



OPEN ACCESS

EDITED BY

Jacqueline Espenilla,
University of the Philippines Diliman,
Philippines

REVIEWED BY

Charles Lawson,
Griffith University, Australia
Youngdawng Moh,
Korea Institute of Ocean Science and
Technology, Republic of Korea

*CORRESPONDENCE

Erik Zhivkoplías
✉erik.zhivkoplías@su.se

RECEIVED 18 March 2026

REVISED 23 April 2026

ACCEPTED 04 May 2026

PUBLISHED 21 May 2026

CITATION

Zhivkoplías E, Dunshirn P, Jouffray J-B,
Jaspars M and Blasiak R (2026) Blue
genes, green promises: linking marine
gene patents to biotechnology and
sustainable development.
Front. Mar. Sci. 13:1833479.
doi: 10.3389/fmars.2026.1833479

COPYRIGHT

© 2026 Zhivkoplías, Dunshirn, Jouffray,
Jaspars and Blasiak. This is an open-
access article distributed under the terms
of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

Blue genes, green promises: linking marine gene patents to biotechnology and sustainable development

Erik Zhivkoplías^{1,2*}, Paul Dunshirn³, Jean-Baptiste Jouffray^{1,4},
Marcel Jaspars⁵ and Robert Blasiak¹

¹Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden, ²Stockholm University Center for Circular and Sustainable Systems, Stockholm University, Stockholm, Sweden, ³Environmental Politics Research Group, Department of Political Science, University of Vienna, Vienna, Austria, ⁴Stanford Center for Ocean Solutions, Stanford, CA, United States, ⁵Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Old Aberdeen, United Kingdom

Marine genetics play an essential role in biodiversity studies, ecosystem monitoring, and biotechnology, yet little is known about how marine genetic sequences are translated into commercial applications. Although international frameworks such as the BBNJ Agreement and the Convention on Biological Diversity address access and benefit sharing, effective implementation requires empirical insight into how marine genes are used. Here we analyze the Marine Bioprospecting Patent Database, a global catalog of marine-derived genes appearing in patents, by combining structural topic modeling and large language model-assisted classification of 1,676 patent texts to map technological themes and actor contributions across healthcare, agriculture, industrial, and foundational biotechnology. Private-sector filings prioritize industrial and agricultural applications, whereas research institutes and governmental actors disproportionately contribute foundational, cross-cutting tools that enable downstream innovation. A curated subset of 221 patent claims mentioning marine enzymes illustrates applications ranging from bioenergy and green chemistry feedstocks to diagnostics and therapeutics. While many inventions align with sustainability objectives, others deliver incremental gains or support applications with ambiguous, if not adverse, environmental implications. Targeted policy, funding, and stewardship are needed to maximize net sustainability benefits and ensure equitable benefit-sharing under international frameworks.

KEYWORDS

BBNJ agreement, digital sequence information, equity, gene patents, marine biotech, marine biotechnology, marine biotechnology and natural products, marine genetic resources

1 Introduction

The global push for sustainable development encourages a holistic vision that recognizes and acts upon the inherent links between our economy, society, and the environment. The environment and its biodiversity provide diverse benefits that underpin humanity's societal and economic well-being. These benefits are multifaceted, ranging from ecosystem functions like climate regulation and food provision to the raw materials for innovation, which

collectively contribute directly or indirectly to all 17 Sustainable Development Goals (SDGs) (Blicharska et al., 2019). For instance, the components of biodiversity - ecosystems, species, and genes - deliver benefits that contribute to food security and improved nutrition by providing crop pollination and pest control (SDG 2), to improved health outcomes by serving as a source for medicines and diluting disease reservoirs (SDG 3), and to economic growth by providing cost-efficient solutions through bioprospecting (SDG 8). Many novel and emerging benefits from biodiversity are associated with the genetic resources of individual species, and often surprising human innovation processes.

