ecosystem and to explore options to increase transparency and equity (15, 16). These efforts led to the construction of the [WiLDSI database](#) (17), which entails structured DSI metadata on origins and user affiliations. The empirical studies building on this database indicate that, in fact, most users rely on DSI from within their own geographic

or economic context (15, 18), questioning the notion of a clear ‘global user-provider relationship’ in which DSI primarily flows from developing to developed countries. In a certain sense, these findings leave it unclear whether access and benefit sharing for DSI is necessary at all.

While these studies provided a much-needed empirical assessment of global trends in the provision and usage of DSI, they did not scrutinize equity dimensions of cases where scientists sourced information from other countries. The fact that researchers are generally likely to source DSI domestically does, in our view, not preclude the parallel existence of unbalanced and extractive user-provider relations. To evaluate this, we defined DSI extractivism as a relationship in which a country, through its scientists, sources economically or scientifically valuable genetic information from other countries without involvement of local researchers. This notion of DSI extractivism mirrors equity issues articulated in the literature on helicopter research (11), but it emphasizes the embeddedness of such practices in persistent global relationships between countries, or groups of countries, and across time (19). Our understanding of the concept thus builds on what (18) described as user-provider relations, but it focuses specifically on the quality and directionality of these exchanges.

To empirically research the structural and temporal dynamics of global DSI extractivism, we used the WiLDSI database (17) to identify 94,598 publications for which DSI was collected and published together with one or, as in most cases, several ‘country of origin’ labels indicating the provenance of underlying samples. For each publication, we aggregated all countries of origin and author affiliations into economic groups, namely the Organization for Economic Cooperation and Development (OECD), the Group of 77 (G77), or the group of Brazil, Russia, India, China, and South Africa (BRICS). We considered a publication to be extractive if it was exclusively authored by scientists of one economic group but sourced DSI from another group. This allowed us to measure and compare shares of DSI extractivism across different group-group constellations, to trace developments over time, and to statistically explore the effects of contextual variables clustered under ‘research types’ (e.g. biotechnology or basic research), ‘regulatory contexts’ (e.g. marine genetic resources or pathogens), ‘socio-ecological contexts’ (mega-biodiverse countries or IUCN threatened species), and ‘taxonomy’ (assessed at the superkingdom level).

Results

Who extracts DSI from where?

We found 33,072 publications that rely on ‘internationally sourced’ DSI (i.e. DSI that was collected within a relationship between a provider group and a user group). The other 61,526 publications only sourced DSI from within the same group as their author affiliations. Amongst all group-group relationships, OECD researchers were the ones who sourced most DSI from others, namely for 18,870 publications from G77 countries and for 12,861 publications from BRICS countries (Fig. 1). In turn, BRICS researchers sourced 4,175 publications from the OECD and 3,619 from G77 countries. G77 researchers registered 2,993 publications with OECD origins and 2,232 with BRICS origins.

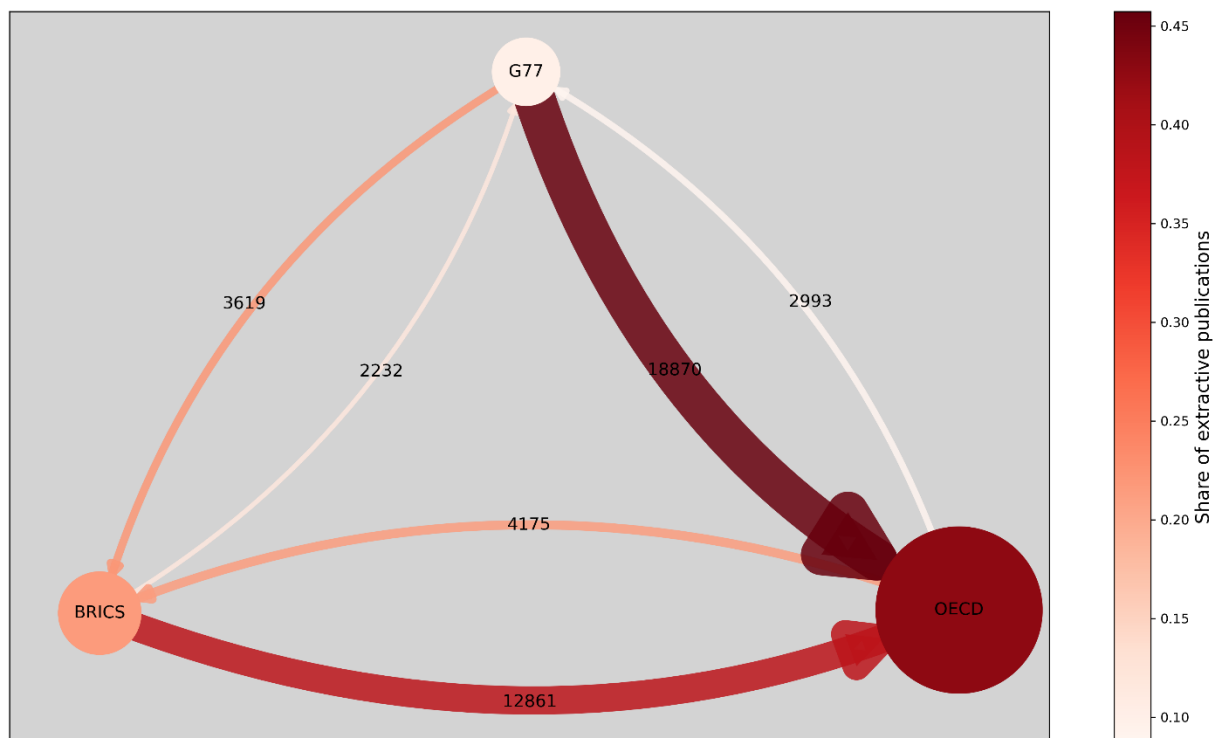


Fig. 1. Directionality of international DSI sourcing between economic groups and associated shares of extractivism. Edge thickness corresponds to the number of publications sourcing DSI from a group and edge color to the share of extractive publications within the relationship. Node color shows a group's extractivism share and node size reflects a group's number of internationally sourced DSI publications.

Out of the 18,870 publications in the G77-OECD relationship, we identified 45.7% to have been exclusively authored by OECD researchers and thus extractive, which marks the highest share in all group-group relationships (Fig. 1). This was followed by the BRICS-OECD relationship with an extractivism share of 38.3%. G77-BRICS had an extractivism share of 22.1%, OECD-BRICS a share of 21.4%, BRICS-G77 a share of 11.5%, and OECD-G77 a share of 8.7%. Overall, 39.7% of all internationally sourced DSI publications were linked to extractivism.

Contextualizing DSI extractivism

Figure 2 displays the distribution of DSI publications across taxonomic groups, research types, and socio-ecological or regulatory contexts. Scientists were most likely to internationally source DSI in the contexts of research on threatened species (72.1%), mega-biodiverse countries of origin (59.1%), basic research (56.5%), and eukaryotic species (52%). Within these distributions, shares of extractive publications were highest for threatened species (25.4%), basic research (18.5%), mega-biodiverse origins (17.5%), marine (16.6%), and eukaryotic species (16.4%) (Fig. 2).

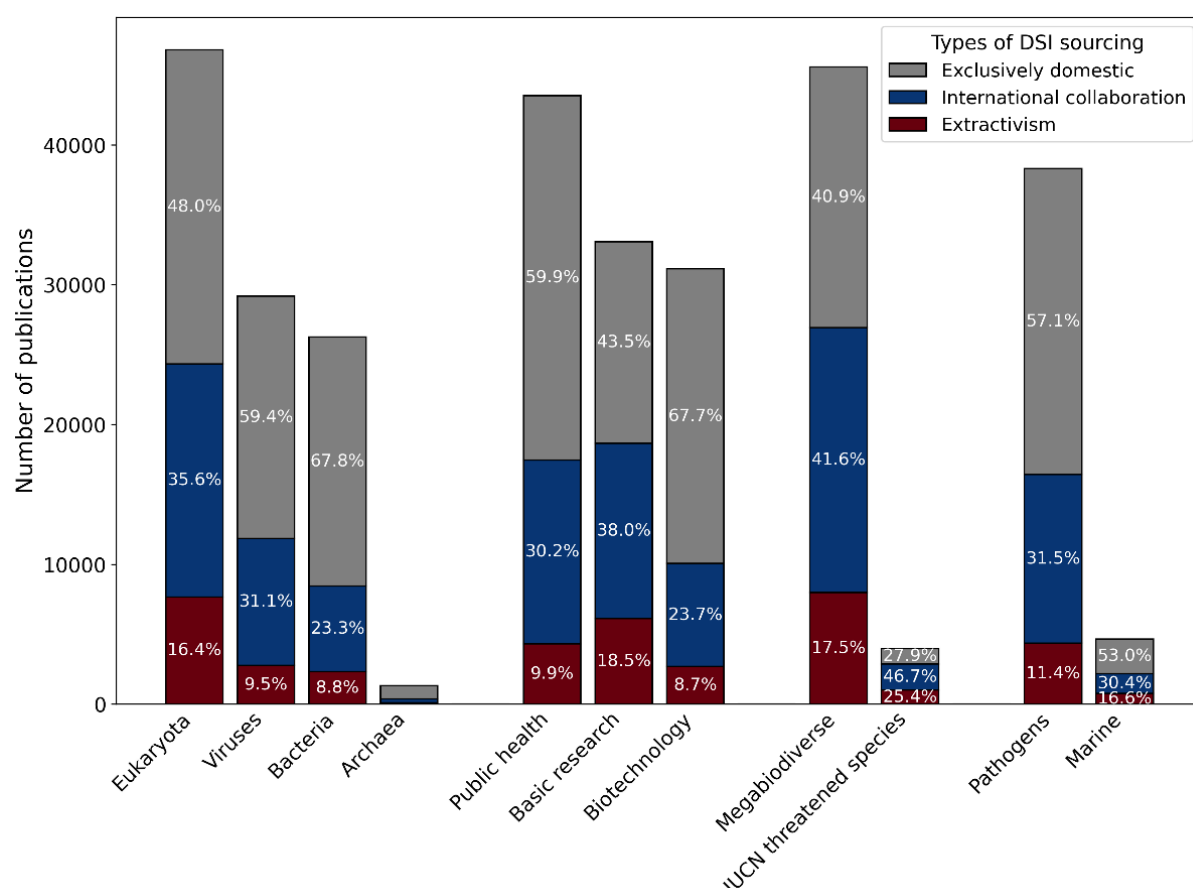


Fig. 2. Distribution of DSI publications and sourcing types across variables. Extractive publications (red) sourced DSI internationally but without involvement of local researchers as authors while collaborative publications (blue) included at least one local researcher. Exclusively domestic publications (grey) only sourced DSI from within their own economic group. Each publication was linked to one unique research type (public health, biotechnology, basic research). The other variables of taxonomic groups (Eukaryota, Viruses, Bacteria, Archaea), regulatory contexts (pathogens, marine), and socio-ecological contexts (mega-biodiverse, IUCN threatened species) were assigned if a publication relied on at least one genetic sequence of that type.

We used logistic regression to understand how these variables affected the statistical likelihood of DSI extractivism in the subset of internationally sourced publications. The probability of extractivism was 9.5 percentage points higher for basic research compared to biotechnology, holding all other variables constant (Fig. 3). In turn, the probability decreased by 3.6 percentage points for public health-related research compared to biotechnology. Sourcing DSI from mega-biodiverse countries increased the probability of extractivism by 4.1, from threatened species by 7, and from marine species by 8.8 percentage points. All effects were significant at a ($p < 0.001$) significance level, apart from pathogens, which showed no significant effect.

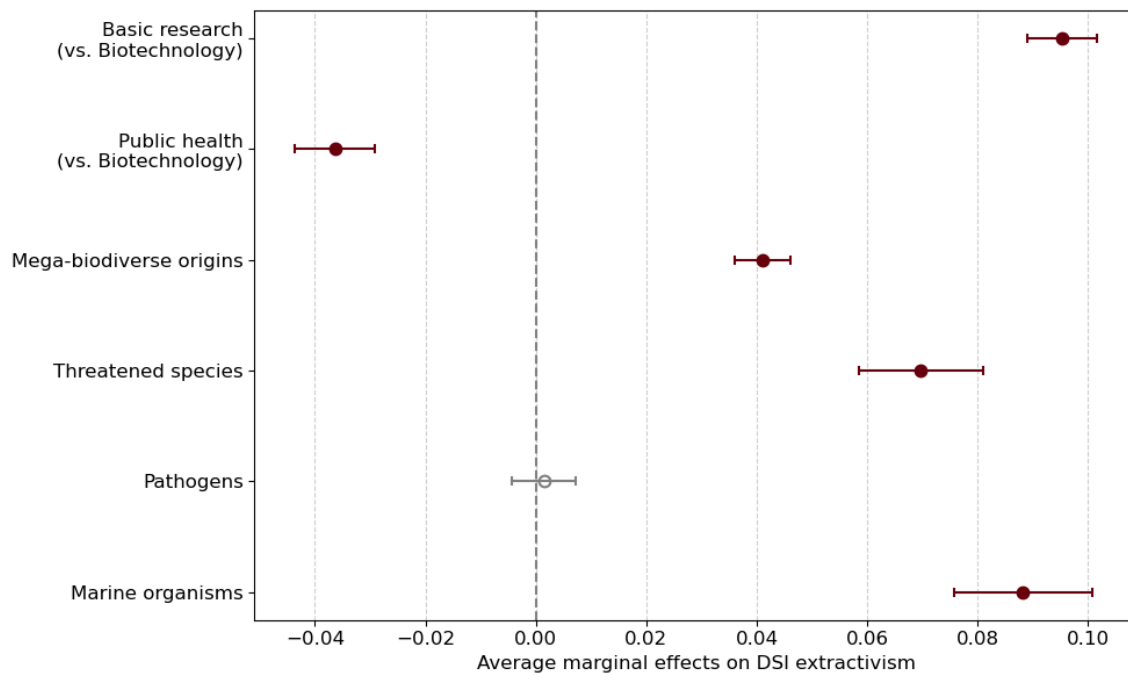


Fig.

3. Average marginal effects of contextual variables on the probability of DSI extractivism. Red markers indicate a significance level of ($p < 0.001$). The effects for basic research and public health were both calculated in comparison to their baseline category, which was biotechnology. The other variables show the effects of a change from their non-condition (e.g. a publication that did not involve marine species) compared to that condition applying. For the underlying regression table, see SI Appendix, Table S2.

Collaborative and extractive sourcing of DSI over time

The share of extractive DSI publications decreased over time (Fig. 4). Out of the 902 publications published in the year 2000, 25% were extractive. In 2020, only 9% out of 3,024 publications were extractive. We found that the share of collaborative publications increased from 14% in 2000 to 29% in 2020. Both international sourcing types account for an average of 35% of all DSI publications over time with some divergences but without any clear linear trends.

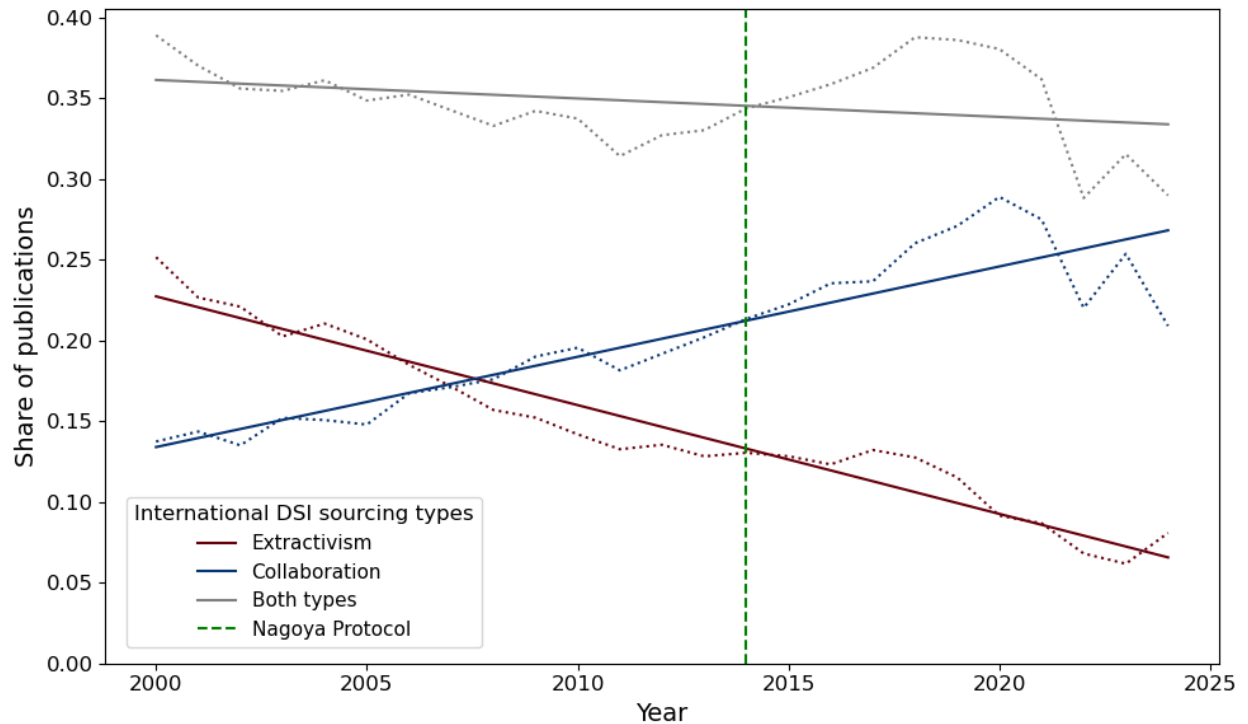


Fig. 4. Extractive (red), collaborative (blue), and total (grey) international DSI sourcing over time, measured in shares of all yearly DSI publications. Dotted lines show exact timelines while strong lines show projected linear trends.

The Nagoya Protocol's entry into force in 2014 did not have a large impact on overall international DSI sourcing nor on the sub-types. Even though the share of extractivism was, on average, lower after 2014 than before, the relatively steady decline since 2000 indicates that factors other than the Protocol shaped this development. Similarly, rates of collaborative publications have steadily increased since 2000 without any apparent impact of the Nagoya Protocol. Collaboration in international DSI sourcing dropped around the start of the Covid-19 pandemic in 2020 and has not recovered since then.

Discussion

Our study introduced a conceptual and empirical framework for studying global information extractivism in the DSI context. We identified publications that reflected extractive relationships and analyzed how these patterns evolved across contexts and time. The following section delves deeper into the observed structural and temporal dynamics and explores their policy implications.

A large share of DSI originates in extractive research practices

Previous studies demonstrated that scientists generally use DSI from within their economic groups, which led to questioning the hypothesis of an imbalanced global provider-user relationship (15, 18). We aimed to complement these observed 'net provision-use balances' by investigating the quality and directionality of cases where DSI was sourced internationally.

We found that around 35% of publications sourced DSI internationally, out of which 40% were extractive. Scientists affiliated with OECD countries were by far the most active in both sourcing DSI internationally and engaging in extractivism. 45% of studies that involved DSI from G77 countries and were conducted by OECD researchers did not include any local authors. These findings suggest that, while scientists generally tend to source DSI domestically, international sourcing remains an important yet potentially problematic aspect of the global DSI landscape.

Basic research is the most international, but also the most extractive research field

Tendencies to rely on international DSI origins differed considerably across scientific fields. While 56.5% of basic research publications relied on international sources, only 40.1% of public health and 32.3% of biotechnology publications did so. Amongst these international DSI publications, the statistical likelihood of extractivism was significantly higher for basic research compared to the more applied research types of biotechnology and public health.

Two main factors likely shape how international a research field's DSI sources are. Firstly, regulations for research of commercial or bio-safety relevance are stricter than for basic research in many countries, including Nagoya Protocol Parties (Brazil, Costa Rica, Ecuador, Peru, Namibia) and non-Parties (e.g. Chile, Colombia) (20, 21). Secondly, basic research often requires the collection of materials from a wide variety of domestic and foreign locations. For instance, this is the case for phylogeography, which seeks to understand species distributions across a variety of geographic contexts. In turn, biotechnological research tends to rely on genetic materials of microbial species that are often highly cosmopolitan, making it less of a necessity for researchers to enter specific (foreign) natural environments (22).

If DSI gets sourced internationally, national funding priorities may play a big role in shaping the likelihood of extractivism across research fields. Investments into basic research are generally considered as 'high risk' in comparison to applied research, which contributes to significant underfunding and to fundamental gaps in our understanding of the planet's biodiversity (23–25). This lack of basic research funding is even stronger in developing economies (26, 27). It might thus be the case that the little funding available in developing countries flows towards the applied research types of public health and biotechnology, positioning local scientists to collaborate more actively in those areas. A competing explanation is that the already mentioned regulatory differences across research fields also affect authorship constellations. Countries like Argentina, Brazil, and Peru put concrete obligations on foreign researchers to cooperate with local institutions and scientists when accessing genetic resources under their jurisdiction (21). However, further research is required to understand how exactly this shapes authorship constellations and, if so, if that affects basic research differently compared to applied research.