A powerful recent example is the development of the drug semaglutide, marketed as Ozempic and Wegovy. This drug, used to treat type 2 diabetes and obesity, is a synthetic version of a hormone discovered in the saliva of *Heloderma suspectum*, a large lizard endemic to arid regions of the United States and Mexico. The hormone exendin-4, extracted from *H. suspectum*, was found to be a long-lasting analog of the human hormone GLP-1, which regulates blood sugar and appetite (Kolterman et al., 2003). This discovery and subsequent bio-inspired innovation, originating unexpectedly from a terrestrial reptile, exemplifies the immense, often untapped, potential of the planet's biodiversity to yield revolutionary and more sustainable tools for human well-being. The economic impact has been so profound that some commentators have noted its significant contribution to the GDP of Denmark, home to the pharmaceutical company Novo Nordisk, highlighting how a single genetic discovery can influence a national economy (<https://www.npr.org/sections/planet-money/2024/07/26/g-s1-13534/ozempic-biggest-side-effect-denmark-pharmastate>). Albeit an exception rather than the rule in biotechnology, the Ozempic case illustrates the commercial potential of terrestrial biodiversity.

All evidence points to the global ocean being a rich and largely untapped repository of genetic resources with vast potential (Sigwart et al., 2021). The role these marine genetic resources (MGR) can play in biotechnological advances, however, remains far less clear. This gap is reflected in the small number of FDA-approved drugs directly linked to marine biomolecules (<https://www.marinepharmacology.org/approved>; Francesch et al., 2024), while non-medical applications are even rarer. Around a quarter of all clinically approved drugs (n=1,881 as of September 2019) are based on natural products (Newman and Cragg, 2020), yet fewer than 10% of identified natural products derive from marine organisms. Nevertheless, marine natural products show up to fourfold higher success rates in discovery pipelines (Sigwart et al., 2021). This disparity arises not from a lack of potential but from complex barriers, including limited confidence that nature can provide tractable solutions, slow and high-risk development processes with low success rates and repeated rediscovery, and persistent challenges in tracking and documenting genetic resources through R&D and commercialization - complicating efforts to link downstream innovations to their genetic origins and to demonstrate compliance with ABS provisions under the CBD (Ruiz Muller and Lapeña, 2007), which can be a necessary condition for providers and users to enter equitable research partnerships and commercialization pathways that support innovation and broaden market participation. At the same time, structural problems in

marine biodiscovery persist: chronic underfunding of taxonomy and organismal biology, short grant cycles ill-suited to high-risk efforts, and limited collaboration between academics and industry in the early stages of the marine biodiscovery pipeline (Sigwart et al., 2021). Large-scale production also remains a critical bottleneck. Marine lead compounds are often present at extremely low abundance in source organisms, often because highly potent metabolites are metabolically expensive to produce, making it difficult to secure a sustainable supply and to recreate complex structures at scale (Donia and Hamann, 2003).

A further disconnect exists in understanding not only the commercial benefits of marine biotechnology but also how it contributes to sustainable development by delivering broader economic and societal benefits (Deep-Ocean Stewardship Initiative (DOSI), 2025). Marine biotechnology is frequently conceptualized as a crucial component of the "blue economy" - an aspirational ocean economy that is sustainable and equitable, but this potential remains largely unrealized (Blasiak et al., 2023). This is particularly relevant for the Global South, where endemic marine biodiversity offers important opportunities for more equitable and sustainable growth, though realizing these benefits depends on governance, capacity, and fair benefit-sharing rather than guaranteed "resource riches" (Cisneros-Montemayor et al., 2021; Bennett et al., 2019).