Increased likelihood of DSI extractivism in research involving threatened species and mega-biodiverse countries of origin

IUCN-assessed threatened species are a small but highly international source of DSI. We found that 72.1% of publications involving threatened species were sourced internationally. This is not surprising, as most of them live in the Global South (28) while most scientists studying them are based in the Global North (26). The same pattern presumably applies for the publications that sourced materials from mega-biodiverse countries (59.1%).

DSI publications that involved threatened species were significantly more likely to be extractive than publications that did not. This was not necessarily an anticipated finding; many countries have stricter regulations on research in protected areas or on endangered species (29). This could theoretically lead to more rather than less collaboration with local researchers (as is the case for applied research). But it appears that other mechanisms are at play. An explanation may be found in socially or culturally shaped research priorities, such as the known bias towards 'charismatic' and threatened species (30, 31). This specific bias might be more prevalent for OECD-funded research, while funding for BRICS or G77 scientists might follow other trajectories, thereby reducing local involvement in research on threatened species.

The significantly increased probability of extractivism for mega-biodiverse DSI origins fits into the often-voiced narrative by these countries' representatives according to which they provide a considerable amount of DSI to developed countries' researchers without sufficient levels of cooperation. Previous research showed that developing countries often provide access to study sites and data while being dependent on researchers and institutions in the Global North for analysis and dissemination (29, 32). Such global dependencies are particularly strong for mega-biodiverse regions (26), which offers an explanation for these findings.

Marine organisms

47% of the publications related to marine organisms sourced DSI internationally. Their likelihood of extractivism was significantly higher compared to non-marine publications. This finding corresponds to our initial expectations – marine genetic research requires a higher level of research capacity than terrestrial research, such as funding and equipment (23). These capacities are highly concentrated in a small number of developed states (33, 34). It is thus reasonable to assume that economically less developed countries face greater challenges in collaborating in marine research.

DSI extractivism declines over time

Interestingly, and importantly, we observed that DSI extractivism has declined steadily over time. Rates of international research collaborations have simultaneously increased while the overall rate of international DSI sourcing remained relatively stable at 35% out of all DSI publications.

The lack of concrete events that changed these trajectories, besides a dip in collaboration at the time of the Covid-19 pandemic, indicates that the Nagoya Protocol did not have an observable impact.

However, the decline in extractivism could be related to the adoption of the CBD in 1992. A recent study on capacity building in marine genetic research showed that several developing countries substantially increased their capacities after the CBD's adoption in 1992 (35). Similar trends affect biological research more broadly (29), which likely contributed to the observed decline in DSI extractivism.

Beyond regulatory measures, the trends are likely related to economic development, general increases in international scientific collaboration, as well as increased relevance of ethical standards within the scientific community. Economic capacity significantly impacts research capacity in developing countries (35), which, in turn, might shape extractivism and collaboration rates. Furthermore, international cooperation in biological research has generally become stronger in recent decades due to scientists recognizing that more collaboration is needed to improve the quality of scientific research (36, 37). This has made research funding schemes more international, contributed to the global pooling of research capacities, and led to an increase in interdisciplinary research (26), all of which likely contributed to the ratio between extractivism and collaboration in our data.

Our measurement of collaboration applied to any publication that involved at least one local author, which results in higher observed collaboration rates than when applying stricter criteria, such as whether local researchers were involved as lead authors (29). To account for this difference, we took the additional step of contrasting 'foreign lead authors' and 'strong collaboration' types of international DSI publications (SI Appendix, Fig. 4). Although the slope for strong collaboration is lower compared to the collaboration slope in the main text's Figure 4, the overall trajectories of declining extractivism and increasing collaboration remain the same even if using this stricter measurement.

Policy implications

The existence of imbalanced relationships in the international collection and use of DSI poses questions to scientists, funding bodies, and policymakers alike. There is a wide body of literature that stresses the importance of engaging local contexts when conducting research, including its people and infrastructures. Collaborating with local researchers can substantially improve the quality of research questions and increase research impact (11, 38). Relying on local infrastructure can ease sample storage, which is particularly relevant in the study of temperature-sensitive or protected species (11). Local-global partnerships can play a key role in expanding the scope of biodiversity data coverage to currently underrepresented regions (39, 40), particularly in terms of building up the required bioinformatics infrastructures and providing necessary training and technologies (41).

Our findings of high levels of extractivism in basic research may be of particular concern to policymakers working in the nexus of conservation, science and innovation. Basic research activities, such as taxonomy and phylogeography are recognized as the foundations for effective conservation efforts (27). Lacking collaboration in these fields prevents the production and dissemination of essential knowledge to support such conservation efforts in developing countries (42). Additionally, taxonomic research provides essential organismal knowledge for novel biodiscoveries (23, 43). Lacking basic research funding and cooperation thus also hinders social and economic innovation in developing countries.

Policymakers may also need to consider how to not only encourage scientific collaboration as such but rather forms of collaboration that have positive long-term effects on developing countries' scientific capacities and produce locally relevant knowledge (32). Funding bodies in OECD countries often require researchers to engage in collaboration with local researchers but do not necessarily allow funds to be spent on measures that would integrate local researchers and institutions in a co-productive manner (29). For instance, local researchers often help foreign researchers in accessing field sites and in data collection, but are not involved in the conceptualization, analysis, and dissemination phases of research. The research is thereby less likely to contribute to long-term knowledge and capacity building in local contexts (11, 29). This disconnect was visible in our data in the persistent share of exclusively foreign-led publications amongst collaborative publications (SI Appendix, Fig. 4). In this sense, the level of research leadership may be the next frontier in tackling imbalanced forms of global research.

Finally, our study highlights both the political and scientific value of high quality DSI metadata. Studying global-provider relationships, as we did here, requires accurate country of origin as well as taxonomic information. Many future research directions require additional information. For instance, opening the black box of user-provider relations as it relates to the usage of DSI from Indigenous territories requires geo-coordinates and biocultural labelling within metadata (14). Increasing transparent and in-depth metadata is considered key in operationalizing the FAIR and CARE principles, which provide complimentary frameworks that ensure data is both scientifically useful and aligned with Indigenous rights and interests, allowing them to stay connected to data held about their lands, waters and territories (44, 45). Additionally, the recently adopted UNCLOS Treaty on Biodiversity Beyond National Jurisdiction (BBNJ Treaty) introduced obligations on researchers to report a 'BBNJ batch identifier', which is expected to establish more transparency on the usage of marine genetic resources sourced from the High Seas. Beyond the socio-political relevance of metadata, poor metadata standards are harmful to biodiversity research and monitoring efforts (46–48), as effective biodiversity monitoring relies on robust, contextual and inclusive datasets (42). The INSDC has recognized this and is introducing new metadata standards, which are expected to strengthen scientific and regulatory efforts (49).

Limitations and future research

Future research may fruitfully extend our analysis of global DSI extractivism by extending it to a wider range of DSI users. For instance, researchers who cite already published DSI in follow-up scientific publications ('secondary DSI publications') or in patent applications might be integrated into the analysis to construct a more extensive analysis of user-provider relations (18). At the same time, moving from the global perspective to concrete case studies of countries to understand how different types of national regulation or other factors, including funding schemes or collaborative programs, have influenced user-provider relationships over time would be an important step to explore mechanisms for successful policymaking in a systematic manner, a point that was previously noted (35). Finally, building a methodology to analyze user-provider relations in the context of DSI of Indigenous origin would be an

important research endeavor, yet requiring a different ‘sub-national’ operationalization of the concept of DSI extractivism.

Materials and methods

DSI metadata and bibliometric information

We used the continuously updated WiLDSI data portal (17) to download information on a total of 36,770,689 unique genetic sequences (i.e. DSI entries) with given country of origin information (downloaded on July 27th, 2024). The dataset also linked sequences to taxonomic IDs (taxIDs), indicating the species the DSI was sourced from, and to publication IDs. We standardized country of origin names and, where possible, linked them to one of the following economic groups: OECD, BRICS (using their original membership list consisting of Brazil, Russia, India, China, and South Africa), or G77 (SI Appendix, Supporting text 1). To reduce the complexity of the analysis, we prioritized BRICS over G77 and G77 over OECD membership for countries that were members of multiple groups. To subsequently merge this data with bibliometric information, we cut the dataset down to all sequences that were linked to at least one unique publication ID. This resulted in a dataset of 12,040,057 sequences linked to a total of 101,111 unique publications, which can be accessed as SI Appendix, Dataset S1.

We used the scholarly API of the database [The Lens](#) to collect detailed bibliometric information on paper titles, publication years, as well as author information, including their institutional affiliation and affiliation country (SI Appendix, Supporting text 2). We managed to download bibliometric information for 96,842 out of the 101,111 publications with varying levels of coverage for the different bibliometric indicators (downloaded on December 12th, 2024). For authors whose institutional affiliations but not their affiliation countries were known, we text-mined their institution names and linked them to countries using a rules-based approach (43). We then linked all affiliation countries to economic groups following the same procedure as described above for the countries of origin. These steps reduced the final scholarly dataset to 94,772 publications with at least one known author affiliation country and group, which can be accessed as SI Appendix, Dataset S2.

Variables

DSI sourcing type

We merged the scholarly dataset with the DSI metadata and iterated through the list of publications to identify a) whether DSI was sourced from one or several economic groups, and b) whether authors were affiliated with one or several economic groups. If a publication’s authors were exclusively affiliated with one group and exclusively sourced DSI from the same group, the publication was labelled as having a ‘domestic’ DSI source (see Fig. 2). If the authors were exclusively from one group and sourced DSI from any other group, the publication was labelled as ‘extractive’. If author and origin groups did not match, but authors were not exclusively from one group, we labelled the publication source as ‘international collaboration’. To contrast this ‘low threshold’ measurement of collaboration with a stricter

version, we further distinguished collaborative publications into an 'exclusively foreign lead authors' and 'strong collaboration' type for SI Appendix, Fig. S4.

Research types

We used word co-occurrence analysis and network community detection (50) to classify publications in terms of research types, using semantic information contained in their titles (Appendix SI, Supporting text 3). We first pre-processed these titles to cut less meaningful words and to reduce computational run-times, which is a common procedure in computational text analysis (51, 52). This included word lemmatization, stop word removal, removal of country and region names, cutting words that occurred less than twice in the entire corpus, and cutting most generic words using frequency-inverse document frequency (tf-idf) weighting (53). We then constructed a word co-occurrence network by iterating through each processed title, counting all co-occurring terms and assigning a weight of 1 for each encountered connection. For instance, if a specific word-word connection occurred in five titles, the connection would have a weight of 5. Once the co-occurrence network was constructed, we further reduced it to its structurally most meaningful connections by applying network backbone, again to reduce noise and computational run-times (54).

Next, we applied the Leiden community detection algorithm (55) to identify community structures in the word co-occurrence network, a common strategy for finding structures in semantic network data (50, 56). We ran the algorithm at different resolution settings and analyzed the resulting partitions using a combination of modularity maximization and qualitative interpretation to find a statistically viable but also human-readable classification of DSI research types (57). We ultimately chose a partition of three communities, for which we inductively analyzed the highest inner-community weighted terms and determined the labels 'basic research', 'biotechnology', and 'public health' (SI Appendix, Fig. S2). We triangulated this labelling approach with a close reading of publication abstracts for those publication titles that were best matching a community's words, thereby affirming the three labels. We used the same community-title matches to eventually link the research types back to the scholarly dataset in the form of separate binary variables to streamline calculations for the logistic regression analysis. We could not re-link 174 publications, which was due to their titles being left without any words to link to after the data cleaning steps. As such, we continued our analysis with a reduced corpus of 94,598 publications.

Control variables

We linked all taxIDs present in the DSI metadata to taxonomy lineage data of the National Center for Biotechnology Information (NCBI) using NCBI's command-line tool to understand which taxonomic groups a publication sourced DSI from. We managed to collect lineage data on 1,525,119 out of the 1,532,757 unique taxIDs present in the WiLDSI data. 1,327,944 of these taxIDs had a taxonomic rank at the species level and had information filled in at the superkingdom level. We linked the identified categories back to the WiLDSI data via the taxIDs. Around 54% of the 36,002,632 sequences under analysis were of eukaryotic, 30% of viral, 11% of bacterial, and 1% of archaeal origin. Finally, we

evaluated for each publication whether it sourced at least one sequence from any of these taxonomic categories and, if so, labelled a publication accordingly.

We used a pre-defined list of taxIDs linked to pathogens that are likely to be included in a future WHO Pandemic Preparedness Framework, as defined by (2), to identify publications that involved pathogen DSI. We labelled each publication as 'pathogen' if it included at least one sequence of a matching taxID. Next, we searched for DSI that is linked to taxIDs of crop types that are covered by FAO's Seed Treaty following the same approach as for pathogens (2). However, we decided to drop this variable as we found an unreasonably small associated share of crop DSI. This is likely related to a bias specific to crops in the WiLDSI data, as countries of origin rarely get reported for them (2).

We identified marine genetic resources following a name matching approach described in previous work (43), which involves matching species names linked to DSI with the World Register of Marine Species (WoRMS) and, to achieve better coverage of microbial species, an additional list of patented marine genetic sequences (58). The name matching approach was more useful than working with taxIDs in the case of marine organisms, as we were not aware of any pre-defined list of marine genetic resource taxIDs. We also relied on name matching when evaluating whether a publication involved DSI of threatened species, as the IUCN Red List does not use NCBI taxIDs as their primary taxonomic identifiers. We searched for matches between the DSI species names and the IUCN Red List's 'threatened' meta-category, which resulted in 61,914 matched species and 938,325 matched sequences. We considered any of these labels to apply to a publication if it relied on at least one sequence from that category (e.g. if it relied on at least one sequence from a marine species).

We further controlled whether DSI was extracted from a mega-biodiverse country, using the definition of the 'likeminded' mega-biodiverse countries to the CBD (59). This alliance includes Bolivia, Brazil, China, Colombia, Costa Rica, Democratic Republic of Congo, Ecuador, Ethiopia, Guatemala, India, Indonesia, Iran, Kenya, Madagascar, Malaysia, Mexico, Peru, Philippines, South Africa, and Venezuela. Most frequent sample origins were China (46%), Brazil (14%), and India (14%). From the sub-group of mega-biodiverse G77 countries, highest shares were Mexico (24%), Iran (15%), and Malaysia (13%). We assigned the label 'mega-biodiverse origin' if a publication sourced at least one sequence from any country on the list.

The resulting dataset, which is based on the scholarly dataset but includes all variable labels can be accessed as SI Appendix, Dataset S3. SI Appendix, Table S1 provides an overview of the constructed variables.

Logistic regression analysis

We used logistic regression to calculate the effects of research types and control variables on the binary outcome of DSI extractivism for the subset of 33,072 internationally sourced DSI publications. We checked for potential multicollinearity issues, which led us to exclude the taxonomic categories due to their relatively high correlations with research types (SI Appendix, Fig. S3). We chose 'biotechnology'

as the baseline against which to calculate effects for public health and basic research. Changing the baseline category did not affect the results. For instance, if using public health as a baseline, biotechnology would show the same strength of effect as public health does, just in the opposite direction. We visualized the findings in terms of average marginal effects, keeping all other variables constant at 0 (Fig. 3). SI Appendix, Table S2 shows the regression results.

Data and code availability

All data used in this study is accessible as SI Appendix, Datasets S1-S3. All Python code used in this study is available at: <https://github.com/pduns/global-dsi-extractivism.git>.

Acknowledgments

Paul Dunshirn is funded by the Research Platform Governance of Digital Practices at the University of Vienna. Paul Dunshirn and Alice Vadrot are funded by the European Research Council project TwinPolitics, project number 101124903; grant holder: Alice Vadrot. Erik Zhivkoplis is funded by FORMAS, project number 2020-01048. Katie Lee Riddle is funded by the Tikanga in Technology project of the Ministry of Business and Enterprise, New Zealand. The funders played no role in the study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Literature

1. C. Hess, E. Ostrom, *Understanding knowledge as a commons: from theory to practice* (MIT Press, 2007).
2. S. Sett, *et al.*, Harmonize rules for digital sequence information benefit-sharing across UN frameworks. *Nat. Commun.* **15**, 8745 (2024).
3. International Chamber of Commerce, Digital Sequence Information and the Nagoya Protocol. (2017).
4. B. E. Kreiken, B. J. M. Arts, 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).
5. K. Raustiala, D. G. Victor, The Regime Complex for Plant Genetic Resources. *Int. Organ.* **58** (2004).
6. A. Vadrot, *The Politics of Knowledge and Global Biodiversity*, 0 Ed. (Routledge, 2014).
7. F. Rabitz, *The Global Governance of Genetic Resources: Institutional Change and Structural Constraints* (Routledge, 2017).
8. M. R. Bond, D. Scott, Digital biopiracy and the (dis)assembling of the Nagoya Protocol. *Geoforum* **117**, 24–32 (2020).
9. M. A. Bagley, “De-materializing genetic resources: Synthetic biology, intellectual property and the ABS bypass” in *Routledge Handbook of Biodiversity and the Law*, (Routledge, 2017).
10. S. Laird, *et al.*, Rethink the expansion of access and benefit sharing. *Science* **367**, 1200–1202 (2020).

11. D. Haelewaters, T. A. Hofmann, A. L. Romero-Olivares, Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South. *PLOS Comput. Biol.* **17**, e1009277 (2021).
12. P. V. Stefanoudis, *et al.*, Turning the tide of parachute science. *Curr. Biol.* **31**, R184–R185 (2021).
13. M. Bockarie, S. Machingaidze, T. Nyirenda, O. F. Olesen, M. Makanga, Parasitic and parachute research in global health. *Lancet Glob. Health* **6**, e964 (2018).
14. J. Golan, *et al.*, Benefit sharing: Why inclusive provenance metadata matter. *Front. Genet.* **13** (2022).
15. F. Rohden, S. Huang, G. Dröge, A. H. Scholz, “Combined study on digital sequence information in public and private databases and traceability” (2020).
16. W. Houssen, R. Sara, M. Jaspars, “Digital sequence information on genetic resources: concept, scope and current use” (2020).
17. M. Lange, *et al.*, Quantitative monitoring of nucleotide sequence data from genetic resources in context of their citation in the scientific literature. *GigaScience* **10**, giab084 (2021).
18. A. H. Scholz, *et al.*, Myth-busting the provider-user relationship for digital sequence information. *GigaScience* **10**, giab085 (2021).
19. C. W. Chagnon, *et al.*, From extractivism to global extractivism: the evolution of an organizing concept. *J. Peasant Stud.* **49**, 760–792 (2022).
20. E. Chege Kamau, Ed., *Global Transformations in the Use of Biodiversity for Research and Development: Post Nagoya Protocol Implementation Amid Unresolved and Arising Issues* (Springer International Publishing, 2022).
21. Y. C. Colmenarez, *et al.*, Regulatory Frameworks for the Access and Use of Genetic Resources in Latin America. *Neotrop. Entomol.* **52**, 333–344 (2023).
22. H. Salem, M. Kaltenpoth, The Nagoya Protocol and its implications for microbiology. *Nat. Microbiol.* **8**, 2234–2237 (2023).
23. J. D. Sigwart, R. Blasiak, M. Jaspars, J.-B. Jouffray, D. Tasdemir, Unlocking the potential of marine biodiscovery. *Nat. Prod. Rep.* **38**, 1235–1242 (2021).
24. M. S. Engel, *et al.*, The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. *Zool. J. Linn. Soc.* **193**, 381–387 (2021).
25. Q. Tahseen, Taxonomy-The Crucial yet Misunderstood and Disregarded Tool for Studying Biodiversity. *Biodivers. Endanger. Species* (2014). <https://doi.org/10.4172/2332-2543.1000128>.
26. L. Tydecks, J. M. Jeschke, M. Wolf, G. Singer, K. Tockner, Spatial and topical imbalances in biodiversity research. *PLOS ONE* **13**, e0199327 (2018).
27. A. Ahrends, *et al.*, Funding begets biodiversity. *Divers. Distrib.* **17**, 191–200 (2011).
28. M. B. J. Harfoot, *et al.*, Using the IUCN Red List to map threats to terrestrial vertebrates at global scale. *Nat. Ecol. Evol.* **5**, 1510–1519 (2021).
29. J. C. Habel, *et al.*, Towards more equal footing in north–south biodiversity research: European and sub-Saharan viewpoints. *Biodivers. Conserv.* **23**, 3143–3148 (2014).
30. J. Troudet, P. Grandcolas, A. Blin, R. Vignes-Lebbe, F. Legendre, Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* **7**, 9132 (2017).