Here, we address these gaps by applying a novel mixed-methods approach to analyze patent data, a recognized indicator of innovation and technological progress (Straccamore et al., 2023; Pugliese et al., 2019). We build on initial work using patent analyses to identify the organizations and applications driving market-oriented solutions (Chiarolla et al., 2025; Oldham et al., 2025; Hall et al., 2014; Arrieta, 2010), and advance these efforts by conducting an exhaustive analysis of the MARine Bioprospecting PATent database (MABPAT), a publicly available catalogue of marine gene sequences referenced in patent filings (Zhivkoplías et al., 2024), using recent advances in computational text-mining and topic modeling approaches (Table 1) (Heberling, 2021). Our findings offer new insights into the technological applications of MGRs and their potential contribution to sustainable development. Specifically, we (1) map the semantic topics involving marine genes and their temporal trends; (2) provide a categorical analysis of the types of biotechnological innovation with respect to applicants filing marine gene patents and their headquarter locations; (3) provide predictive assessment of product types based on semantic analysis; and (4) estimate the primary contribution of MGR to sustainable development.

2 Methods

2.1 Data collection and wrangling

We extracted patent application numbers from the MABPAT database, a publicly available catalogue of marine gene sequences referenced in patent filings, focusing on patents written in English. From MABPAT, we also gathered detailed information about the patent applicants, including their type and country of origin, as well as any genetic sequences attached to the patent applications.

TABLE 1 Summary of methods used.

Method name	Method description	Data used	Resulted in
Structural topic modeling	An unsupervised probabilistic topic-modeling method that identifies latent topics by grouping words that co-occur across documents and, unlike basic topic models, allows document metadata to influence topic prevalence and content	Full text patent data (1,676 patent applications)	12 topics: Bioluminescence, Crop traits, Programmable bacteria, Omega-3 production, Bio-based chemicals, Crop engineering, Biopharmaceuticals, Solar microbes, Protein engineering, Molecular engineering, DNA enzymes, Unnatural amino acids
LLM categorization	Supervised classification using prompted large language model to identify keywords and assign documents to predefined thematic categories	Abstracts and claims (1,273 patent applications)	4 biotechnological themes: Foundational, Industrial, Healthcare, Agricultural
Potential product types estimation	Semi-automated extraction using fuzzy string matching of enzyme names followed by manual validation to infer potential product or method types from claim language.	Claims section of patents and enzyme names extracted from <i>title</i> column of MABPAT (221 patent applications)	13 potential product or method types: Food, feed & nutraceuticals; Molecular biology toolbox; Green chemistry & environmental technologies; Therapeutics & treatments; Bioenergy & carbon utilization; Vaccines & production support; GM crops; Blue biotechnology; Clinical diagnostics; Biosensors; Nanotechnology & advanced materials; Optimized microbial systems; Unclear or unspecified

A patent may be filed in multiple jurisdictions, resulting in multiple patent documents but with more or less the same content. Such sets of related patents are considered to be patent families. Accounting for such family ties is a common technique used to avoid duplicates in datasets (e.g., Oldham et al., 2025). We used the patent API of the database The Lens (<https://www.lens.org/>) to gather the same information as mentioned above for all patent documents within the families tied to the initial documents. Within each of the families, we selected only the most relevant documents by prioritizing 1) granted patents, 2) granted patents in English, and 3) the earliest granted patent in English. If a family had no granted patent, we would run through the same ranking criteria for patent applications (i.e. prioritizing language over date of publication). This reduced our dataset to 1,676 family-checked patent documents for which we then scraped full texts from Google Patents.

Different parts of this collection were then interrogated with complementary methods to provide overviews at varying scales and levels of granularity, summarized in Table 1. Data constraints also prevented the use of a single method (e.g., potential product-type estimation) across the entire corpus, so we applied each method where it was most appropriate for data quality and scope (Table 1).

2.2 Structural topic modeling

We used Structural Topic Model (Roberts et al., 2019) to identify prevalent topics within the patent texts. This approach allowed us to test the influence of document metadata, such as the country of patenting or the type of applicant, on topic distribution. Prior to conducting the analysis, we transformed the complete text of the patent data into a document-term matrix format, which displays the frequency of each word for each document. We applied common data-cleaning procedures (Birkholz et al., 2021; Uglanova and Gius, 2020; Chen et al., 2016). Specifically, we removed words that appeared in fewer than 1% of the patents, common stopwords that do not carry significant meaning, and technical or academic terms that simply describe invention outcomes. Finally, we identified and removed the 100 most frequently used words from the corpus. These steps eliminated terms that are either too specific (and thus unrepresentative of broader topics) or too general (and

therefore uninformative), and improved the computational efficiency of the model estimation step.

To estimate the optimal number of topics, we fit models with a range of topic numbers (from 2 to 25) and evaluated them using held-out likelihood, residuals, and semantic coherence scores (Supplementary Figure 1). We selected a 12-topic model because it provided the best balance between statistical metrics and the semantic interpretability of the resulting topics. We then labeled the topics based on their most unique keywords and the titles of the top 30 most representative patents for each topic (function `findThoughts` in `stm` package (Roberts et al., 2019)). Full topic names, titles of most representative patents, and unique keywords are available in Supplementary Table 1.

In some cases, topic labeling was easy to do solely based on the most frequent keywords, e.g., “Bioluminescence” (‘*gfp*’, ‘*fluorescent*’, ‘*enzyme*’, ‘*luciferase*’) and “Biopharmaceuticals” (‘*antibody*’, ‘*human*’, ‘*chain*’, ‘*cancer*’). In other cases, we consulted the titles of the most representative patents. For example, for the “Bio-based chemicals” topic, titles such as ‘Carbon pathway optimized production hosts for the production of isobutanol,’ ‘Fermentative production of alcohols,’ and ‘Glycerol 3-phosphate dehydrogenase for butanol production’ were indicative that these inventions centered around green chemistry as an alternative to petrochemical processes.

We incorporated multiple parameters that could influence topic distribution, including the type of applicant, as well as the inclusion of genetic sequences from deep-sea species. Linear regression was used to assess the significance of these covariates on topic prevalence, with the prevalence of each topic set as a dependent variable and the covariates as independent variables.

2.3 LLM-assisted categorization

Our approach involves training and evaluating a LLM to identify cases where patent applicants frame their invention as a specific type of biotechnological innovation. This method is inspired by a previous study on public value expressions in patents (Pelaez et al., 2024).

Patent abstracts provide a concise, standardized overview of an invention’s purpose, core components, and applications, making

them useful for rapid screening. Claims, in contrast, use precise legal language to convey both explicit facts and implicit technical knowledge about an invention's scope and R&D context (Chen et al., 2016). Together, abstracts enable initial screening, while claims provide the granular detail needed for thorough semantic analysis.

We used a locally installed LLM (DeepSeek-R1-7b) to extract relevant keywords (e.g., “vaccine,” “enzyme,” “transgenic plant”) from patent abstracts and claims. These keywords were mapped to four biotechnological themes (Healthcare, Agriculture, Industrial, and Foundational Biotechnology) using a predefined description of each theme's potential uses for product and method development (Text S1). Prompts included a concise system role (“Acting as a very knowledgeable biotechnology researcher...”), explicit task instructions (“Identify keywords indicative of potential product or method development, and assign the most likely biotechnological theme”), and a clear output format (JSON with fields “biotechnological theme” and “keywords of support”). Where helpful, we used a chain-of-thought style and iteratively refined prompts on a small subset of the data, following recommendations in the literature (Törnberg, 2024; Weber and Reichardt, 2023). To consolidate variations, the 100 most frequent keywords were then mapped to standardized categories (Text S2).

To confirm that the LLM-identified keywords were actually present in the patent text, we used regular-expression checks. This computational technique searches text for specific character sequences, allowing precise, automated verification of whether a word or phrase appears in a document. For this step, we considered only patents previously selected for topic-modeling analysis that also had both claims and abstract sections completed, yielding 1,310 patents.