31. I. E. Hendriks, C. M. Duarte, Allocation of effort and imbalances in biodiversity research. *J. Exp. Mar. Biol. Ecol.* **360**, 15–20 (2008).
32. N. Boshoff, Neo-colonialism and research collaboration in Central Africa. *Scientometrics* **81**, 413–434 (2009).
33. P. C. Sierra-Correa, *et al.*, “Research capacity and infrastructure” in *Global Ocean Science Report 2020: Charting Capacity for Ocean Sustainability*, (United Nations, 2020).
34. A. Lara-Lopez, L. Valdés, R. de Pinho, H. Enevoldsen, “Analysis of ocean science production and impact” in *Global Ocean Science Report 2020: Charting Capacity for Ocean Sustainability*, (United Nations, 2020).
35. M. C. Leal, *et al.*, Fifty years of capacity building in the search for new marine natural products. *Proc. Natl. Acad. Sci.* **117**, 24165–24172 (2020).
36. R. H. G. Jongman, A. K. Skidmore, C. A. Múcher, R. G. H. Bunce, M. Metzger, “Global Terrestrial Ecosystem Observations: Why, Where, What and How?” in *The GEO-Handbook on Biodiversity Observation Networks*, (Springer, 2017), pp. 19–38.
37. R. F. Noss, Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conserv. Biol.* **4**, 355–364 (1990).
38. A. M. Mc Cartney, *et al.*, Indigenous peoples and local communities as partners in the sequencing of global eukaryotic biodiversity. *Npj Biodivers.* **2**, 1–12 (2023).
39. H. Farooq, J. A. R. Azevedo, A. Soares, A. Antonelli, S. Faurby, Mapping Africa's Biodiversity: More of the Same Is Just Not Good Enough. *Syst. Biol.* **70**, 623–633 (2021).
40. R. T. Corlett, A Bigger Toolbox: Biotechnology in Biodiversity Conservation. *Trends Biotechnol.* **35**, 55–65 (2017).
41. E. Zhivkoplías, J. da Silva, R. Blasiak, How smarter biotechnology use will unlock biodiversity research. *Trends Biotechnol.* (Forthcoming).
42. H. S. Kühl, *et al.*, Effective Biodiversity Monitoring Needs a Culture of Integration. *One Earth* **3**, 462–474 (2020).
43. P. Dunshirn, E. Zhivkoplías, Conducting marine genetic research for whom? Mapping knowledge flows from science to patents. *Npj Ocean Sustain.* **3**, 1–11 (2024).
44. S. R. Carroll, E. Herczog, M. Hudson, K. Russell, S. Stall, Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Sci. Data* **8**, 108 (2021).
45. S. Rainie, *et al.*, Issues in Open Data: Indigenous data sovereignty. (2019).
46. R. H. Toczydlowski, *et al.*, Poor data stewardship will hinder global genetic diversity surveillance. *Proc. Natl. Acad. Sci.* **118**, e2107934118 (2021).
47. S. Hoban, *et al.*, Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biol. Rev.* **97**, 1511–1538 (2022).
48. J. Wieczorek, *et al.*, Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE* **7**, e29715 (2012).
49. International Nucleotide Sequence Database Collaboration, INSDC spatiotemporal metadata – minimum standards update (03-03-2023). (2023).

50. L. Leydesdorff, A. Nerghe, Co-word maps and topic modeling: A comparison using small and medium-sized corpora ($N < 1,000$). *J. Assoc. Inf. Sci. Technol.* **68**, 1024–1035 (2017).
51. S. P. Borgatti, E. Quintane, Techniques: Dichotomizing a Network. *Connections* **38**, 1–11 (2019).
52. L. Price, M. Thelwall, The clustering power of low frequency words in academic Webs. *J. Am. Soc. Inf. Sci. Technol.* **56**, 883–888 (2005).
53. S. Qaiser, R. Ali, Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *Int. J. Comput. Appl.* **181** (2018).
54. M. Coscia, F. M. H. Neffke, Network Backboning with Noisy Data in *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, (2017), pp. 425–436.
55. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
56. F. Rabitz, A. Olteanu, J. Jurkevičienė, A. Budžytė, A topic network analysis of the system turn in the environmental sciences. *Scientometrics* **126**, 2107–2140 (2021).
57. S. Fortunato, D. Hric, Community detection in networks: A user guide. *Phys. Rep.* **659**, 1–44 (2016).
58. E. Zhivkoplías, J.-B. Jouffray, P. Dunshirn, A. Pranindita, R. Blasiak, Growing prominence of deep-sea life in marine bioprospecting. *Nat. Sustain.* **7**, 1027–1037 (2024).
59. Convention on Biological Diversity, Declaration by the Group of Likeminded Megadiverse Countries (LMMC) at the fifteenth meeting of the Conference of Parties of the Convention for Biological Diversity (CBD COP15). (2022).

7. Paper 2: Conducting marine genetic research for whom? Mapping knowledge flows from science to patents

<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.

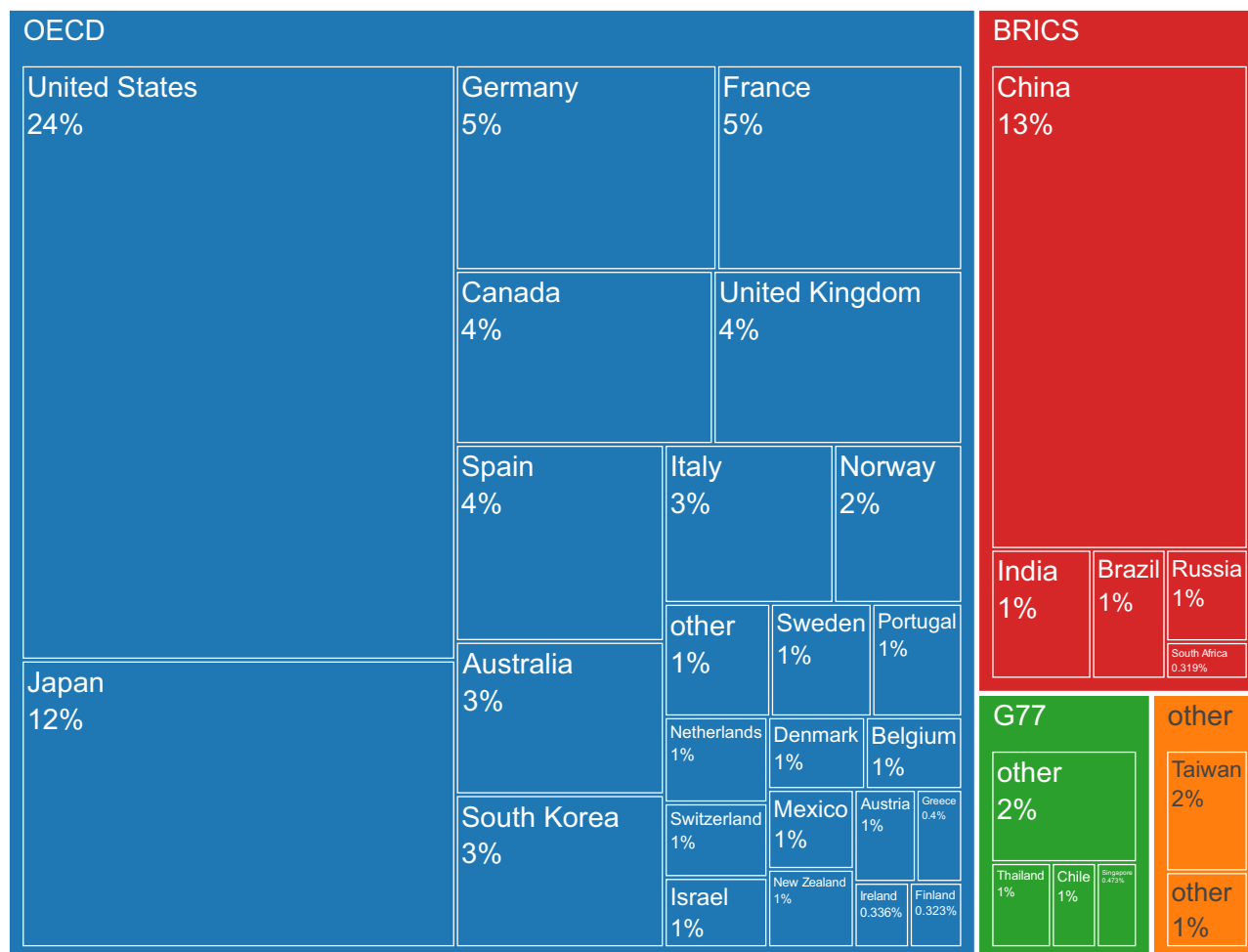


Fig. 1 | Marine genetic studies per country and economic group. Box sizes represent shares of total publications.

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

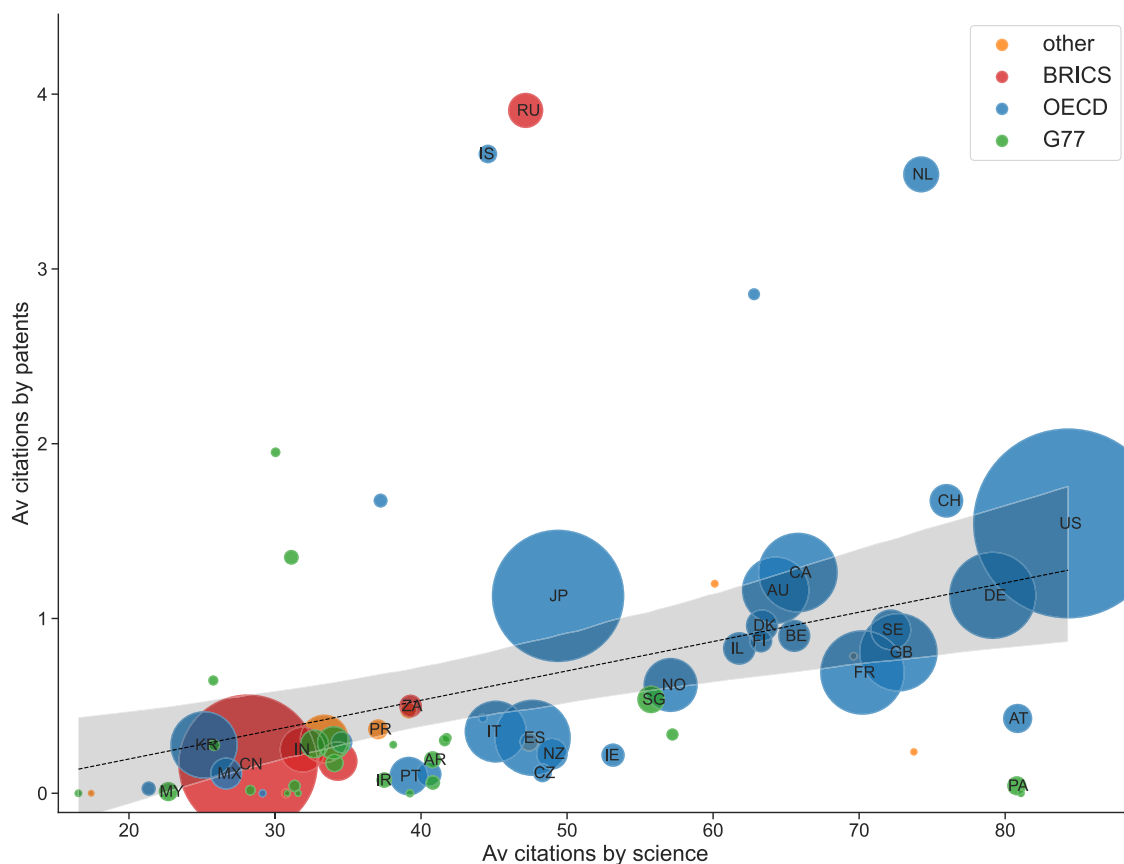


Fig. 2 | Relationship between average citations by secondary science and average citations by patents per country. Figure is based on countries with more than 5 publications ($n = 70$). Node size displays the number of publications and color displays economic group. 95% confidence intervals are displayed. 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 Prokaryote in Iceland (<https://prokaryote.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 'cion' (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

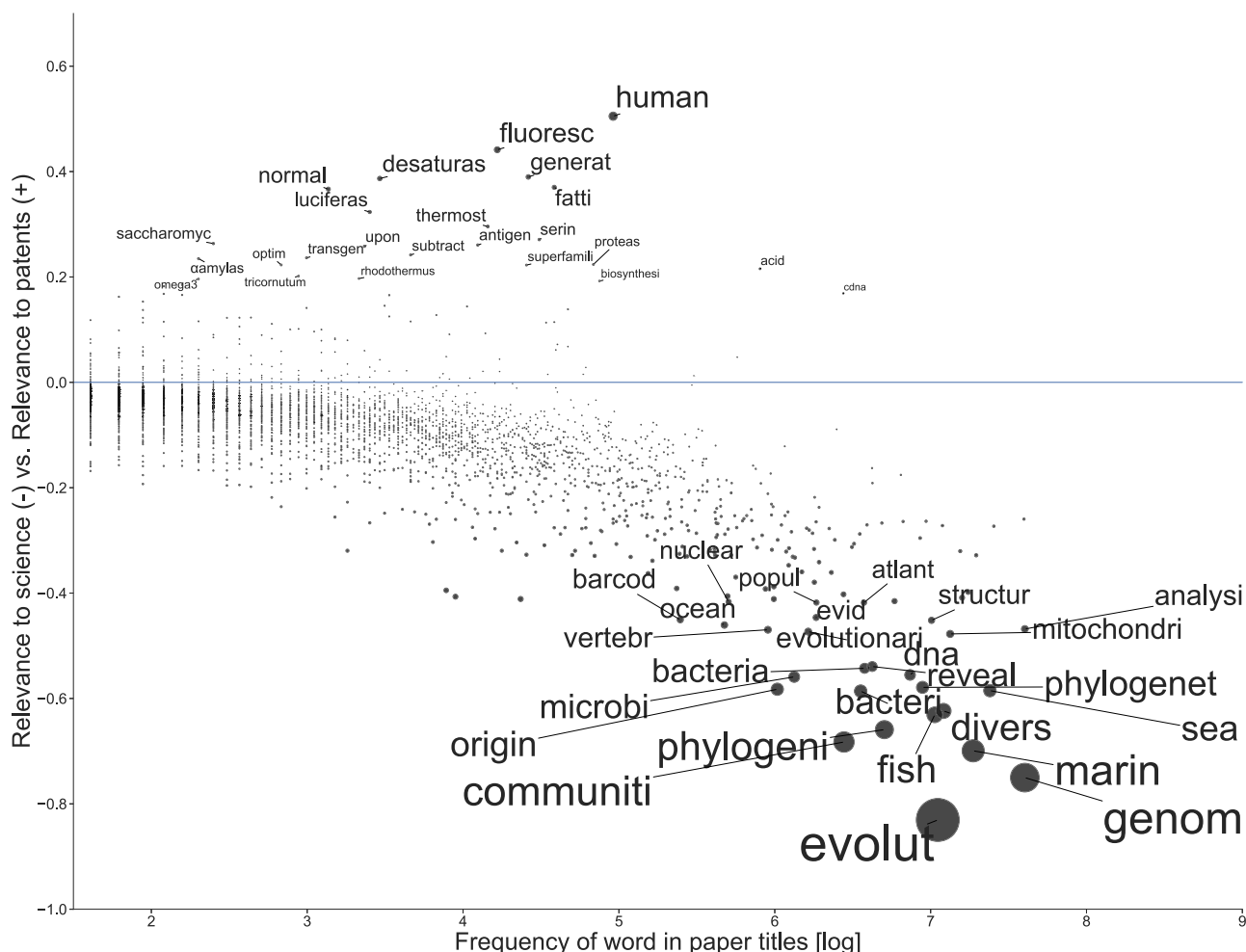


Fig. 3 | Relevance to science versus relevance to patents for words in paper titles. Node size corresponds to how polarized a word was according to the relevance indicator. The indicator ranges from -1 , indicating exclusive relevance to science, to

$+1$, 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 future innovation 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

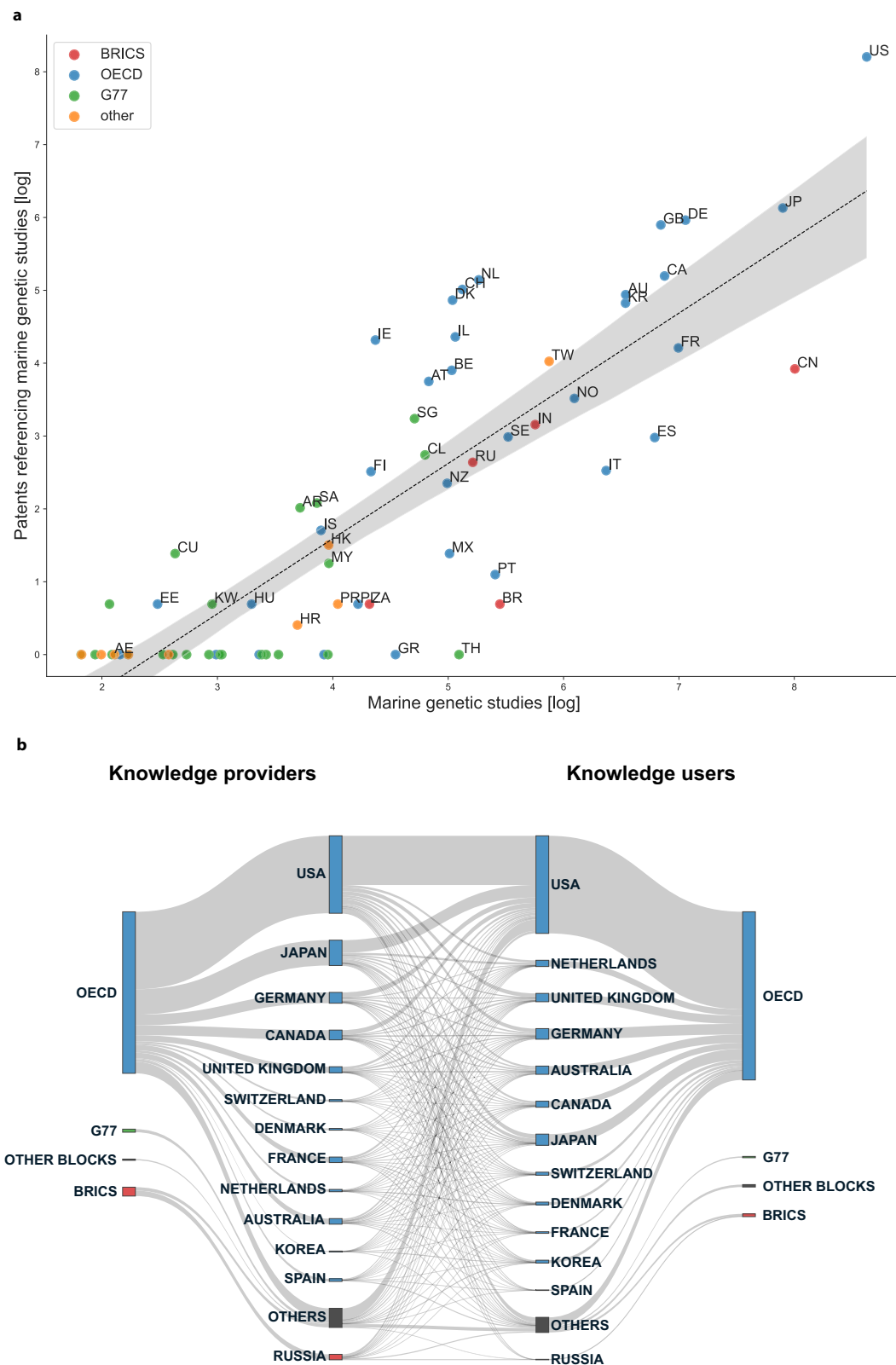


Fig. 4 | National science-patent relations. **a** Comparing counts of marine genetic studies and patents for countries with more than 5 publications ($N = 70$). Counts were transformed using $\log(x + 1)$ and $\log(y + 1)$ to avoid hiding countries without patents. A 95% confidence interval is displayed. 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

8. Paper 3: Growing prominence of deep-sea life in marine bioprospecting

Growing prominence of deep-sea life in marine bioprospecting

Received: 3 July 2023

Accepted: 25 June 2024

Published online: 8 August 2024



Erik Zhivkoplías¹✉, Jean-Baptiste Jouffray^{1,2}, Paul Dunshirn³, Agnes Pranindita^{1,4} & Robert Blasiak^{1,5}

Marine bioprospecting, which involves the exploration of genetic and biochemical material from marine organisms, can be used towards addressing a broad range of public and environmental health applications such as disease treatment, diagnostics and bioremediation. Marine genetic resources are important reservoirs for such bioprospecting efforts; however, the extent to which they are used commercially for natural product discovery and the marine sources from which they are derived are not well understood. Here we introduce a comprehensive database of marine genes referenced in patent filings, the Marine Bioprospecting Patent database. It includes 92,550 protein-coding sequences associated with 4,779 patent filings, identified by analysing all relevant records from genetic sequence databases. Three companies alone—BASF, IFF and DuPont—included sequences from 949 species (more than half of referenced species with identified marine origin). Microbial life in the deep sea, a vast and remote biome predominantly beyond national jurisdiction, is already attracting substantial economic interest; the top ten patent holders have all filed marine gene patents referencing sequences from deep-sea life. Our findings provide an updated understanding of the marine bioprospecting landscape, contribute to the sustainable use of marine biodiversity and underscore the need for policymakers to ensure stewardship of deep-sea ecosystems.