To ensure robustness, we ran the keyword extraction process three times. A patent was only included for further analysis if a keyword's presence was confirmed in at least two of the three runs. Similarly, the categorization into product types was repeated three times, and we retained only the 1,273 patents that received the same category label in two out of three iterations. This rigorous filtering process led to the exclusion of 37 patents due to inconsistencies in either keyword presence or category prediction.

Finally, we performed a manual validation on a random sample of 200 patents that met these criteria. To ensure that the validation set included all predicted categories, we randomly selected 50 patents for each category. By manually reading the title, abstract, and claims, we assessed the accuracy of the LLM's predictions and found a good agreement with the model's assigned category (Weighted-F1 score = 0.9016; F1 scores for individual categories: 0.9515, Foundational biotechnology = 0.8387; Healthcare biotechnology = 0.9231; Industrial biotechnology = 0.8800). F1 scores were calculated in Python with `classification_report` function imported from `sklearn` library (Pedregosa et al., 2011).

2.4 Product type prediction and potential contribution to sustainable development

Our analysis began by identifying patents where marine-derived enzymes were central to the invention. First, we performed a regular expression search to match enzyme names from the MABPAT

database [originally sourced from UniProt (UniProt Consortium, 2023)] against patent titles, resulting in an initial set of 395 potential matches.

Next, we conducted a manual verification of these candidates to confirm that the contribution of the marine-derived enzyme was substantial to the core claims of the invention. This validation process, which involved a detailed reading of each patent, narrowed the set to 221 confirmed cases. To support this manual assessment, we performed a sequence-similarity search for each confirmed enzyme in the InterPro database (<https://www.ebi.ac.uk/interpro/>). This allowed us to link the enzymes to specific protein families and their associated functional descriptions, providing an independent line of evidence for their role in the invention.

A patent was ultimately included if it met two criteria: (1) the enzyme name or class was explicitly mentioned in the claims, and (2) manual evaluation confirmed that the invention was substantively reliant on that enzyme. For each confirmed patent, we also summarized the invention's potential use in developing specific methods or products.

To evaluate each invention's potential contribution to sustainable development, we used the invention's intended use from the patent text and then contextualized it with scientific literature that explains the relevance of the described products or methods to sustainable development goals. We included these supporting references to link the specific technological innovation to its broader sustainability impact (Supplementary Table 4).

Our use of the SDG framework in this analysis is intentionally indicative rather than evaluative. We use SDG alignment as a reference point for mapping the intended applications of marine genetic resources as expressed in patent texts, rather than as a definitive measure of sustainability outcomes. We do not claim that patenting activity or biotechnological innovation is inherently sustainable, nor do we treat stated alignment with SDG objectives as a sufficient condition for sustainability. Realizing any such contribution in practice requires governance frameworks, equitable benefit-sharing arrangements, and active policy choices that extend well beyond the act of innovation itself.

3 Results

3.1 Significant abundance of deep-sea sequences in enzyme development

By applying structural topic modeling to 1,676 patent texts, we created a 12-topic model for enzyme development. The two most abundant topics relate to biological enzymes: molecular tools that link genes to measurable light signals (i.e., reporter assays) for studying cellular processes (Bioluminescence, 19.31%) and tools for manipulating molecular biology machinery (DNA enzymes, 12.39%) (Figure 1). Inventions targeting crop improvement are next (Crop traits, 12.22%), followed by Biopharmaceuticals (12.19%) and Bio-based chemicals (9.06%). Interestingly, these topics also show different temporal dynamics. Bioluminescence and DNA enzymes peaked in the late 1990s, while Biopharmaceuticals have become the most abundant topic post-2020 (Supplementary Figure 2).

We also tested whether the presence of sequences from deep-sea species was a co-factor affecting topic distribution (see Methods: Topic identification). We found this influence to be significant for four topics: DNA enzymes ($p_{adj_BH} = 4.78e-10$), Bio-based chemicals ($p_{adj_BH} = 3.25e-03$), Protein engineering ($p_{adj_BH} = 8.66e-05$), and Unnatural amino acids ($p_{adj_BH} = 5.29e-04$). Deep sea sequences are thus more likely to get listed in patents associated with these four topics.