Biodiscovery—the exploration and use of genetic and biochemical properties of biological materials—has a long and rich history. For instance, centuries before the discovery of penicillin from mould in a laboratory, skin diseases were already being treated in the Kingdom of Jordan via red soils with potent antibacterial properties that have only recently been confirmed¹. Other examples include traditional medicines extracted from evergreen shrubs for cancer treatment², derivatives of the foxglove plant used to treat heart problems³, anti-malarial quinine⁴ and fungi-extracted podophyllotoxin to treat sexually transmitted diseases⁵. However, recent advances in genetics and sequencing innovations have spurred an unprecedented growth in the scale of discoveries. Today, bioprospecting—the search for potential products with scientific and industrial value derived from biological

resources such as animals, plants and microorganisms—often involves large-scale screening, analysis and prediction of prospective biological compounds through the exploration of databases with sequencing data, including DNA extracted directly from environmental samples⁶.

In this context, the ocean is considered a promising but largely untapped frontier for biodiscovery⁷. Marine organisms have evolved over millions of years to adapt to extreme conditions of temperature, salinity, light, pressure and water flow⁸. These conditions as well as a far longer evolutionary history have contributed to substantially greater taxonomic and functional diversity in marine habitats than in other biomes⁹. Nearly one million eukaryotic species are believed to inhabit the ocean¹⁰, and the number of archaea and bacteria may be ten thousand times higher¹¹, yet most remain undescribed by science.

¹Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden. ²Stanford Center for Ocean Solutions, Stanford, CA, USA. ³Research Platform Governance of Digital Practices, University of Vienna, Vienna, Austria. ⁴Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden.

⁵Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. ✉e-mail: erik.zhivkoplías@su.se

Despite these knowledge gaps, marine biotechnology—the use of marine organisms and their compounds for a wide range of applications in industrial sectors—has managed to distinguish itself from the broader biotechnology landscape. For instance, while nearly half of the approved pharmaceuticals are based on biological compounds produced by living organisms, success rates are two to four times higher for compounds from marine organisms^{7,12}. Annual sales and licensing revenues from marine drugs have exceeded US\$1 billion annually since 2011¹³, and prospects for greater commercial growth are substantial: in 2020 alone, more than 1,400 new compounds were isolated from marine species¹⁴. Biomolecules extracted from marine bacteria and other products developed from sequences of larger marine organisms are widely used in food production, diagnostics, bioremediation and disease treatment¹⁵. Some notable examples include the discovery of a thermostable enzyme required for the production of lactose-free milk in *Archaea Pyrococcus furiosus*¹⁶, seawater cyanobacteria toxins developed into anticancer treatment products¹⁷ and the extensive use of green fluorescent protein found in jellyfish *Aequorea victoria*¹⁸ as a molecular marker, both in medical and diagnostic contexts and fundamental research.

Establishing a regulatory landscape that keeps pace with rapid advances in biotechnology, while also promoting transparency, equitable access and benefit-sharing mechanisms, has proven challenging¹⁹. The adoption of the Convention on Biological Diversity (CBD) in 1993 was a crucial milestone, as it defined genetic resources as ‘any material of marine plant, animal, microbial or other origin containing functional units of heredity of actual or potential value’, and established the fair and equitable sharing of benefits from their use as one of the convention’s three core objectives²⁰. In 2014, the convention’s Nagoya Protocol provided a framework to regulate the access and benefit sharing of marine genetic resources (MGR) sampled in national jurisdictions²¹. Yet, some two-thirds of the ocean lies beyond national jurisdiction, and it was not until 2023, following protracted negotiations, that the ‘High Seas Treaty’ was agreed upon, including provisions to address MGR from areas beyond national jurisdiction (ABNJ)²².

Despite these encouraging developments, the actual and potential value of MGR for marine bioprospecting remains poorly understood. Studies have focused on counting referenced marine species in patents²³ or GenBank²⁴, examining sequences in international patent applications^{25–28} or exploring biological compounds for natural product discovery^{29,30}. A common aspect to all these studies, however, is their lack of focus on the connection between the actors involved in the use of MGR and the potential sources for natural product discovery. They also suffer from limited information in patent and GenBank records about the geographical origin of gene sequences, which in many cases are referenced without naming the source species. The unevenness of these data presents a challenge for interpreting the true scale, scope and trajectory of marine bioprospecting.

Here we address these gaps by creating a comprehensive database of genetic sequences and related patent applications from 1989 to 2022 in marine bioprospecting. In addition to systematically compiling and presenting key data about the sequences, coded proteins, date of deposition and patent holders, we also address significant data gaps by developing and applying a BlastX sequence similarity model to consider sequences from unnamed species. We also assess the biodiversity data of species currently considered unique to ABNJ and highlight the special importance of deep-sea conservation for future biotechnology focused on the innovation and development of naturally derived products.

Results

Our analysis of patent filings revealed 29,065 nucleotide sequences from 1,474 disclosed marine species across 3,636 unique patents, representing approximately 1% of all gene patents submitted to the International Nucleotide Sequence Database Collaboration (INSDC).

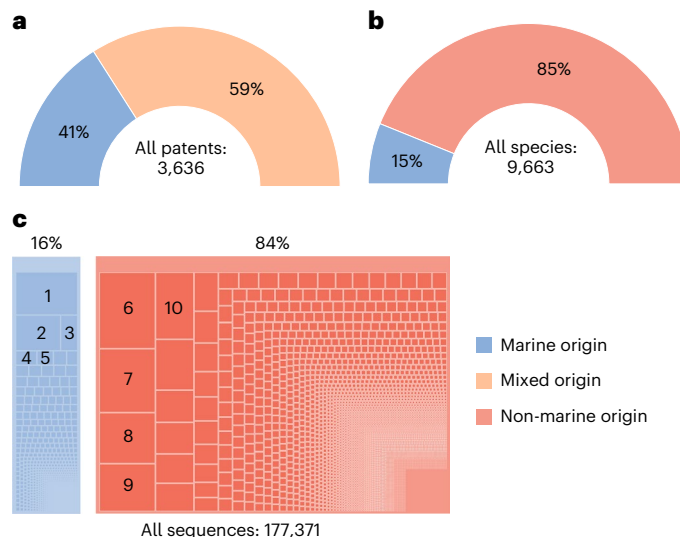


Fig. 1 | Patent applications associated with marine species. **a**, Number of patents that contain sequences associated with species of marine origin only, and patents with species from both marine and mixed origins. **b**, Number of unique marine and non-marine species referenced in patent applications that include at least one sequence of marine origin. **c**, Sequence frequency aggregated by species they originate from. Only 16% of all sequences attached to patent applications originate from marine species. The top five species with the highest frequency in each class are indicated. Marine species: (1) *C. intestinalis* (2.96%), (2) *Gadus morhua* (1.76%), (3) *Anguilla japonica* (0.67%), (4) *Salmo salar* (0.34%) and (5) *Oncorhynchus mykiss* (0.26%). Non-marine species: (6) *Arabidopsis thaliana* (4.31%), (7) *Zea mays* (3.61%), (8) *Glycine max* (2.87%), (9) *Homo sapiens* (2.71%) and (10) *Oryza sativa* (2.60%).

Many patents referenced multiple sequences, with a majority including both marine and non-marine sequences (Fig. 1a). Overall, marine sequences and species represented only 16% and 15%, respectively, of all sequences and species identified within the 3,636 patents (Fig. 1b,c). For comparison, approximately 242,000 marine species have been described to date (World Register of Marine Species (WoRMS), 2022), corresponding to roughly 10% of the 2.1 million species described by science³¹. This suggests considerable untapped potential of marine bioprospecting (Fig. 1b and Supplementary Table 1).

Types of sequence in marine gene patents

The patent applicants who referenced the highest number of unique genetic sequences included both protein-coding and non-coding sequences, with the former having a higher potential for natural product discovery (Fig. 2). Most of the companies with a large number of applications referenced protein-coding genes that originate from multiple species, with an average length between 500 and 2,000 nucleotides. Some applicants specifically focused on MGR from a single species and predominately referenced non-coding sequences. For instance, the Fisheries Research Agency of the National Research and Development Agency in Japan included 1,179 sequences in their patent applications, mostly originating from Japanese eel (*Anguilla japonica*), yet only 127 are protein-coding sequences. Similarly, the Japan Science and Technology Agency has referenced 5,190 sequences from the sea vase tunicate (*Ciona intestinalis*), only 150 of which are protein-coding genes.

Most short non-coding sequences of identical length, originating from the same species, exhibit a wide range of GC content (that is, the percentage of two DNA basic building blocks), which is typical for artificially modified sequences used in amplification or as probes for detecting specific sequences of DNA or RNA (Supplementary Fig. 1). Out of all the patents that include at least one sequence from disclosed marine species, 71% contain nucleotide sequences that are potentially

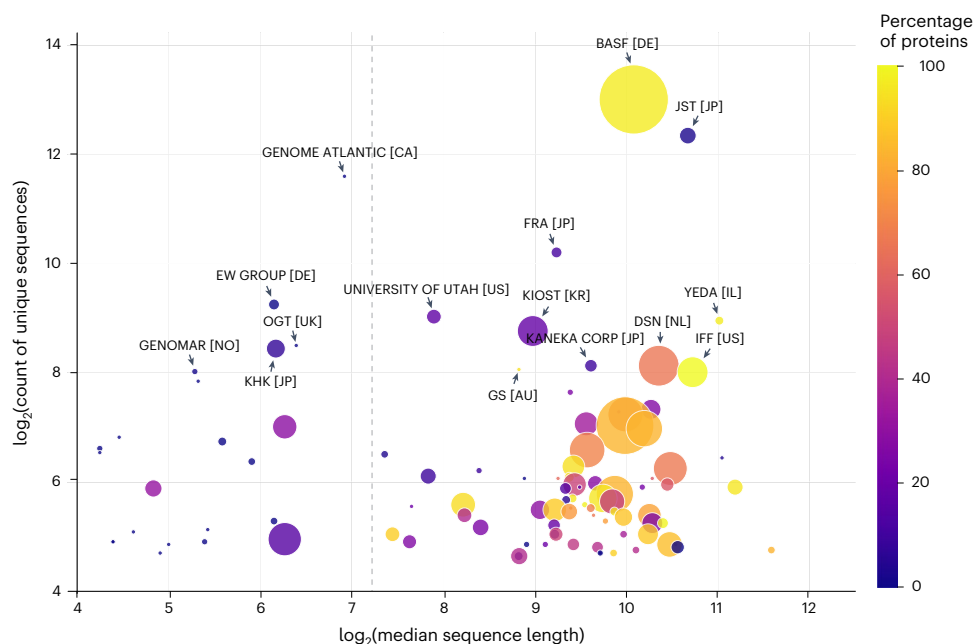


Fig. 2 | Key actors in marine biotechnology. Applicants that submitted at least 25 nucleotide sequences in their patent claims (81% of all sequences) are shown. Companies that submitted at least 250 sequences are indicated. The size of the dots represents the number of patents submitted by each applicant. The dotted grey line indicates the shortest protein length estimation (150 base pairs). The continuous colour bar indicates the percentage of protein-coding sequences submitted in applicant claims. BASF [DE], BASF; JST [JP], Japan Science

and Technology Agency; GENOME ATLANTIC [CA], Genome Atlantic; FRA [JP], Fisheries Research Agency; EW GROUP [DE], EW Group GmbH; UNIVERSITY OF UTAH [US], The University of Utah; YEDA [IL], Yeda Research and Development Company Ltd.; KIOST [KR], The Korea Institute of Ocean Science and Technology; DU PONT [US], DuPont; OGT [UK], Oxford Gene Technology Ltd.; KHK [JP], Kyowa Kirin Co., Ltd.; DSM [NL], DSM N.V.; KANEKA CORP [JP], Kaneka Corporation; GS [AU], Gene Stream Pty Ltd.; GENOMAR [NO], GenoMar.

protein-coding genes. This suggests that most MGR are used in bioprospecting (Fig. 3a). For sequences of particular interest (that is, those submitted to all patent systems), we provide examples illustrating the conversion of DNA molecules into products of value (Box 1).

Marine Bioprospecting Patent database

While INSDC records provide considerable insight into the genes referenced in patents, only 37.3% of records include the name of source species, primarily filed under the World Intellectual Property Organization (WIPO), the European Patent Office, the Patent Office of Japan and the Korean Intellectual Property Office. Most of the remaining records are from the US Patent and Trademark Office, which does not share species names in its records (Supplementary Fig. 2).

To address this gap, we developed a sequence similarity model and BlastX search tool to query all genetic sequences with unknown origins against the UniProtKB protein sequence database. This model retrieved an additional 60,636 sequences, which can be said with a high degree of certainty to originate from marine organisms. Together with the 31,914 protein-coding sequences of confirmed marine species, this resulted in a comprehensive database of 92,550 sequences, which form the basis for all subsequent analysis in this paper and were used to construct the Marine Bioprospecting Patent (MABPAT) database (<https://mabpat.shinyapps.io/main/>).

Key actors in marine biotechnology

We found that 100 applicants accounted for 58% of all patents that contain protein-coding sequences with identified marine origin (that is, bioprospecting patents). The remaining 42% were associated with applicants who filed fewer than two patents on average. For companies in the top 100 (Supplementary Table 3), the total number of patent applications would have been underestimated by at least one-third if we had not applied the sequence similarity model. Transnational corporations (1,675 applications) are the most frequent type of applicant,

although roughly one-fifth of filings are from research institutes and their commercialization centres (634 applications) (Fig. 3b). In total, 78% of all bioprospecting patents filed by the top 100 were submitted by actors headquartered in the USA, Germany or Japan (Fig. 3c).

The number of patents registered by each applicant is correlated with the total count of unique species included in such patents ($r = 0.8168$, $P = 2.17552 \times 10^{-318}$). To illustrate how much biological diversity each of these applicants is drawing upon, we connected patent holders and unique species included in patent claims and aggregated on the domain (Fig. 4a) and phylum level of biological taxonomy (Supplementary Fig. 6). For each flow diagram, we also added information if the corresponding marine species had been observed in a deep-sea environment. The most active users of MGR are primarily dependent on sequences from bacteria and archaea (Fig. 4a). The ten largest actors, including eight multinational corporations and two public research bodies (Fig. 4a), collectively registered more than one-third of all patents in the top 100. Deep-sea marine species have attracted interest from all ten of the largest users of MGR.

The opacity of marine bioprospecting in ABNJ

Issues of access and benefit sharing related to genetic material from ABNJ are of particular interest as they fall outside the scope of the Nagoya Protocol of the Convention on Biological Diversity and were at the core of negotiations for the High Seas Treaty adopted in June 2023. It is therefore notable that among 1,639 species of identified marine origin referenced in INSDC patent records, 281 species have been observed in ABNJ, with only 5 of them being exclusive. This contrasts with the 5,889 species found exclusively in ABNJ, predominantly from the Arthropoda, Foraminifera and Nematoda phyla, according to our analysis of species observation data available in the Ocean Biodiversity Information System (OBIS), a global open-access database on marine biodiversity (<https://obis.org>). The complete taxonomic distribution is given in Supplementary Fig. 7. According to the records from the World

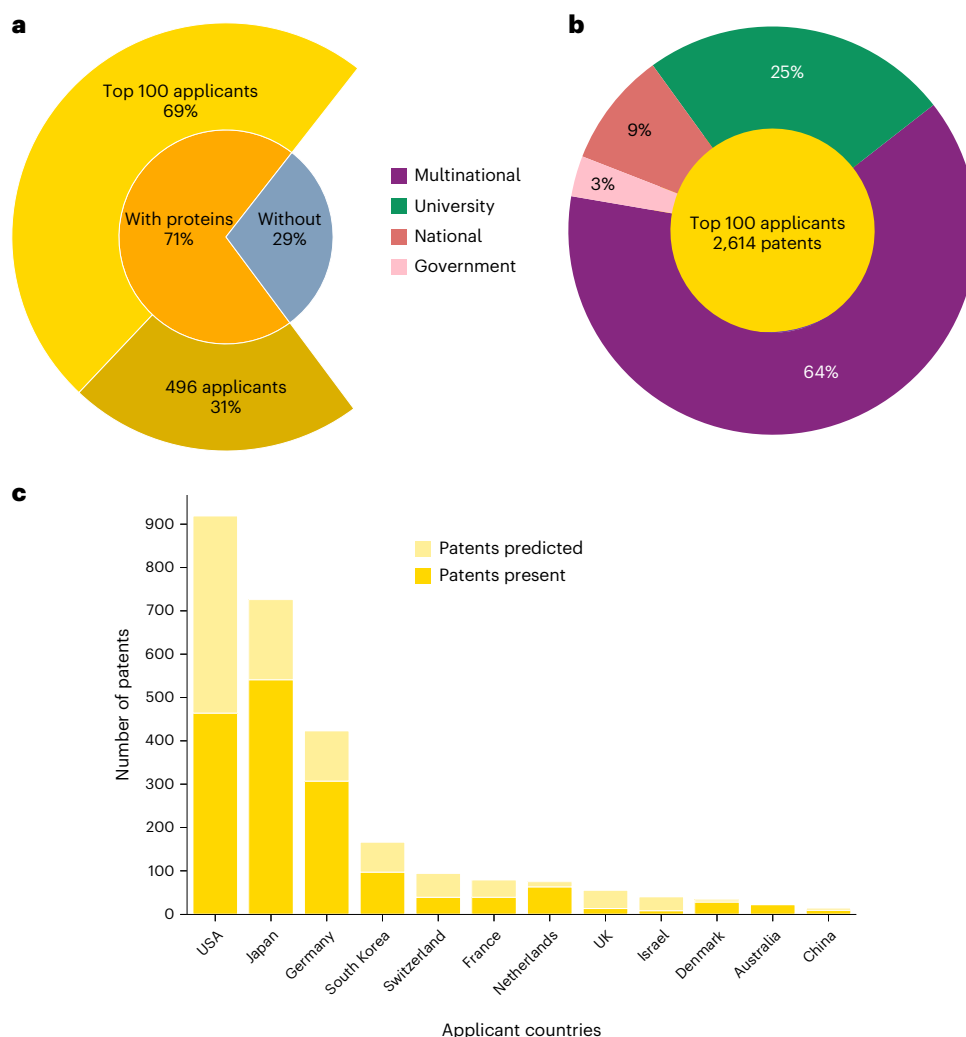


Fig. 3 | Patent applications in global marine bioprospecting. **a**, Share of companies that submitted patents with at least one protein-coding sequence (bioprospecting patents) and patents with non-coding sequences only. **b**, Top 100 largest patent applicants in marine bioprospecting, aggregated by applicant

type. The terms ‘multinational’ and ‘national’ denote the extent of company presence in more than two countries or less, respectively. **c**, Top 100 largest patent applicants in marine bioprospecting, aggregated by country of origin (the country of headquarters).

Register of Deep-Sea Species (WoRDS), 39% of marine species were exclusively found to inhabit deep-sea environments, in contrast to only 15% of all species listed in WoRMS (Fig. 4b). The spatial distribution of ABNJ-specific species (Supplementary Table 4) is predominantly in the sub-Antarctic and Antarctic latitudes (Supplementary Fig. 8).

ABNJ account for 64% of the ocean surface area and 95% of its volume. Once thought to be largely devoid of life, the deep-sea habitats and the water column have been found to harbour many marine species. While many of these species are thought to be considerably cosmopolitan, hotspots of endemism are found throughout the deep sea, perhaps most strikingly around hydrothermal vent systems³². According to geolocations of active hydrothermal vents (721 in total), more than half (363) are located in ABNJ.

Discussion

Marine biotechnology is mainly focused on species that serve as model organisms in basic research and as a backbone in genetic engineering, allowing the creation of new drugs and increasing the efficiency of biotechnological processes for food and energy production, plant agriculture or the invention of new materials³³. Marine species currently represent a small, but important, share that is used as a source for natural product discovery^{7,30}. Unravelling the global scope of economic

interest in MGR is a crucial first step towards understanding the value that rests in the biological functions encoded in genetic sequences and pathways to fair and equitable sharing of benefits from its use.

Patent data are a valuable source of information in examining innovation and technological advancements, which are widely acknowledged as key drivers of firm performance and economic growth^{34,35}. Aggregate patent application counts in particular are useful for studying national patenting activity³⁶. Patent data also provide insights into the scope of ‘pre-emptive patenting’ to block competitors, to increase the market price of existing products or to ensure operational freedom³⁷—strategies that biotechnology corporations are known to use³⁸. While estimating the market value of patents or establishing links to commercialization is challenging³⁹, patent data are a useful indicator for gaining insights into the long-term economic interest of societal actors in MGR applications on a global level, in the form of either knowledge production or market control.