While this analysis provides an overview of the semantic landscape, its temporal dynamics, and the influence of deep-sea sequences, we found that some topics were difficult to delineate from one another (e.g., Crop traits vs. Crop engineering; Protein engineering vs. Molecular engineering). This inspired us to use large language model-assisted categorization (see Methods: LLM-assisted categorization for details).

3.2 Sector-specific language and public-private divergence in biotechnology patenting

By applying an LLM-based extraction method to 1,273 patent texts, we extracted keywords indicating the type of biotechnological innovation and linked each patent to predefined biotechnological themes, informed by our topic modeling results (healthcare, industrial, agricultural applications, and foundational biotechnology; see full description in [Supplementary Table 2](#)). We visualized relationships between keywords and themes as a network, which illustrates the degree to which keywords are exclusive to a theme ([Figure 2](#)). From this network, we observed that patents belonging to the healthcare and agriculture themes are more distinct, using exclusive sets of words to frame their applications. For instance, healthcare patents frequently use terms like ‘vaccine,’ ‘antibody,’ and

‘treatment,’ while agriculture patents use ‘transgenic plant,’ ‘yield,’ and ‘nitrogen use efficiency.’ In contrast, patents within the industrial and foundational biotechnology themes share more common language (e.g., ‘enzyme,’ ‘host cell’), making them harder to distinguish from one another based on keywords alone. Only two terms, ‘nucleic acid’ and ‘polypeptide,’ were used across all four themes.

Next, we aggregated the patents by biotechnological theme and found that the foundational, cross-cutting theme was the largest, while the agricultural theme was the smallest ([Figure 3A](#)). Using patent applicant types from MABPAT, we then estimated the relative contribution of public actors (government agencies, universities, and research centers) versus private actors (companies) across these themes ([Figure 3B](#)). We found statistically significant differences in focus between public and private actors: public actors tend to prioritize foundational and healthcare biotechnology in their patent applications, whereas private actors focus more on industrial and agricultural biotechnology ([Supplementary Table 3](#)).

Finally, we evaluated country-level contributions for each actor type and found a relatively similar distribution among the top contributing nations, with one exception: Germany showed a notably higher contribution from private actors ([Figure 3C](#)), driven by a large number of agricultural biotechnology patents filed by BASF, a German multinational company and one of the largest chemical producers in the world.

3.3 Marine enzymes as enablers of sustainability-relevant biotechnological innovation

Although our initial computational analysis provided insight into innovation priorities, it remained difficult to predict the specific