The MABPAT database offers a global catalogue of patent sequences derived from marine species over the past three decades. It includes in depth information on patent applications, the genetic sequences attached to them and the marine species from which the sequences were derived, effectively connecting the resources and users of marine bioprospecting. In doing so, the MABPAT database not only

BOX 1

Marine gene patents needing extra protection

Global actors often seek patent protection for their inventions in multiple countries. Filing fees for patent applications extend into the thousands of US dollars per application, and the inventor still has to pay additional fees for each filing. For instance, the average cost of filing in the USA, including attorney's fees, has been estimated at around US\$50,000 (ref. 78). Therefore, it is more likely that protection will be sought for highly promising products, methods or associated biotechnological processes. A review of patent filings that include identical nucleotide sequences submitted to all national patent bureaus (13 sequences in total) reveals that the scope of patented commercial biotechnological applications is wide and usually involves transferring specific enzymes to cell metabolic pathways to maximize the production of a specific compound. Examples include applications of biotechnology in medicine (enzymes used for skin care), the food industry (enzymes used in baking and dairy products), agriculture (production of herbicide-tolerant transgenic crops), industrial production (metal nanomaterials used in products such as creams, shampoos, clothing, footwear and plastic containers) and the production of biofuel (isobutanol production and hydrocarbon biosynthesis) (Supplementary Table 2).

One sequence originating from the methanogenic marine archaea species *Methanococcus maripaludis* has been the subject of a series of lawsuits between Butamax (now a subsidiary of International Flavors & Fragrances (IFF)) and Gevo, a transnational biofuels company. The enzyme found in *M. maripaludis* was essential for isobutanol production, and these companies fought over the production of isobutanol instead of ethanol using yeast. After many years of patent disputes, the issue was finally settled through the splitting of the parties' licences in all fields of isobutanol production⁷⁹. In 2021, all Butamax-owned patents were completely acquired by Gevo⁸⁰.

fills an important research gap but also contributes to the transparency and interoperability of MGR use. By making it publicly available, we hope to enable further research efforts to inform improved policymaking. The analysis that generated this database also resulted in three key insights that are addressed below.

Rapid technological advances and data governance

Scholars have suggested that the earliest form of a patent system can be traced back 2,500 years ago to ancient Greece and that the first modern patent law dates back to the year 1474⁴⁰. Little surprise then that the patent system has struggled to keep pace with the rapid advances in genetics and genomics research of the past decades, as seen, for instance, in the considerable variation in ground rules for patenting genetic sequences across jurisdictions²⁷. Key developments over the past 30 years have focused on jurisdictional norms and compliance standards. In 1998, international applications introduced a mandatory data element for sequence description ('organism'), which aimed to indicate biological origin⁴¹. Yet, current international standards⁴² still allow the inclusion of custom organism names not listed in the Integrated Taxonomic Information System (<https://www.itis.gov/>), including 'unknown', 'unidentified' and 'artificial sequence'. The new requirements of INSDC⁴³, announced in November 2021, aim to ensure

correct origin disclosure for all incoming sequences. But the effect on the 24.5 million patent sequences already stored in the databases as well as new depositions remains uncertain given that it is ultimately up to patent offices to define standards for the sequences attached to patent applications (https://www.ncbi.nlm.nih.gov/education/patent_and_ip_faqs/).

The analysis of patents therefore often depends on either accepting considerable data gaps or developing methods to reconstruct missing data. In this study, for instance, 17.2 million sequences would have been excluded from the analysis owing to the lack of species names (primarily from the US Patent and Trademark Office, the largest repository of biological sequences and patents). Instead, our sequence similarity model allowed us to reasonably and more comprehensively estimate the patent shares across national states and actor types. This reconstruction allowed us to identify marine origin, focusing on molecular similarities of biological molecules instead of relying on disclosed species names, and to confirm with higher confidence than previous work that Japan, the USA and Germany are the headquarters location for the world's primary MGR patent applicants^{25,27}. The disproportionate importance of these three states suggests a corresponding responsibility to work towards innovative benefit-sharing and capacity-building mechanisms. These could include, for instance, the establishment of a multilateral fund for the equitable sharing of benefits between providers and users of digital sequence information (DSI), which has been agreed to be finalized at CBD COP16 (ref. 44).

Importance of microorganisms and deep-sea life for bioprospecting

Marine viruses, although having been recognized as being highly prevalent in ocean ecosystems, contributing to the largest pool of genetic diversity⁴⁵, have seen little commercial activity to date beyond a limited focus on those that affect commercial aquaculture production. However, the potential role of viruses in creating proteins of interest for marine bioprospecting could be bigger than we think. Viruses have shaped the majority of the genomes of Archaea and Bacteria via horizontal gene transfer, the exchange of genetic material between organisms that do not form parent–offspring relationships⁴⁶. Bacterial and archaeal species often live in symbiosis and exchange genes with microbial eukaryotes, protists⁴⁷, and together constitute the vast majority of organisms used in marine bioprospecting. Importantly, many archaeal and bacterial species used in bioprospecting live in deep-sea habitats, most of which are located in ABNJ. The diversity of microbial marine species is still highly underrepresented in databases that document the distribution and abundance of marine life (Box 2). This underrepresentation may account for the lack of patenting interest in species found exclusively in ABNJ. However, even with limited data, our findings show that ABNJ-specific species are 2.5 times more likely to inhabit the deep ocean compared with marine species in general.

Our analysis of the past three decades of global gene patents indicates that deep-sea species have become an important source for marine bioprospecting. All of the ten largest actors in marine bioprospecting are already using deep-sea species. As a result, there is a logic for benefit sharing from MGR utilization to flow into conservation projects aimed at protecting at-risk deep-sea habitats⁴⁸, not least as a vital source for future biotechnology focused on innovation and development of naturally derived products. More advanced biodiversity models that put emphasis on safeguarding entire communities with unique functional roles, including microbial species, should also be better integrated into conservation plans⁴⁹.

With the successful conclusion of the High Seas Treaty and the recognition of DSI in the legally binding agreement, MGR used for bioprospecting and product discovery opens a new opportunity to protect biodiversity in deep-sea habitats. However, the INSDC database, the largest data repository of DSI, is currently missing from the biodiversity informatics landscape⁵⁰; therefore, genetic diversity and information

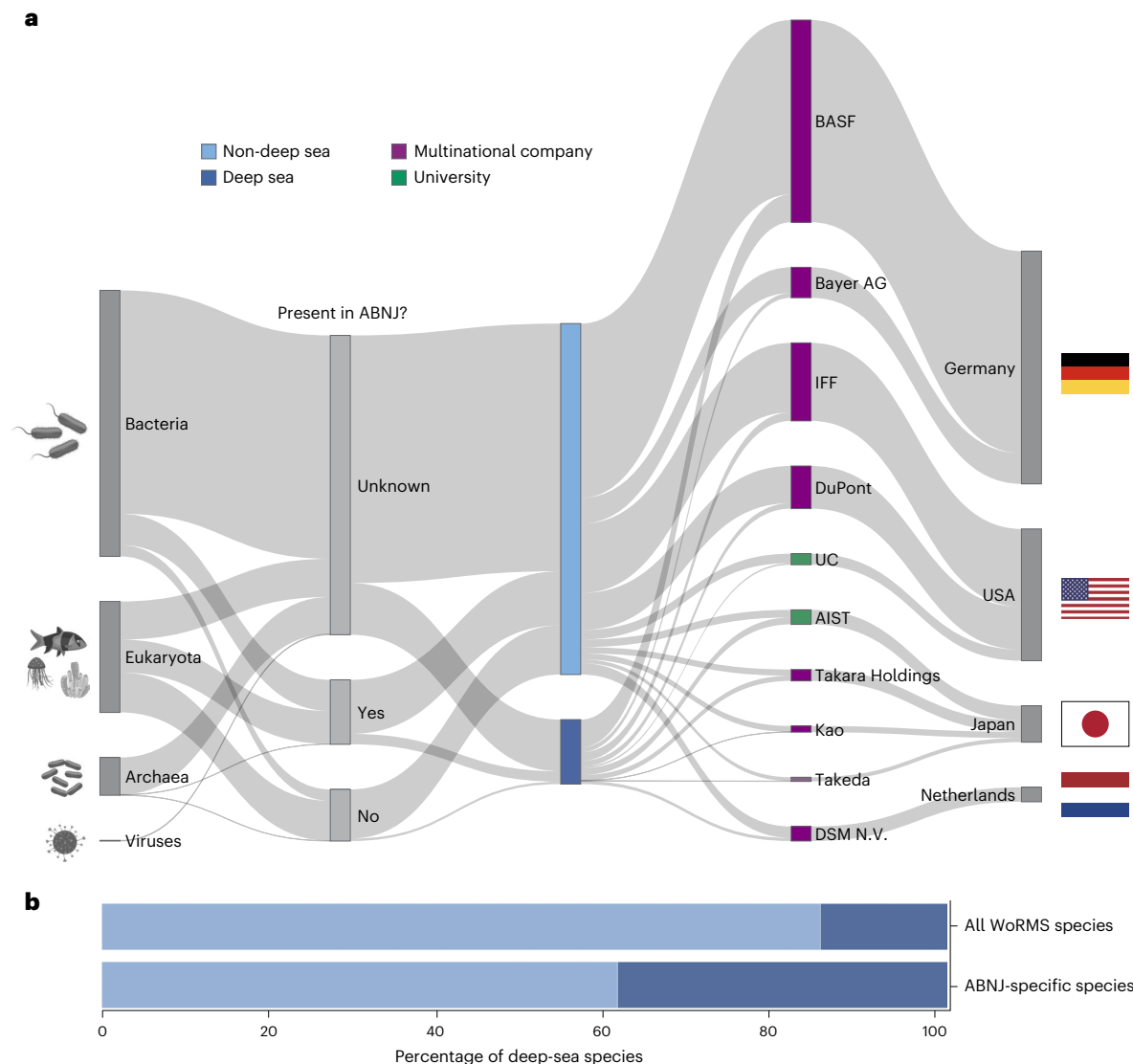


Fig. 4 | Species of interest to marine bioprospecting. a, Species of interest in bioprospecting connected to a company of reference (top 10 largest patent applicants) in patent application grouped by domain and potential presence in ABNJ and deep-sea habitats. AIST, The National Institute of Advanced Industrial

Science and Technology; UC, The University of California. The interactive version of this plot is available at <https://mabpat.shinyapps.io/main/>. **b,** Percentage of deep-sea species present exclusively in ABNJ and all species living in the ocean according to WoRMS. Credit: flags in **a**, flagpedia.net; icons in **a**, [FreePik.com](https://www.freepik.com/).

on the spatial origin of genetic information are not available on a global level. Adoption of the principles of Open and Responsible Data Governance and the development of MGR data repositories⁵¹ will be a necessary step to overcome the lack of information on MGR in ABNJ.

Intellectual property questions are not discussed within the High Seas Treaty, yet commercial sensitivities and national patent regulations are important for benefit sharing related to MGR sourced from the deep sea in ABNJ. While the agreed text of the treaty includes a voluntary mechanism to ensure traceability of MGR collected from ABNJ to end product, the treaty implementation will not affect sequences already used in marine bioprospecting up to date. As there are no legal requirements for patent holders to disclose commercialization of their patents, the scale of commercial products developed and marketed from deep-sea organisms will remain poorly understood. A continued increase in corporate interest along current trajectories would lead to unequal opportunities for new developments in biotechnology.

Multi-stakeholder collaboration in MGR protection

Analysis of bioprospecting patents yielded an asymmetrical distribution of patent registrations, consistent with previous findings^{25,27}.

The sector is dominated by transnational corporations, which have a higher capacity to undertake genomic research. One-third of all patents were held by the ten largest actors, eight of which are large multinational corporations and none of which conduct marine research themselves but instead rely on public gene databases for sequences with potential commercial applications. While many multinational pharmaceutical companies have marine biology departments⁵², their total share of bioprospecting patents is modest (Supplementary Table 3). Still, a fair estimate of corporate engagement in marine species discovery is hard to calculate. Marine scientists who study microbial diversity often engage in collaboration with the oil and gas industry for the collection of samples in deep-sea oil wells^{53,54}. With the rising popularity of using remotely operating vehicles for the inspection and maintenance of offshore oil and gas development sites, it is likely that more science–industry partnerships will emerge to support collection of biological data in the deep sea⁵⁵.

The disproportionate role of a small number of actors also suggests the potential for science–industry collaboration in the spirit of previous efforts with so-called keystone actors, which consists in engaging the largest companies in a given sector to enable transformative

BOX 2

A vast sea of unknown microbial diversity in the ocean

Microbial species from all domains of life account for more than 95% of total marine biomass and play a pivotal role in the functioning of marine ecosystems, as a foundational level of food webs, climate regulation elements and the backbone of immense genetic diversity⁸¹. However, as of November 2022, the OBIS database contains only 499 bacterial (<https://obis.org/taxon/6>) and 7 archaeal species (<https://obis.org/taxon/8>). The total diversity of microbial marine species, including those uniquely present in ABNJ, is highly underrepresented. Similarly, the WoRDS list contains almost no bacteria (18 species) or archaea (3 species) as both databases use WoRMS taxonomy as a backbone, and INSDC and National Center for Biotechnology Information (NCBI) taxonomy is not supported as one of the data sources in OBIS.

Thanks to the continuous effort of TARA Oceans, Malaspina, BioGEOTrACES and other projects contributing towards a deeper understanding of microbial diversity, more data have become available to project world map microbial distribution in the global ocean⁸². Some regional diversity studies, including a study in the Clarion–Clipperton Fracture Zone, have already shown that the molecular diversity of deep-sea species is equivalent to levels found in coral reef ecosystems⁸³. Ocean datasets are currently missing a global map of microbial functions⁸⁴. As most marine species listed in patent claims are of microbial origin, our understanding of how many of them are uniquely present in ABNJ is far from complete^{20,21,31,42–44,57,58,70–76,78–80}.

change⁵⁶. Constructive efforts to promote sustainable management in ABNJ have also been undertaken by partnerships such as the Deep Seas Project (<https://www.deep-seas.eu>) and the Common Oceans ABNJ Project⁵⁷, as well as regional bodies such as OSPAR Commission, the North East Atlantic Fisheries Commission and the Sargasso Sea Commission⁵⁸, which have addressed challenges related to illegal, unreported and unregulated fishing, and pollution, based on integrated and holistic approaches. The International Seabed Authority, empowered by UNCLOS (Supplementary Text 1) to manage the resources of the seabed in ABNJ, has begun to apply tools such as Regional Environmental Management Plans (REMPs) and designated associated Areas of Particular Ecological Interest (APEIs) aimed at conserving ecosystem function and biodiversity. The impact of such measures could be further amplified by seeking a coordinated approach in accordance with overarching environmental goals⁵⁹. Such initiatives can foster cross-sectoral dialogue and capacity-building activities that improve the capacity of national governments and local communities to engage in sustainable resource use in ABNJ.

Corporate efforts to safeguard intellectual property rights, significant data gaps and the heterogeneity of data standards have contributed to the use of ambiguous terminology and a lack of precision in discussions concerning MGR and bioprospecting in ABNJ. This has shaped perceptions of the scale and nature of commercial interest in MGR from ABNJ, feeding expectations of a lucrative ‘deep-sea gold rush’ without adequate empirical support for such claims^{60,61}. While the conclusion of the High Seas Treaty has laid the foundation for improved management in ABNJ, its entry into force and full implementation are a remote prospect and, in the meantime, voluntary collaborative efforts based on the best available science can help inform future binding

mechanisms to ensure conservation and sustainable use. By filling the crucial knowledge gap in understanding the potential of MGR, the MABPAT database represents a first step in that direction.

Methods

Summary statistics of patents that include MGR

The GenBank patent division, the European Bioinformatics Institute database (EMBL-EBI) and the DNA DataBank of Japan (DDBJ) exchange their data daily and together form the INSDC. Genetic sequences associated with patents were retrieved from the Patent division of GenBank from the NCBI (GenBank database) on 10 November 2022; this included 24,600,503 annotated sequences. All files (from `gbpat1.seq.gz` to `gbpat254.seq.gz`) were downloaded and processed following the methodology of ref. 25 to create database entries with information on the nucleotide sequence of DNA, species name, patent number, patent data and the party registering the patent. This was done by splitting each file into individual sequences and by extracting the data in the ‘origin’ field (nucleotide sequence), ‘organism’ field (species name) and ‘journal’ field (patent application number, year of application, patent system and patent applicant name) for each sequence. Unlike previous studies^{25,27} that restricted their analysis to sequences submitted in a given patent system, here we considered both patents submitted in national jurisdictions and those filed under the Patent Cooperation Treaty (‘international’ patents) of WIPO.

As of November 2022, sequences from a total of 14,708 different species were included in the GenBank database. To determine the subset of ‘marine species’ within the database, the taxon match tool of the WoRMS was used for all database entries, resulting in a filtered list of 4,000 species. Web searches were conducted for each of these species to verify the marine origin and to collect further information about the nature of each species. More than half of the matched species were subsequently excluded as non-unique to marine environments, resulting in the list of 1,474 marine species, which was used to select patent records associated with disclosed marine species. See ref. 27 for details of marine origin determination and criteria for filtering.

The taxonomy (domain and phylum) of 879 marine species was retrieved from the WoRMS database. In cases in which such taxonomic levels were not available, we obtained species taxonomy from the NCBI taxonomy database (<https://www.ncbi.nlm.nih.gov/taxonomy>) and Wikipedia (<https://en.wikipedia.org/wiki/>) (220 and 356 species, respectively). We did not succeed in matching 19 of the marine species (predominantly marine bacterial strains) into related taxonomic groups owing to lack of certainty in organism names. The complete list of marine species selected for this study is given in Supplementary Table 5a.

MABPAT construction

Marine biotechnology pipelines usually focus on the search for biological compounds that encode a new functionality⁶². There are two types of nucleotide sequences encoded in DNA: protein-coding sequences and non-coding sequences. The latter could have either a functional or a non-functional role in genome regulation, including DNA fragments that code for proteins involved in all cell functions. Except for short peptides like cone snail peptide toxins⁶³, most natural products are derived from proteins, which are polypeptide chains of a certain length. While identifying the shortest polypeptide chain length to form a protein is still controversial, it is currently estimated at 50 (ref. 64) to 100 (ref. 65) amino acids or 150 to 300 DNA base pairs, respectively.

Another important metric widely used to analyse genome composition variation in molecular biology and genomics is nucleotide usage, which is normally calculated as GC content—the percentage of certain nucleotide bases (guanine and cytosine) that form stronger chemical bonds in DNA strings. Modern genetic engineering techniques such as CRISPR⁶⁶ have proven to be very useful at enhancing important functions of proteins by altering DNA makeup. This could involve

changing individual nucleotides or introducing short sequences that control gene regulation and protein synthesis. Hence, GC content for modified proteins with similar functionality remains the same. Short DNA sequences, below the shortest DNA length required for protein formation, have various functions, including in the amplification of a specific gene sequence (as PCR primers), and usually have a wide range of GC content.

To predict whether genetic sequences are protein coding or not, we applied two filtering criteria: sequence length threshold and the presence of an open reading frame (ORF)—a gene region that has the potential to be transcribed into RNA and, after, translated into proteins. Sequences with an ORF longer than 150 base pairs have been considered protein-coding sequences. As most natural products are derived from proteins, we reason that at least one protein-coding sequence has to be included in a patent application, to be related to marine bioprospecting. Following that, we selected 31,914 protein-coding sequences associated with 1,039 marine species together with 112,115 of other sequences that have been submitted as a part of the same application.

For all companies that have registered patents associated with MGR, we counted the total number of nucleotide sequences and calculated the average sequence length (Fig. 2). Based on the shortest protein length estimation, the number of protein-coding or non-coding sequences for each company was identified. In each category, for the ten companies with the highest counts of genetic sequences attached to patent claims, we calculated the length and DNA composition (GC content) of each sequence, and coloured by distinct species origin (Supplementary Fig. 1).

For each sequence that was included in patent applications submitted in national jurisdictions as well as ‘international’ patents (sequences of special patenting interest), we collected the description of the invention and the protein function, if a nucleotide sequence search (BlastX) resulted in a significant match to a protein with annotated function. Web searches were conducted for each of these proteins to collect further information about protein function and potential application. The resulting information about the sequences of special patenting interest is available in Supplementary Table 2.

Patents owned by subsidiaries were replaced with ultimate owner names of controlled subsidiaries as stated in the Orbis company database, which contains information on around 400 million companies worldwide (Orbis; <https://orbis.bvdinfo.com/>). For jointly owned patents, the ownership was assigned to the first company on the list. After filtering and removing duplicate entity names and aggregating subsidiaries, we identified a total of 1,125 applicants and collected information about each through web searches, including the country where it is headquartered and the type of entity that it represents. Our classification resulted in five major entity types: multinational (presence in more than two countries) or national companies, universities and their commercialization centres, governmental agencies and ‘other’ (predominantly applications submitted by private individuals). We also included patent applications from 201 entities that contained protein-coding sequences with identified marine origins, which we were unable to classify under any specific entity type (‘none’).

Each record in the MABPAT database includes the following: (1) patent applicant name, (2) type of applicant, (3) country where it is headquartered, (4) year of application, (5) patent application number, (6) patent system, (7) genetic sequence identification, (8) marine species name associated with the sequence, (9 and 10) species taxonomy, (11) taxonomic source, (12) whether species can be classified as ‘deep-sea’ species, (13) source of deep-sea presence, (14) whether species were observed in ABNJ, (15) genetic sequence, (16) GC content, (17) sequence length, (18) whether the sequence originated from a marine organism, (19) whether the marine origin of the sequence was disclosed by the patent applicant or bioinformatically predicted, (20) whether the sequence contains protein-coding information and (21) sequence prediction source. If the marine origin was predicted,

the following information about the most similar protein entry in the reference database is provided: (22) protein entry header, (23) protein entry sequence identification, (24) protein entry title, (25) *E*-value, (26) hit identity and (27) query coverage.

Deep-sea presence of marine species

The search for presence of species in deep-sea habitats was conducted based on multiple sources. For species in the Eukarya domain of life, we used the WoRDSS, a taxonomic database of deep-sea species. As Bacteria and Archaea species are not present in WoRDSS, we used web search based on the PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) and Integrated Microbial Genomes and Microbiomes (<https://img.jgi.doe.gov/>) databases to establish their potential presence in deep-sea habitats, whether within or beyond national jurisdiction. Samples of species collected from deep-sea environments that have already been found to be associated with international patent applications²⁷ are also marked as ‘deep-sea’ species. For the definition of deep-sea marine species, we followed the inclusion criteria in WoRDSS, that is, that the biological material was sampled in depths greater than 500 m.

BlastX sequence similarity model and patent share estimation

Sequence similarity models are widely used to identify newly sequenced data or unknown species⁶⁷. To conduct sequence similarity BlastX searches (translated nucleotide versus protein) against the database of annotated protein sequences, we created the reference database of all proteins belonging to 627 genera of previously confirmed marine species in Supplementary Table 5a. A total of 24,024,531 proteins from all species within those genera were selected from UniProt Knowledgebase (UniProtKB/Swiss-Prot; UniProt Consortium 2023) which included Swiss-Prot (the expertly curated protein records) and TrEMBL (bioinformatically predicted proteins).

BlastX searches with a specific set of search parameters (*E*-value $\leq 10^{-5}$, query coverage $\geq 80\%$, hit identity $\geq 99\%$) were used to verify that marine sequences could be identified to a genus level with at least 95% confidence (correct hit) (Supplementary Fig. 3a). We also tested whether correct hits and searches with confidence below 95% tend to be included in certain patent applications, patented by certain actors or in certain patent systems, but did not find any preference (Supplementary Fig. 3b,c). Using the sequence search tool DIAMOND⁶⁸, we queried 12,716 protein-coding sequences with disclosed marine origin against the selected records from UniProtKB, which resulted in 10,514 correct hits (82.68% recovery rate).

We then queried 7,467,396 sequences with unknown taxonomic origin (‘unknown’, ‘unidentified’ and ‘synthetic construct’ species tag)—62.7% of all GenBank records—against the selected records from UniProtKB, and found 234,836 sequences originating from 1,368 species not previously disclosed in patent records. All matched species were subsequently verified to be exclusively present in marine habitats, resulting in a final list of 561 additional marine species (Supplementary Table 5b). Overall, we have recovered 60,636 previously unknown protein-coding sequences with marine origin and 144,545 other sequences that have been submitted as a part of the same patent application (2,257 patent applications in total).

Finally, we compared summary statistics (number of sequences, number of patents and median year of application) for the top 10 largest patent applicants that referenced sequences with disclosed marine origin and top 10 applicants that referenced sequences with predicted marine origin (Supplementary Figs. 4 and 5, respectively), and found that both lists contained the two largest patent applicants (Bayer and BASF, respectively).

Hydrothermal vent presence and ABNJ-unique species counts

The geolocation of hydrothermal vents was collected from the Inter-Ridge Vents Database. The maritime boundary map of World High Seas was downloaded from Marine Regions (<https://marineregions.org/>).

Each set of hydrothermal vent coordinates was checked for presence within any of the High Seas polygons. Spatial vector data were analysed with the R package *sf* version 1.0-9 (ref. 69).

To establish the list of species uniquely present in ABNJ, we used species geographical abundance data from OBIS. We first retrieved all 28,375 species with at least one occurrence record in ABNJ (<https://obis.org/area/1>). For each ABNJ-present species, we checked if it was also observed in the territorial waters of any country. Species with at least one occurrence record were excluded. Data were obtained from the OBIS database (2022) using the R package *robis* version 2.11.0 (ref. 70) and *parallel* version 3.6.2. (ref. 71).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data were collected from publicly available data sources: INSDC (<ftp.ncbi.nih.gov/genbank/>)⁷², UniProtKB⁷³, WoRMS (<https://www.marinespecies.org>)⁷⁴, WoRDS⁷⁵, PubMed (<https://pubmed.ncbi.nlm.nih.gov>) and the Integrated Microbial Genomes and Microbiomes database (<https://img.jgi.doe.gov>). Species observation records were obtained from OBIS (<https://obis.org>). The geolocations of hydrothermal vents were collected from the InterRidge Vents database (<http://vents-data.interridge.org>)⁷⁶. The maritime boundaries map of World High Seas was downloaded from Marine Regions (World EEZ v.11) (<https://marineregions.org>). The resulting MABPAT database is available at <https://mabpat.shinyapps.io/main> and via figshare at <https://doi.org/10.6084/m9.figshare.25289404.v3> (ref. 77).

Code availability

Analysis scripts are available via GitHub at <https://github.com/zhiwokopias/mabpat>.

References

- Falkinham, J. O. et al. Proliferation of antibiotic-producing bacteria and concomitant antibiotic production as the basis for the antibiotic activity of Jordan's red soils. *Appl. Environ. Microbiol.* **75**, 2735–2741 (2009).
- Cragg, G. M. & Pezzuto, M. Natural products as a vital source for the discovery of cancer chemotherapeutic and chemopreventive agents. *Med Princ. Pract.* **25**, 41–59 (2016).
- Hournan, P. C. H., Hertog, M. G. L. & Katanc, M. B. Analysis and health effects of flavonoids. *Food Chem.* **57**, 43–46 (1996).
- Achan, J. et al. Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. *Malar. J.* **10**, 144 (2011).
- Shah, Z. et al. Podophyllotoxin: history, recent advances and future prospects. *Biomolecules* **11**, 603 (2021).
- Atanasov, A. G. et al. Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).
- Sigwart, J. D. et al. Unlocking the potential of marine biodiversity. *Nat. Prod. Rep.* **38**, 1235–1242 (2021).
- Beraldi-Campesi, H. Early life on land and the first terrestrial ecosystems. *Ecol. Process.* **2**, 1 (2013).
- Román-Palacios, C., Moraga-López, D. & Wiens, J. J. The origins of global biodiversity on land, sea and freshwater. *Ecol. Lett.* **25**, 1376–1386 (2022).
- Appeltans, W. et al. The magnitude of global marine species diversity. *Curr. Biol.* **22**, 2189–2202 (2012).
- Eguiluz, V. M. et al. Scaling of species distribution explains the vast potential marine prokaryote diversity. *Sci. Rep.* **9**, 18710 (2019).
- Gerwick, W. H. & Moore, B. S. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.* **19**, 85–98 (2012).
- Blasiak, R. et al. A forgotten element of the blue economy: marine biomimetics and inspiration from the deep sea. *PNAS Nexus* **1**, pgac196 (2022).
- Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A. & Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **38**, 362–413 (2021).
- Blasiak, R. et al. Making marine biotechnology work for people and nature. *Nat. Ecol. Evol.* **7**, 482–485 (2023).
- Li, B. et al. Preparation of lactose-free pasteurized milk with a recombinant thermostable β -glucosidase from *Pyrococcus furiosus*. *BMC Biotechnol.* **13**, 73 (2013).
- Aesoy, R. & Herfindal, L. in *Principles of Cancer Treatment and Anticancer Drug Development* (ed. Link, W.) 137–139 (Springer International Publishing, 2022).
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
- Wynberg, R. & Laird, S. A. Fast science and sluggish policy: the Herculean task of regulating biodiscovery. *Trends Biotechnol.* **36**, 1–3 (2018).
- Convention on Biological Diversity (Secretariat of the CBD, UN Environment Programme, 2011); <https://www.cbd.int/convention/text>
- 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 (Secretariat of the CBD, UN Environment Programme, 2011); <https://www.cbd.int/abs/text>
- UN General Assembly. *Draft Agreement under the United Nations Convention on the Law of the Sea on the Conservation and Sustainable Use of Marine Biological Diversity of Areas Beyond National Jurisdiction*; https://www.un.org/bbnj/sites/www.un.org.bbnj/files/draft_agreement_advanced_unedited_for_posting_v1.pdf (2023).
- Oldham, P., Hall, S. & Barnes, C. *Patent Landscape Report on Animal Genetic Resources* (World Intellectual Property Organization, 2014); https://www.wipo.int/edocs/pubdocs/en/wipo_pub_947_3.pdf
- Scholz, A. H. et al. Myth-busting the provider-user relationship for digital sequence information. *Gigascience* **10**, giab085 (2021).
- Arnaud-Haond, S., Arrieta, J. M. & Duarte, C. M. Marine biodiversity and gene patents. *Science* **331**, 1521–1522 (2011).
- Arrieta, J. M., Arnaud-Haond, S. & Duarte, C. M. What lies underneath: conserving the oceans' genetic resources. *Proc. Natl Acad. Sci. USA* **107**, 18318–18324 (2010).
- 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).
- 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).
- Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).
- Jaspars, M. et al. The marine biodiversity pipeline and ocean medicines of tomorrow. *J. Mar. Biol. Assoc.* **96**, 151–158 (2016).
- The IUCN Red List of Threatened Species* Version 2024-1 (IUCN, 2022).
- Van Dover, C. L. et al. Scientific rationale and international obligations for protection of active hydrothermal vent ecosystems from deep-sea mining. *Mar. Policy* **90**, 20–28 (2018).
- Khan, I., Akmal, K. F., Chong, W. S., Maran, B. A. V. & Shah, M. D. in *Marine Biotechnology: Applications in Food, Drugs and Energy* (eds Shah, M. D. et al.) Ch. 1 (Springer Nature, 2023).
- Hasan, I. et al. The innovation–economic growth nexus: global evidence. *Res. Policy* **39**, 1264–1276 (2010).
- Lara-Lopez, A., Valdés, L., de Pinho, R. & Enevoldsen, H. In *Global Ocean Science Report 2020: Charting Capacity for Ocean Sustainability* (ed. Isensee, K.) 135–173 (UNESCO Publishing, 2020).

36. Haščič, I. et al. *Public Interventions and Private Climate Finance Flows: Empirical Evidence from Renewable Energy Financing*; OECD Environment Working Papers no. 80 (2015).
37. Guellec, D., Martinez, C. & Zuniga, M. P. Pre-emptive patenting: securing market exclusion and freedom of operation. *Econ. Innov. New Technol.* **21**, 1–29 (2012).
38. Gurgula, O. Strategic patenting by pharmaceutical companies—should competition law intervene? *IIC Int. Rev. Ind. Prop. Copyr. Law* **51**, 1062–1085 (2020).
39. Hall, B. H., Jaffe, A. & Trajtenberg, M. Market value and patent citations. *RAND J. Econ.* **36**, 16–38 (2005).
40. Adams, J. N. in *Research Handbook on Patent Law and Theory* (ed. Takenaka, T.) Ch. 1, 2–26 (Edward Elgar Publishing, 2019).
41. Jefferson, O. A., Köllhofer, D., Ajikuttira, P. & Jefferson, R. A. Public disclosure of biological sequences in global patent practice. *World Pat. Inf.* **43**, 12–24 (2015).
42. *Standard ST.26: Recommended Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings Using XML (Extensible Markup Language)* (WIPO, 2023); <https://www.wipo.int/export/sites/www/standards/en/pdf/03-26-01.pdf>
43. *Spatio-Temporal Annotation Policy* (INSDC, 2021); <https://www.insdc.org/news/spatio-temporal-annotation-policy-18-11-2021/>
44. *COP15: Nations Adopt Four Goals, 23 Targets for 2030 in Landmark UN Biodiversity Agreement* (Secretariat of the CBD, UN Environment Programme, 2022); <https://www.cbd.int/article/cop15-cbd-press-release-final-19dec2022>
45. Suttle, C. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
46. Sobecky, P. A. & Hazen, T. H. Horizontal gene transfer and mobile genetic elements in marine systems. *Methods Mol. Biol.* **532**, 435–53, (2009).
47. Husnik, F. et al. Bacterial and archaeal symbioses with protists. *Curr. Biol.* **31**, 862–877 (2021).
48. Cordes, E. E. & Levin, L. A. Exploration before exploitation. *Science* **359**, 719 (2018).
49. Pollock, L. J. et al. Protecting biodiversity (in all its complexity): new models and methods. *Trends Ecol. Evol.* **35**, 1119–1128 (2020).
50. Bingham, H. et al. The biodiversity informatics landscape: elements, connections and opportunities. *Res. Ideas Outcomes* **3**, e14059 (2017).
51. Oldham, P., Chiarolla, C. & Thambisetty, S. *Digital Sequence Information in the UN High Seas Treaty: Insights from the Global Biodiversity Framework-related Decisions*; LSE Law School Policy Briefing Series 53/2023. Available at SSRN <https://doi.org/10.2139/ssrn.4343130> (2023).
52. Trevisanut, S. & Bonfanti, A. *Intellectual Property Rights Beyond National Jurisdiction: Outlining a Regime for Patenting Products Based on Marine Genetic Resources of the Deep-Sea Bed and High Sea*. Available at SSRN <https://doi.org/10.2139/ssrn.1861020> (2011).
53. Alexander, J. B. et al. Complementary molecular and visual sampling of fish on oil and gas platforms provides superior biodiversity characterisation. *Mar. Environ. Res.* **179**, 105692 (2022).
54. Franco, N. R. et al. Bacterial composition and diversity in deep-sea sediments from the southern Colombian Caribbean Sea. *Diversity* **13**, 10 (2020).
55. McLean, D. L. et al. Enhancing the scientific value of industry remotely operated vehicles (ROVs) in our oceans. *Front. Mar. Sci.* **7**, 00220 (2020).
56. Österblom, H. et al. Scientific mobilization of keystone actors for biosphere stewardship. *Sci. Rep.* **12**, 3802 (2022).
57. *The Common Oceans: ABNJ Deep Seas Project* (FAO, 2018); <https://www.fao.org/3/CA2245EN/ca2245en.pdf>
58. Wright, G. & Rochette, J. *Regional Ocean Governance of Areas Beyond National Jurisdiction: Lessons Learnt and Ways Forward* (STRONG High Seas Project, 2019); https://www.prog-ocean.org/wp-content/uploads/2019/03/STRONG-HS_Lessons-Learnt-Report.pdf
59. Amon, D. J. et al. Assessment of scientific gaps related to the effective environmental management of deep-seabed mining. *Mar. Policy* **138**, 105006 (2022).
60. Leary, D. & Juniper, S. K. in *The Limits of Maritime Jurisdiction* (eds Schofield, C. et al.) Ch. 34, 769–785 (Martinus Nijhoff Publishers, 2014).
61. Leary, D. Marine genetic resources in areas beyond national jurisdiction: do we need to regulate them in a new agreement? *Marit. Saf. Secur. Law J.* **5**, 22–47 (2018).
62. Rotter, A. et al. The essentials of marine biotechnology. *Front. Mar. Sci.* **8**, 629629 (2021).
63. Terlau, H. & Olivera, B. M. Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.* **84**, 41–68 (2004).
64. Woolfson, D. N., Baker, E. G. & Bartlett, G. J. How do miniproteins fold? *Science* **357**, 133–134 (2017).
65. Brunet, M. A., Leblanc, S. & Roucou, X. Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Exp. Cell Res.* **393**, 112057 (2020).
66. Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum. Mol. Genet.* **23**, 40–46 (2014).
67. Pearson, W. R. An introduction to sequence similarity (“homology”) searching. *Curr. Protoc. Bioinformatics* **Chapter 3**, 3.1.1–3.1.8 (2013).
68. Buchfink, B., Xie, C. & Huson, D. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
69. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439 (2018).
70. Provoost, P., Bosch, S. & Best, B. iobis/robin: 2.11.0. Zenodo <https://zenodo.org/doi/10.5281/zenodo.1489948> (2022).
71. parallel: support for Parallel computation in R. R version 3.6.2 <https://rdocumentation.org/packages/parallel/versions/3.6.2>
72. Arita, M. et al. The international nucleotide sequence database collaboration. *Nucleic Acids Res.* **49**, D121–D124 (2021).
73. The Uniprot Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
74. *World Register of Marine Species* (WoRMS Editorial Board, accessed 15 November 2022); <https://doi.org/10.14284/170>
75. Glover, A. G., Higgs, N. & Horton, T. *World Register of Deep-Sea Species* (WoRDSS) (accessed 15 November 2022); <https://doi.org/10.14284/352>
76. Beaulieu, S. E. & Szafranski, K. *InterRidge Global Database of Active Submarine Hydrothermal Vent Fields* Version 3.4 (InterRidge, accessed 1 February 2023).
77. Zhivkoplis, E. MARine Bioprospecting PATent dataset. *figshare* <https://doi.org/10.6084/m9.figshare.25289404.v3> (2024).
78. How much does a patent cost? *BlueIron* (16 January 2022); <https://blueironip.com/how-much-does-a-patent-cost>
79. Butamax, Gevo settle patent dispute. *Biomass Magazine* (24 August 2015); <https://biomassmagazine.com/articles/butamax-gevo-settle-patent-dispute-12339>
80. Gevo acquires Butamax patent estate. *Yahoo Finance* (23 September 2021); <https://finance.yahoo.com/news/gevo-acquires-butamax-patent-estate-130000249.html>
81. Abida, H. et al. Bioprospecting marine plankton. *Mar. Drugs* **11**, 4594–4611 (2013).
82. Delmont, T. O. et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genom.* **2**, 100123 (2022).

83. Haward, M. G. & Rogers, A. D. Marine genetic resources in areas beyond national jurisdiction: promoting marine scientific research and enabling equitable benefit sharing. *Front. Mar. Sci.* **8**, 667274 (2021).
84. Tara Ocean Foundation, T. O. & Oceans, T. Priorities for ocean microbiome research. *Nat. Microbiol.* **7**, 937–947 (2022).

Acknowledgements

We thank D. Khvostovets for providing valuable consultancy in the Patent Cooperation Treaty and European Patent Convention. E.Z., A.P. and R.B. are funded by FORMAS, project number 2020-01048. A.P. is also funded by FORMAS, project number 2019-01220. P.D. is funded by the Research Platform Governance of Digital Practices at the University of Vienna. J.-B.J. is funded by the Knut and Alice Wallenberg Foundation (2021.0343).

Author contributions

E.Z. and A.P. collected the raw data. E.Z., J.-B.J. and R.B. designed the research and analysed the data. E.Z., P.D., J.-B.J. and R.B. wrote the paper.

Funding

Open access funding provided by Stockholm University.

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/s41893-024-01392-w>.

Correspondence and requests for materials should be addressed to Erik Zhivkoplis.

Peer review information *Nature Sustainability* thanks Peter McGarvey and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at 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

9. Discussion

This section synthesizes the contributions made in this thesis. It first discusses empirical contributions and reflects on how the underlying research approach extends the literature on marine genetic resources and commons governance (9.1). It then reflects on policy implications, discussing pathways for multilateral agreements, national governments, and local institutions to counteract persisting equity issues in the ocean genome commons (9.2). Finally, it discusses limitations and future research (9.3).

9.1. Empirical and theoretical contributions

Paper 1 presents evidence of imbalanced global relationships in the provision and use of DSI. It finds that OECD countries are the biggest users of foreign-sourced DSI and that a high share of their foreign-sourced publications did not involve collaboration with local authors. While international collaboration has increased over time, certain types of genetic research, namely basic research, research in mega-biodiverse countries, and marine genetic research, continue to be significantly more exclusive towards authors from countries of origin. In observing these nuances, the paper establishes an understanding of global user-provider relationships as shaped by historical inequities and dependent on specific research contexts. These insights establish important nuances to the previous research by Scholz et al. (2021) on DSI user-provider relationships.

Paper 2 addresses the research gap on relationships between marine genetic research and patenting practices. While the paper reiterates evidence on global research and innovation capacity imbalances between developing and developed countries (Blasiak et al. 2018; P. Oldham et al. 2014; Tolochko and Vadrot 2021), its analysis of patent references to marine genetic research articles establishes a more nuanced view. It identified ‘net-providers’ and ‘net-users’ of marine genetic research, pointing to diverging provider-user balances even amongst developed countries: some countries benefit comparatively more from others’ provisioning of marine genetic research than they provide themselves. As such, the paper describes a typical freeriding problem in the science-patent nexus, a collective action problem identified in many other cases of commonly shared resources (Hess and Ostrom 2007).

Paper 3 extends previous studies on marine genetic resource patents (Arnaud-Haond, Arrieta, and Duarte 2011; Blasiak et al. 2018). It introduces the Marine Bioprospecting Patent (MABPAT) database, a comprehensive dataset of 92,550 protein-coding sequences linked to 4,779 patent filings from 1989 to 2022. Using a sequence similarity model to supplement incomplete patent records, the study offers the most extensive mapping to date of marine genetic resources in patents. It shows that microbial and deep-sea species, the latter being predominantly found in Areas Beyond National Jurisdiction, are central to marine bioprospecting. It identifies key corporate actors and geographic concentrations of patent activity in only a few countries, notably in the USA, Germany, and Japan. By quantifying the

extent and origins of marine genetic resource use in patents, the paper provides empirical evidence of corporate interest in marine genetic resources, including from the deep-sea.

Beyond the individual papers, the thesis provides a holistic understanding of how marine genetic resources circulate through interconnected stages of scientific research and bioprospecting pipelines. The research reveals significant imbalances in capacities to use and benefit from marine genetic resources across these stages: a small number of developed countries and transnational corporations dominate downstream use, while several providers of genetic materials and research, both in developing and developed countries, remain relatively inactive in patenting activities. As argued in this thesis, these imbalances indicate the existence of a freeriding problem in the ocean genome commons and raise important equity concerns. Such systematic and holistic insights could not be drawn from previous research on marine genetic resource collection and use (Blasiak et al. 2018; Jaspars, Humphries, and Rabone 2021), which has either not considered interactions between actors across genetic sample collection, marine genetic research, and subsequent corporate activities or not generalized these beyond individual cases.

The insights from the presented studies may prove beneficial to research on other knowledge or hybrid commons, as they potentially hint at generalizable practices or tendencies. The context of genetic resource governance appears particularly fruitful for holistic explorations of resource system dynamics, due to the high availability of metadata to study various provision and use activities. The thesis also provides an example of how the traditional commons literature may inform international relations and vice versa (Keohane and Ostrom 1995), due to its focus on nation states in their facilitating roles within the ocean genome commons.

9.2. Policy implications and pathways towards improved equity

The thesis points to a series of structural and institutional asymmetries in the global governance of marine genetic resources and DSI. These asymmetries have implications for ongoing international policy debates in the CBD and the BBNJ Treaty, as well as for more general considerations on data governance and research funding.

The findings of Paper 1 suggest that current national research policies, international capacity-building, and access and benefit-sharing arrangements need to pay attention to the significantly higher likelihood of non-involvement of local researchers in basic, marine, and ecologically sensitive genetic research contexts. In the case of marine genetic research, this divergence may be due to the fact that it is more costly to collect and study marine genetic materials compared to terrestrial resources (Sigwart et al. 2021), making it more difficult for universities and researchers in developing countries to participate. While the paper does not consider DSI sourced from Areas Beyond National Jurisdiction, the observed

lack of collaboration might equally shape research regulated under the BBNJ Treaty. Across legal contexts, policymakers and scientists may thus need to reassess international partnerships and the allocation of resources to establish more equitable and inclusive scientific practices involving marine genetic resources.

Paper 2 highlights imbalances between the provision and use of marine genetic research in the patent system, revealing countries that act primarily as users rather than providers of scientific knowledge. These imbalances are not addressed by current benefit-sharing mechanisms, which typically focus either on scientific activities, patenting, or commercialization, but not on relationships between these domains. This raises questions about how benefit-sharing mechanisms might evolve to reflect the full lifecycle of scientific knowledge, particularly when that knowledge becomes relevant for corporate activities (Brown 2025). The paper advances the idea of incorporating indicators related to science-patent balances into multilateral benefit-sharing frameworks. The BBNJ Treaty's article 14.7 explicitly considers future monetary benefit-sharing rates to be calculated 'based on a diversified set of indicators measuring the aggregate level of activities by a Party' (United Nations General Assembly 2023). Science-patent balances could be part of these sets of indicators, which would provide incentives for the continued provisioning of scientific research. Additional quality checks would be necessary to avoid counting 'non-valuable contributions', for instance, in the form of articles published in predatory journals. The notification information system constructed under the BBNJ Treaty (Lawson et al. 2025) could more generally benefit from the inclusion of science-patent citations linkages to establish transparency.

Paper 3's findings on the high concentration of intellectual property in the United States, Germany, and Japan raise questions about the adequacy of existing governance frameworks to address global imbalances in corporate use of marine genetic resources. Besides this, it addresses discussions about the difficulty of tracing DSI (Raposo et al. 2023) and introduces the methodological idea of focusing on 'most relevant' DSI, such as protein-coding sequences. This line of thinking may be useful in the design of transparency or traceability mechanisms under the BBNJ Agreement and the CBD, which could focus on specific DSI to increase the efficiency of investments and efforts into such mechanisms. The paper also provides ideas of linking innovation outcomes with conservation financing (Blasiak et al. 2023), for example, through earmarked contributions to marine protection.

All three papers point to the importance of improving current metadata reporting practices in administrative, scientific, and patent databases. Inaccurate or incomplete metadata, such as missing country of origin information, taxonomic identifiers, or biocultural labels, hinder efforts to monitor the movement and use of marine genetic resources and DSI and complicate compliance with international agreements (Rohden et al. 2020; Blasiak et al. 2019; Golan et al. 2022). The ongoing efforts by INSDC

to standardize metadata fields (International Nucleotide Sequence Database Collaboration 2023) could, in coordination with specific Treaty mechanisms such as the BBNJ standardized batch identifier (Lawson et al. 2025), significantly enhance traceability and facilitate effective implementation of access and benefit-sharing mechanisms.

9.3. Limitations and future research

The thesis relied on an analytical framework that considered countries of origin, science, and patenting as central stages of the marine bioprospecting pipeline. While this allowed for a holistic and systematic analysis, the approach was also constrained by several limitations, which I will turn to now.

A key limitation relates to relevant activities of marine genetic resource collection and use that are not included in the analytical framework. Importantly, the framework and analyses have not extended to commercial products. As argued in Paper 3, patent practices provide a vantage point for the study of corporate interest, but there is a wide gap from there to commercialization, as products need to traverse many additional stages along the biodiscovery pipeline, such as clinical trials (Jaspars, Humphries, and Rabone 2021). Additionally, patents often fulfill primarily strategic or speculative functions such as gatekeeping (Heller 2013; Guellec, Martinez, and Zuniga 2012) and financial hedging (Kang 2020) rather than indicating actual commercial activities. Methodological innovation and improved transparency standards will be needed to achieve traceability to commercial products. Besides commercialization, researchers may also want to extend the research approach to include specific information on samples (Oldham 2020), secondary scientific activities (Scholz et al. 2021), or data from research cruise reports.

Another limitation concerns the explicit focus on marine genetic resource provision and use at the nation state level. While this is desirable in terms of generalization and systematic understanding, it did not study relevant activities at smaller scales, as is typically the case in research on commons governance. It will thus be crucial to align the systematic findings with research on sub-national stakeholders and their interactions (Humphries, Rabone, and Jaspars 2021; Jaspars, Humphries, and Rabone 2021; Heinrich et al. 2020; Noreña – P et al. 2018; Geary and Bubela 2019; Nyika 2009; Rayne et al. 2022; Mc Cartney et al. 2023). Such case studies can provide valuable insights into causal mechanisms shaping global patterns and outcomes. This is a crucial step in the IAD framework's action arena, where concrete scenarios can be analyzed to understand how patterned interactions emerge, thus providing more detailed insights into how particular equity issues could be addressed (Hess and Ostrom 2007). The large-scale analysis and framework presented in this thesis may indeed strengthen such smaller-scale analyses, as it can provide criteria for case selection and a setup for cross-case comparison, for instance, by following established approaches for mixed methods research (Lieberman 2005).

10. Conclusion

As marine genetic resources become increasingly central to research and innovation in public health, environmental monitoring, and biotechnology, they also emerge as a new frontier of global environmental governance. This frontier entails a complex landscape of actors involved in their collection and use, constituting a global resource system that I conceptualized as the ocean genome commons. By empirically studying this system, this thesis provides a unique perspective on contemporary challenges and opportunities arising from the governance of such globally shared resources.

These challenges are manifold and, as this thesis shows, policy interventions must engage in difficult balancing acts between principles of open science and equity, environmental stewardship and corporate interest, and public control and bureaucratic efficiency. The ocean genome commons provides a unique space for developing, testing, and refining new models of traceability, benefit-sharing, and accountability to address these challenges. If these efforts are approached in a collaborative spirit and with foresight, the ocean genome commons has the potential to serve as a model for how humanity can equitably share Earth's vital resources.

11. References

- Adelman, David E. 2005. "A Fallacy of the Commons in Biotech Patent Policy." *Berkeley Technology Law Journal* 20 (2): 985–1030. <http://www.jstor.org/stable/24116639>.
- Ardura, Alba, Anastasija Zaiko, Jose L. Martinez, Aurelija Samulioviene, Anna Semenova, and Eva Garcia-Vazquez. 2015. "eDNA and Specific Primers for Early Detection of Invasive Species – A Case Study on the Bivalve *Rangia Cuneata*, Currently Spreading in Europe." *Marine Environmental Research*, Mares 2014, 112 (December):48–55. <https://doi.org/10.1016/j.marenvres.2015.09.013>.
- Arnaud-Haond, Sophie, Jesús M. Arrieta, and Carlos M. Duarte. 2011. "Marine Biodiversity and Gene Patents." *Science* 331 (6024): 1521–22. <https://doi.org/10.1126/science.1200783>.
- Arora, Ashish, Sharon Belenzon, and Andrea Pataconi. 2018. "The Decline of Science in Corporate R&D." *Strategic Management Journal* 39 (1): 3–32. <https://doi.org/10.1002/smj.2693>.
- Aubry, Sylvain, Christine Frison, Jorge C. Medaglia, Emile Frison, Marcel Jaspars, Muriel Rabone, Aysegul Sirakaya, Devanshi Saxena, and Esther van Zimmeren. 2022. "Bringing Access and Benefit Sharing into the Digital Age." *PLANTS, PEOPLE, PLANET* 4 (1): 5–12. <https://doi.org/10.1002/ppp3.10186>.
- Beygmoradi, Azadeh, and Ahmad Homaei. 2017. "Marine Microbes as a Valuable Resource for Brand New Industrial Biocatalysts." *Biocatalysis and Agricultural Biotechnology* 11 (July):131–52. <https://doi.org/10.1016/j.bcab.2017.06.013>.
- Bhatia, Pooja, and Archana Chugh. 2015. "Role of Marine Bioprospecting Contracts in Developing Access and Benefit Sharing Mechanism for Marine Traditional Knowledge Holders in the Pharmaceutical Industry." *Global Ecology and Conservation* 3 (January):176–87. <https://doi.org/10.1016/j.gecco.2014.11.015>.

- Biermann, Frank, and Rakhyun E. Kim, eds. 2020. *Architectures of Earth System Governance: Institutional Complexity and Structural Transformation*. 1st ed. Cambridge University Press. <https://doi.org/10.1017/9781108784641>.
- Blasiak, Robert, Jean-Baptiste Jouffray, Diva J. Amon, Joachim Claudet, Paul Dunshirn, Peter Søgaaard Jørgensen, Agnes Pranindita, Colette C. C. Wabnitz, Erik Zhivkoplis, and Henrik Österblom. 2023. “Making Marine Biotechnology Work for People and Nature.” *Nature Ecology & Evolution*, January, 1–4. <https://doi.org/10.1038/s41559-022-01976-9>.
- Blasiak, Robert, Jean-Baptiste Jouffray, Colette C. C. Wabnitz, Emma Sundström, and Henrik Österblom. 2018. “Corporate Control and Global Governance of Marine Genetic Resources.” *Science Advances* 4 (6): eaar5237. <https://doi.org/10.1126/sciadv.aar5237>.
- Blasiak, Robert, Jean-Baptiste Jouffray, Colette C.C. Wabnitz, and Henrik Österblom. 2019. “Scientists Should Disclose Origin in Marine Gene Patents.” *Trends in Ecology & Evolution* 34 (5): 392–95. <https://doi.org/10.1016/j.tree.2019.02.004>.
- Blasiak, Robert, R. Wynberg, K. Grorud-Colvert, S. Thambisetty, N. M. Bandarra, A. V. M. Canário, J. da Silva, et al. 2020. “The Ocean Genome and Future Prospects for Conservation and Equity.” *Nature Sustainability* 3 (8): 588–96. <https://doi.org/10.1038/s41893-020-0522-9>.
- Bond, Molly R., and Deborah Scott. 2020. “Digital Biopiracy and the (Dis)Assembling of the Nagoya Protocol.” *Geoforum* 117 (December):24–32. <https://doi.org/10.1016/j.geoforum.2020.09.001>.
- Broggiato, Arianna, Paul Dunshirn, Marcel Jaspars, and Sergio Peña-Neira. 2024. “Monetary and Non-Monetary Benefit Sharing under the BBNJ Agreement.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. <https://doi.org/10.2139/ssrn.4897384>.
- Broggiato, Arianna, Thomas Vanagt, Laura E. Lallier, Marcel Jaspars, Geoff Burton, and Dominic Muyldermans. 2018. “Mare Geneticum: Balancing Governance of Marine Genetic Resources in International Waters.” *The International Journal of Marine and Coastal Law* 33 (1): 3–33. <https://doi.org/10.1163/15718085-13310030>.
- Brown, Abbe E. L. 2025. “The Place of Intellectual Property Under the BBNJ Agreement.” In *Decoding Marine Genetic Resource Governance Under the BBNJ Agreement*, edited by Fran Humphries, 213–24. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72100-7_9.
- Cambon-Bonavita, M.-A., G. Raguénès, J. Jean, P. Vincent, and J. Guezennec. 2002. “A Novel Polymer Produced by a Bacterium Isolated from a Deep-Sea Hydrothermal Vent Polychaete Annelid.” *Journal of Applied Microbiology* 93 (2): 310–15. <https://doi.org/10.1046/j.1365-2672.2002.01689.x>.
- Chen, Jianwei, Yangyang Jia, Ying Sun, Kun Liu, Changhao Zhou, Chuan Liu, Denghui Li, et al. 2024. “Global Marine Microbial Diversity and Its Potential in Bioprospecting.” *Nature* 633 (8029): 371–79. <https://doi.org/10.1038/s41586-024-07891-2>.
- Contreras, Jorge L., and Bartha M. Knoppers. 2018. “The Genomic Commons.” *Annual Review of Genomics and Human Genetics* 19 (1): 429–53. <https://doi.org/10.1146/annurev-genom-083117-021552>.
- Convention on Biological Diversity. 1992. “The Convention on Biological Diversity.” <https://www.cbd.int/doc/legal/cbd-en.pdf>.
- . 2022. *COP Decision 15/19 on Digital Sequence Information on Genetic Resources*. <https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-09-en.pdf>.
- . 2024. “COP Decision 16/2 on Digital Sequence Information on Genetic Resources.”
- Cross, Samuel, Yeanuk Rho, Henna Reddy, Toby Pepperrell, Florence Rodgers, Rhiannon Osborne, Ayolola Eni-Olotu, Rishi Banerjee, Sabrina Wimmer, and Sarai Kestra. 2021. “Who Funded the Research behind the Oxford–AstraZeneca COVID-19 Vaccine?” *BMJ Global Health* 6 (12): e007321. <https://doi.org/10.1136/bmjgh-2021-007321>.

- Dedeurwaerdere, Tom. 2010. "Self-Governance and International Regulation of the Global Microbial Commons: Introduction to the Special Issue on the Microbial Commons." *International Journal of the Commons* 4 (1): 390–403. <https://doi.org/10.18352/ijc.215>.
- Drew, Joshua A. 2005. "Use of Traditional Ecological Knowledge in Marine Conservation." *Conservation Biology* 19 (4): 1286–93. <https://doi.org/10.1111/j.1523-1739.2005.00158.x>.
- Enyew, Endalew Lijalem. 2022. "Sailing with TWAIL: A Historical Inquiry into Third World Perspectives on the Law of the Sea." *Chinese Journal of International Law* 21 (3): 439–97. <https://doi.org/10.1093/chinesejil/jmac028>.
- Eren, A. Murat, and Tom O. Delmont. 2024. "Bioprospecting Marine Microbial Genomes to Improve Biotechnology." *Nature*, September. <https://doi.org/10.1038/d41586-024-02661-6>.
- Eslava, Luis, Michael Fakhri, and Vasuki Nesiah. 2017. *Bandung, Global History, and International Law: Critical Pasts and Pending Futures*. Cambridge University Press.
- Frischmann, Brett M., Michael J. Madison, and Katherine J Strandburg. 2014. "Madison_etal-2014-Governing_Knowledge_Commons.Pdf." Oxford University Press.
- Frison, Christine. 2018. "Redesigning the Global Seed Commons: Law and Policy for Agrobiodiversity and Food Security." *Routledge*. <https://www.tandfonline.com/doi/full/10.1080/09615768.2019.1645475>.
- Geary, Janis, and Tania Bubela. 2019. "Governance of a Global Genetic Resource Commons for Non-Commercial Research: A Case-Study of the DNA Barcode Commons." *International Journal of the Commons* 13 (1): 205. <https://doi.org/10.18352/ijc.859>.
- Golan, Jacob, KatieLee Riddle, Maui Hudson, Jane Anderson, Natalie Kusabs, and Tim Coltman. 2022. "Benefit Sharing: Why Inclusive Provenance Metadata Matter." *Frontiers in Genetics* 13. <https://www.frontiersin.org/articles/10.3389/fgene.2022.1014044>.
- Gollin, Michael. 2008. "Biopiracy Started with a Bounce." *Nature* 451 (7182): 1055–1055. <https://doi.org/10.1038/4511055a>.
- Gottlieb, Hiroko Muraki, Daniel Kachelriess, and Lydia Slobodian. 2025. "Understanding the Preamble, Objectives and Principles of the BBNJ Agreement: A Focus on the Fair and Equitable Sharing of Benefits of Marine Genetic Resources." In *Decoding Marine Genetic Resource Governance Under the BBNJ Agreement*, edited by Fran Humphries, 95–124. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72100-7_4.
- Grotius, Hugo. 1914. *The Freedom of the Seas, Or, The Right Which Belongs to the Dutch to Take Part in the East Indian Trade*. New York, NY, USA: Oxford University Press.
- Guellec, Dominique, Catalina Martinez, and Pluvia Zuniga. 2012. "Pre-Emptive Patenting: Securing Market Exclusion and Freedom of Operation." *Economics of Innovation and New Technology* 21 (1): 1–29. <https://ideas.repec.org/a/taf/ecinnt/v21y2012i1p1-29.html>.
- Halewood, Michael. 2013. "What Kind of Goods Are Plant Genetic Resources for Food and Agriculture? Towards the Identification and Development of a New Global Commons." *International Journal of the Commons* 7 (2). <https://doi.org/10.18352/ijc.412>.
- Hannigan, John. 2015. *The Geopolitics of Deep Oceans*. John Wiley & Sons.
- Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science (American Association for the Advancement of Science)* 162 (3859): 1243–48. <https://doi.org/10.1126/science.162.3859.1243>.
- Heinrich, Michael, Francesca Scotti, Adolfo Andrade-Cetto, Monica Berger-Gonzalez, Javier Echeverría, Fabio Friso, Felipe Garcia-Cardona, et al. 2020. "Access and Benefit Sharing Under the Nagoya Protocol—Quo Vadis? Six Latin American Case Studies Assessing Opportunities and Risk." *Frontiers in Pharmacology* 11 (June). <https://doi.org/10.3389/fphar.2020.00765>.

- Heller, Michael. 2013. "The Tragedy of the Anticommons: A Concise Introduction and Lexicon." *The Modern Law Review* 76 (1): 6–25. <https://doi.org/10.1111/1468-2230.12000>.
- Hess, Charlotte, and Elinor Ostrom. 2007. *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, Mass: MIT Press.
- Humphries, Fran, Todd Berry, and Hiroko Muraki Gottlieb. 2025. "Bridging Divides: The Evolution of Marine Genetic Resource Governance Beyond National Jurisdiction." In *Decoding Marine Genetic Resource Governance Under the BBNJ Agreement*, edited by Fran Humphries, 1–25. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72100-7_1.
- Humphries, Fran, Muriel Rabone, and Marcel Jaspars. 2021. "Traceability Approaches for Marine Genetic Resources Under the Proposed Ocean (BBNJ) Treaty." *Frontiers in Marine Science* 8 (April):661313. <https://doi.org/10.3389/fmars.2021.661313>.
- Hurst, Dermot, Torger Børresen, Lisa Almesjö, Fien Raedemaeker, and Steinar Bergseth. 2016. *Marine Biotechnology Strategic Research and Innovation Roadmap - Insights to the Future Direction of European Marine Biotechnology*. Marine Biotechnology ERA-NET.
- International Nucleotide Sequence Database Collaboration. 2023. "INSDC Spatiotemporal Metadata – Minimum Standards Update (03-03-2023)." <https://www.insdc.org/news/insdc-spatiotemporal-metadata-minimum-standards-update-03-03-2023/>.
- Jackson, Joe. 2008. *The Thief at the End of the World: Rubber, Power, and the Seeds of Empire*. New York: Viking.
- Jaffe, Adam B., and Gaétan de Rassenfosse. 2017. "Patent Citation Data in Social Science Research: Overview and Best Practices." *Journal of the Association for Information Science and Technology* 68 (6): 1360–74. <https://doi.org/10.1002/asi.23731>.
- Jaspars, Marcel, Fran Humphries, and Muriel Rabone. 2021. "Tracing Options for Marine Genetic Resources from within National Jurisdictions." Report prepared for Commonwealth Secretariat. <https://eprints.qut.edu.au/227075/1/102475803.pdf>.
- Jaspars, Marcel, Donatella De Pascale, Jeanette H. Andersen, Fernando Reyes, Alexander D. Crawford, and Adrianna Ianora. 2016a. "The Marine Biodiscovery Pipeline and Ocean Medicines of Tomorrow." *Journal of the Marine Biological Association of the United Kingdom* 96 (1): 151–58. <https://doi.org/10.1017/S0025315415002106>.
- . 2016b. "The Marine Biodiscovery Pipeline and Ocean Medicines of Tomorrow." *Journal of the Marine Biological Association of the United Kingdom* 96 (1): 151–58. <https://doi.org/10.1017/S0025315415002106>.
- Kachelriess, Daniel, Paul Dunshirn, Arne Langlet, Abbe E. L. Brown, and Amber H. Scholz. 2025. "Marine Genetic Resources and Digital Sequence Information Under the BBNJ Agreement—Interlinkages with Other Access and Benefit-Sharing Frameworks." In *Decoding Marine Genetic Resource Governance Under the BBNJ Agreement*, edited by Fran Humphries, 241–51. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72100-7_11.
- Kamau, Evanson, and Gerd Winter. 2013. "An Introduction to the International ABS Regime and a Comment on Its Transposition by the EU." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2387876>.
- Kang, Hyo Yoon. 2020. "Patents as Assets: Intellectual Property Rights as Market Subjects and Objects," July. <https://doi.org/10.7551/mitpress/12075.003.0004>.
- Keller, Evelyn Fox. 2002. *The Century of the Gene*. First Harvard University Press paperback edition. Harvard University Press.
- Keohane, Robert O., and Elinor Ostrom, eds. 1995. *Local Commons and Global Interdependence: Heterogeneity and Cooperation in Two Domains*. Sage Publications.

- Klünker, Irma. 2024. "The tragedy of the access and benefit-sharing anticommons: Property rights on genetic resources and digital sequence informationDie Tragödie des Access and Benefit-Sharings: Nutzungsregeln für genetische Ressourcen und digitale Sequenzinformation." Trier University. <https://doi.org/10.25353/UBTR-D12B-BCFA-544E>.
- Lange, Matthias, Blaise T F Alako, Guy Cochrane, Mehmood Ghaffar, Martin Mascher, Pia-Katharina Habekost, Upneet Hillebrand, et al. 2021. "Quantitative Monitoring of Nucleotide Sequence Data from Genetic Resources in Context of Their Citation in the Scientific Literature." *GigaScience* 10 (12): giab084. <https://doi.org/10.1093/gigascience/giab084>.
- Langlet, Arne, and Paul Dunshirn. 2023. "Traceability Options for Marine Genetic Resource from Areas beyond National Jurisdiction." <https://highseasalliance.org/resources/paper-traceability-options-for-marine-genetic-resource-from-areas-beyond-national-jurisdiction/>.
- Lara-Lopez, Ana, Luis Valdés, Roberto de Pinho, and Henrik Enevoldsen. 2020. "Analysis of Ocean Science Production and Impact." In *Global Ocean Science Report 2020: Charting Capacity for Ocean Sustainability*. United Nations. <https://doi.org/10.18356/9789216040048>.
- Lawson, Charles, Fran Humphries, Marcel Jaspars, and Muriel Rabone. 2025. "Data Management and the 'BBNJ Standardized Batch Identifier' Under the BBNJ Agreement." In *Decoding Marine Genetic Resource Governance Under the BBNJ Agreement*, edited by Fran Humphries, 253–69. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72100-7_12.
- Lawson, Charles, Michelle Rourke, and Fran Humphries. 2020. "Information as the Latest Site of Conflict in the Ongoing Contests about Access to and Sharing the Benefits from Exploiting Genetic Resources." *Queen Mary Journal of Intellectual Property* 10 (1): 7–33. <https://doi.org/10.4337/qmjip.2020.01.01>.
- Leal, Miguel C., Jaime M. Anaya-Rojas, Murray H. G. Munro, John W. Blunt, Carlos J. Melian, Ricardo Calado, and Moritz D. Lürig. 2020. "Fifty Years of Capacity Building in the Search for New Marine Natural Products." *Proceedings of the National Academy of Sciences* 117 (39): 24165–72. <https://doi.org/10.1073/pnas.2007610117>.
- Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52. <https://doi.org/10.1017/S0003055405051762>.
- Marciniak, Konrad. 2020. "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*, edited by Tomas Heidar, 40–64. Brill | Nijhoff. https://doi.org/10.1163/9789004437753_005.
- Marx, Matt, and Aaron Fuegi. 2020. "Reliance on Science: Worldwide Front-page Patent Citations to Scientific Articles." *Strategic Management Journal* 41 (9): 1572–94. <https://doi.org/10.1002/smj.3145>.
- Mauss, Marcell. 1956. "The Gift: Forms and Functions of Exchange in Archaic Societies." *The University Press Aberdeen* 21 (2): 249. <https://doi.org/10.2307/2088549>.
- Maxson Jones, Kathryn, Rachel A. Ankeny, and Robert Cook-Deegan. 2018. "The Bermuda Triangle: The Pragmatics, Policies, and Principles for Data Sharing in the History of the Human Genome Project." *Journal of the History of Biology* 51 (4): 693–805. <https://doi.org/10.1007/s10739-018-9538-7>.
- Mc Cartney, Ann M., M. A. Head, K. S. Tsosie, B. Sterner, J. R. Glass, S. Paez, J. Geary, and M. Hudson. 2023. "Indigenous Peoples and Local Communities as Partners in the Sequencing of Global Eukaryotic Biodiversity." *Npj Biodiversity* 2 (1): 1–12. <https://doi.org/10.1038/s44185-023-00013-7>.
- Morin, Jean-Frédéric. 2014. "Paradigm Shift in the Global IP Regime: The Agency of Academics." *Review of International Political Economy* 21 (2): 275–309. <https://doi.org/10.1080/09692290.2013.819812>.

- Muñoz-García, Melania, Alejandro Lago, and Amber H Scholz. 2024. “Access and Benefit-Sharing Indicators for the Kunming-Montreal Global Biodiversity Framework.” <https://www.cbd.int/doc/c/6920/4e1e/8a6ba925279ea19033eb8ed2/sbstta-26-inf-12-en.pdf>.
- Noreña – P, Alejandra, Andrea González Muñoz, Jeanneth Mosquera-Rendón, Kelly Botero, and Marco A. Cristancho. 2018. “Colombia, an Unknown Genetic Diversity in the Era of Big Data.” *BMC Genomics* 19 (8): 859. <https://doi.org/10.1186/s12864-018-5194-8>.
- Nyika, Aceme. 2009. “Ethical and Practical Challenges Surrounding Genetic and Genomic Research in Developing Countries.” *Acta Tropica*, Health Research in Africa: Ethical and Practical Challenges, 112 (November):S21–31. <https://doi.org/10.1016/j.actatropica.2009.07.034>.
- Oldham, Paul. 2009. “An Access and Benefit-Sharing Commons? The Role of Commons/Open Source Licences in the International Regime on Access to Genetic Resources and Benefit-Sharing.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1438027>.
- . 2020. “Digital Sequence Information - Technical Aspects.” <https://doi.org/10.13140/RG.2.2.15833.24163>.
- Oldham, Paul, Claudio Chiarolla, and Siva Thambisetty. 2023. “Digital Sequence Information in the UN High Seas Treaty: Insights from the Global Biodiversity Framework-Related Decisions.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4343130>.
- Oldham, Paul, Stephen Hall, and Oscar Forero. 2013. “Biological Diversity in the Patent System.” Edited by Francesco Pappalardo. *PLoS ONE* 8 (11): e78737. <https://doi.org/10.1371/journal.pone.0078737>.
- Oldham, Paul, Jasmine Kindness, Emmanuell Davidson, Amelia Westmoreland, Thomas Vanagt, and Marcel Jaspars. 2025. “Study on ‘Marine Genetic Resources’ Market Value and State of the Art of Commercialisation of Related Products in the Context of the BBNJ Negotiations: Final Report.” *European Climate, Infrastructure and Environment Executive Agency*. <https://data.europa.eu/doi/10.2926/7945806>.
- Oldham, Paul, Stephen Hall, Colin Barnes, Catherine Oldham, Mark Cutter, Natasha Burns, and Leonie Kindness. 2014. “Valuing the Deep: Marine Genetic Resources in Areas Beyond National Jurisdiction.” <https://bookdown.org/poldham/valuingthedeep/>.
- Oldham, Paul, and Siva Thambisetty. 2023. “ONEST: The Middle Way for Monetary Benefit Sharing in BBNJ Negotiations.” Zenodo. <https://doi.org/10.5281/zenodo.7573700>.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Political Economy of Institutions and Decisions. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511807763>.
- . 2009. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400831739>.
- Prathapan, K. Divakaran, Rohan Pethiyagoda, Kamaljit S. Bawa, Peter H. Raven, Priyadarsanan Dharma Rajan, and 172 co-signatories from 35 countries. 2018. “When the Cure Kills—CBD Limits Biodiversity Research.” *Science* 360 (6396): 1405–6. <https://doi.org/10.1126/science.aat9844>.
- Puglisi, Melany P., Jennifer M. Sneed, Koty H. Sharp, Raphael Ritson-Williams, and Valerie J. Paul. 2014. “Marine Chemical Ecology in Benthic Environments.” *Nat. Prod. Rep.* 31 (11): 1510–53. <https://doi.org/10.1039/C4NP00017J>.
- Rabitz, Florian. 2017. *The Global Governance of Genetic Resources: Institutional Change and Structural Constraints*. London: Routledge. <https://doi.org/10.4324/9781315271316>.
- Raposo, Débora S, Martha L Cepeda, Barbara Ebert, and Amber H Scholz. 2023. “Challenges and Opportunities of Geographical Origin Information in DSI Benefit-Sharing: A Global Analysis from the Academic Sector.”

- Raustiala, Kal, and David G. Victor. 2004. "The Regime Complex for Plant Genetic Resources." *International Organization* 58 (02). <https://doi.org/10.1017/S0020818304582036>.
- Rayne, Aisling, Stephanie Blair, Matthew Dale, Brendan Flack, John Hollows, Roger Moraga, Riki N. Parata, et al. 2022. "Weaving Place-Based Knowledge for Culturally Significant Species in the Age of Genomics: Looking to the Past to Navigate the Future." *Evolutionary Applications* 15 (5): 751–72. <https://doi.org/10.1111/eva.13367>.
- Relly, Eduardo. 2024. "Brazil's Implementation of Access and Benefit-Sharing and the Nagoya Protocol: Analyzing Some Trends and Positions in the Ongoing Debate." *Genetic Resources* 5 (10): 65–80. <https://doi.org/10.46265/genresj.GKTE3850>.
- Richardson, Katherine, Will Steffen, Wolfgang Lucht, Jørgen Bendtsen, Sarah E. Cornell, Jonathan F. Donges, Markus Drüke, et al. 2023. "Earth beyond Six of Nine Planetary Boundaries." *Science Advances* 9 (37): eadh2458. <https://doi.org/10.1126/sciadv.adh2458>.
- Roach, Michael, and Wesley M. Cohen. 2013. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." *Management Science* 59 (2): 504–25. <https://doi.org/10.1287/mnsc.1120.1644>.
- Robinson, Daniel F., and Evana Wright. 2023. "Access and Benefit-Sharing: What Indicators to Measure 'Success'?" *Trends in Ecology & Evolution* 38 (9): 795–98. <https://doi.org/10.1016/j.tree.2023.05.016>.
- Rohden, Fabian, Sixing Huang, Gabriele Dröge, and Amber Hartman Scholz. 2020. "Combined Study on Digital Sequence Information in Public and Private Databases and Traceability." <https://www.cbd.int/abs/DSI-peer/Study-Traceability-databases.pdf>.
- Rotter, Ana, Michéle Barbier, Francesco Bertoni, Atle M. Bones, M. Leonor Cancela, Jens Carlsson, Maria F. Carvalho, et al. 2021. "The Essentials of Marine Biotechnology." *Frontiers in Marine Science* 8 (March): 629629. <https://doi.org/10.3389/fmars.2021.629629>.
- Salem, Hassan, and Martin Kaltenpoth. 2023. "The Nagoya Protocol and Its Implications for Microbiology." *Nature Microbiology* 8 (12): 2234–37. <https://doi.org/10.1038/s41564-023-01532-6>.
- Scholz, Amber Hartman, Fran Humphries, Thomas Vanagt, and Marcel Jaspars. 2023. "A New Dawn for Global Benefit-Sharing: Capitalizing on the Global Biodiversity Framework for Marine Genetic Resources from Areas Beyond National Jurisdiction." https://www.iucn.org/sites/default/files/2023-02/bbnj_icg5bis_policy_brief_global_benefit_sharing_1.pdf.
- Scholz, Amber Hartman, Matthias Lange, Pia Habekost, Paul Oldham, Ibon Cancio, Guy Cochrane, and Jens Freitag. 2021. "Myth-Busting the Provider-User Relationship for Digital Sequence Information." *GigaScience* 10 (12): giab085. <https://doi.org/10.1093/gigascience/giab085>.
- Sett, Scarlett, W. John Kress, Michael Halewood, David Nicholson, Genuar Nuñez-Vega, Davide Faggionato, Mathieu Rouard, et al. 2024. "Harmonize Rules for Digital Sequence Information Benefit-Sharing across UN Frameworks." *Nature Communications* 15 (1): 8745. <https://doi.org/10.1038/s41467-024-52994-z>.
- Sierra-Correa, Paula Cristina, Kwame A. Koranteng, Itahisa Déniz-González, Christian Wexels Riser, Margareth Serapio Kyewalyanga, Ntahondi Mcheche Nyandwi, J. Magdalena Santana-Casiano, Francesca Santoro, Kirsten Isensee, and Dongho Youm. 2020. "Research Capacity and Infrastructure." In *Global Ocean Science Report 2020: Charting Capacity for Ocean Sustainability*. United Nations. <https://doi.org/10.18356/9789216040048>.
- Sigwart, Julia D., Robert Blasiak, Marcel Jaspars, Jean-Baptiste Jouffray, and Deniz Tasdemir. 2021. "Unlocking the Potential of Marine Biodiscovery." *Natural Product Reports* 38 (7): 1235–42. <https://doi.org/10.1039/D0NP00067A>.

- Sirakaya, Aysegul. 2022. "Where Access and Benefit-Sharing Comes from: A Historical Overview." *Genetic Resources* 3 (6): 74–88. <https://doi.org/10.46265/genresj.PPUF5169>.
- Stange, Madlen, Rowan D. H. Barrett, and Andrew P. Hendry. 2021. "The Importance of Genomic Variation for Biodiversity, Ecosystems and People." *Nature Reviews Genetics* 22 (2): 89–105. <https://doi.org/10.1038/s41576-020-00288-7>.
- Strasser, Bruno J. 2012. "Data-Driven Sciences: From Wonder Cabinets to Electronic Databases." *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 85–87. <https://doi.org/10.1016/j.shpsc.2011.10.009>.
- Tiller, Rachel, Elizabeth De Santo, Elizabeth Mendenhall, and Elizabeth Nyman. 2019. "The Once and Future Treaty: Towards a New Regime for Biodiversity in Areas beyond National Jurisdiction." *Marine Policy* 99 (January):239–42. <https://doi.org/10.1016/j.marpol.2018.10.046>.
- Timmermans, Stefan, and Iddo Tavory. 2012. "Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis." *Sociological Theory* 30 (3): 167–86. <https://doi.org/10.1177/0735275112457914>.
- Tolochko, Petro, and Alice B.M. Vadrot. 2021. "The Usual Suspects? Distribution of Collaboration Capital in Marine Biodiversity Research." *Marine Policy* 124 (February):104318. <https://doi.org/10.1016/j.marpol.2020.104318>.
- Tortorella, Emiliana, Pietro Tedesco, Fortunato Palma Esposito, Grant Garren January, Renato Fani, Marcel Jaspars, and Donatella de Pascale. 2018. "Antibiotics from Deep-Sea Microorganisms: Current Discoveries and Perspectives." *Marine Drugs* 16 (10): 355. <https://doi.org/10.3390/md16100355>.
- Tuerk, Helmut. 2017. "The Common Heritage of Mankind after 50 Years." *Indian Journal of International Law* 57 (3): 259–83. <https://doi.org/10.1007/s40901-018-0085-8>.
- United Nations General Assembly. 2023. "Agreement under the United Nations Convention on the Law of the Sea on the Conservation and Sustainable Use of Marine Biological Diversity of Areas beyond National Jurisdiction." <https://documents-dds-ny.un.org/doc/UNDOC/LTD/N23/177/28/PDF/N2317728.pdf>.
- Vadrot, A.B.M., A. Langlet, and I. Tessnow-von Wysocki. 2021. "Who Owns Marine Biodiversity? Contesting the World Order through the 'Common Heritage of Humankind' Principle." *Environmental Politics* 31 (2): 226–50. <https://doi.org/10.1080/09644016.2021.1911442>.
- Vieira, Monica Brito. 2003. "Mare Liberum vs. Mare Clausum : Grotius, Freitas, and Selden's Debate on Dominion over the Seas." *Journal of the History of Ideas* 64 (3): 361–77. <https://muse.jhu.edu/pub/1/article/47665>.
- Vierros, Marjo, and William Aalbersberg. 2010. *Traditional Marine Management Areas of the Pacific in the Context of National and International Law and Policy*. United Nations University Traditional Knowledge Initiative.
- Walckiers, Pierre, Christine Frison, and Sylvain Aubry. 2024. "A Roadmap to Equity in Pandemic Preparedness, Prevention and Response." SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.4708224>.
- Webster, Michael S., Madhavi A. Colton, Emily S. Darling, Jonathan Armstrong, Malin L. Pinsky, Nancy Knowlton, and Daniel E. Schindler. 2017. "Who Should Pick the Winners of Climate Change?" *Trends in Ecology & Evolution* 32 (3): 167–73. <https://doi.org/10.1016/j.tree.2016.12.007>.
- White, Kimberly N., Karen Tenney, and Phillip Crews. 2017. "The Bengamides: A Mini-Review of Natural Sources, Analogues, Biological Properties, Biosynthetic Origins, and Future Prospects." *Journal of Natural Products* 80 (3): 740–55. <https://doi.org/10.1021/acs.jnatprod.6b00970>.

- Wilkes, R. A., and L. Aristilde. 2017. "Degradation and Metabolism of Synthetic Plastics and Associated Products by *Pseudomonas* Sp.: Capabilities and Challenges." *Journal of Applied Microbiology* 123 (3): 582–93. <https://doi.org/10.1111/jam.13472>.
- Wolfrum, Rüdiger. 2023. "Common Heritage of Mankind." *Oxford University Press*.
- Wolter, Hendrik, and Stefanie Sievers-Glotzbach. 2019. "Bridging Traditional and New Commons: The Case of Fruit Breeding." *International Journal of the Commons* 13 (1): 303. <https://doi.org/10.18352/ijc.869>.
- Wynberg, Rachel. 2023. "Biopiracy: Crying Wolf or a Lever for Equity and Conservation?" *Research Policy* 52 (2): 104674. <https://doi.org/10.1016/j.respol.2022.104674>.
- Zhang, Chuwen, Yongyi Peng, Xinyue Liu, Jieni Wang, and Xiyang Dong. 2024. "Deep-Sea Microbial Genetic Resources: New Frontiers for Bioprospecting." *Trends in Microbiology* 32 (4): 321–24. <https://doi.org/10.1016/j.tim.2024.01.002>.