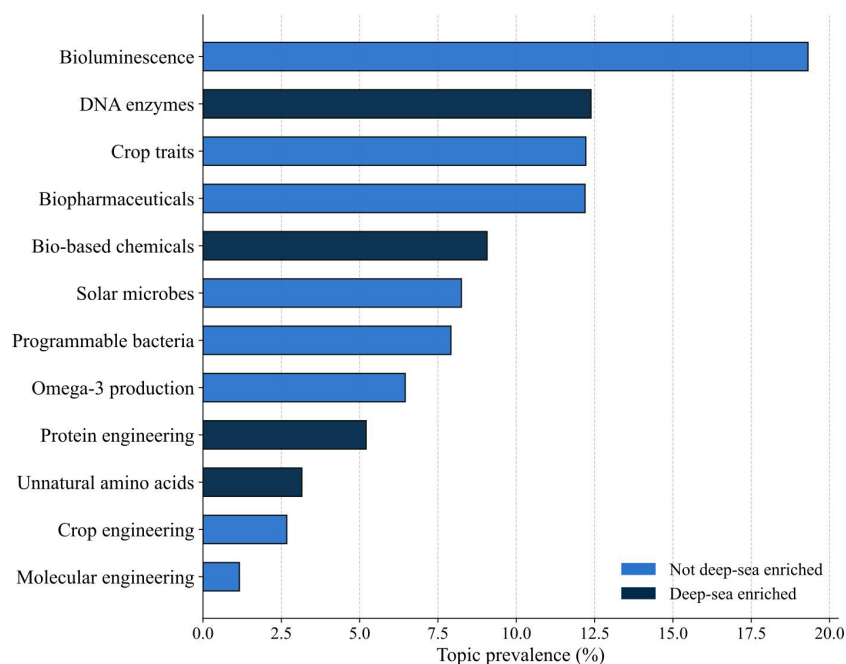


FIGURE 1 Topic prevalence and deep-sea enrichment in marine gene patents. Bars show topic prevalence (%) from a 12-topic structural topic model. Topics are colored by whether they are deep-sea enriched (dark blue) or not deep-sea enriched (blue) based on the estimated effect of deep-sea sequence presence on topic prevalence; dark-blue topics indicate a significant effect.

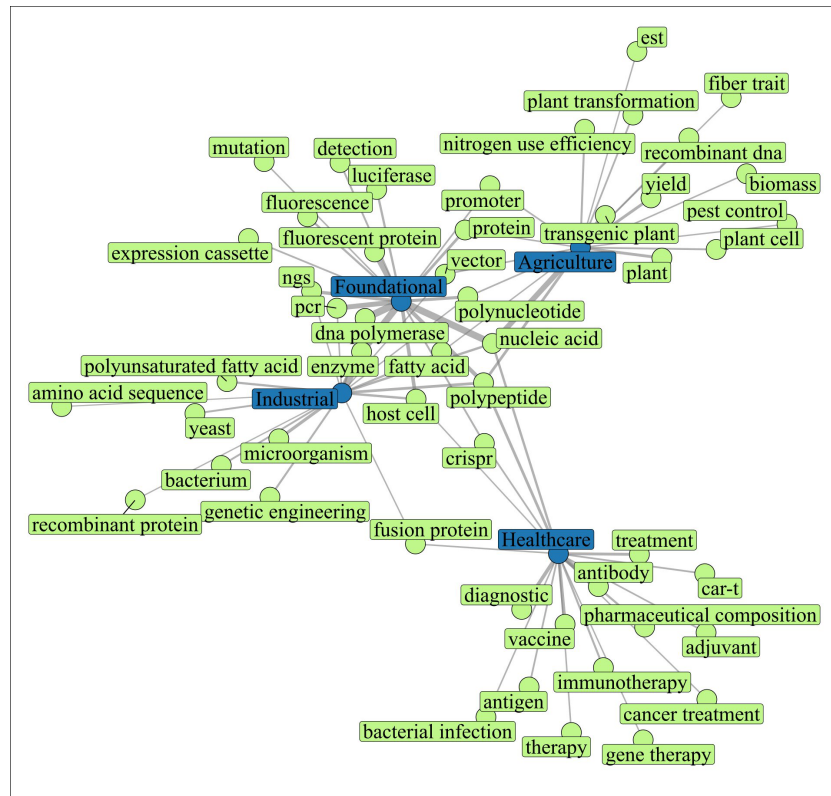


FIGURE 2
 Network of the top 50 biotechnology invention keywords (green) grouped into four emerging biotechnological themes (blue). Edge thickness reflects keyword co-occurrence strength. Network construction: for each theme, the dataset was sorted by keyword frequency and the top 18 keywords were selected, these lists were combined (duplicates merged), a keyword co-occurrence matrix was computed from the merged set, and the resulting co-occurrence weights were used to draw edges between keyword nodes and theme hubs.

products or methods resulting from these inventions using patent text alone. To address this, we leveraged a unique feature of the MABPAT dataset: the names of biological enzymes derived from marine species. We created a subset of patents by selecting only those that mentioned at least one marine enzyme by name within their claims section. The claims section defines the legal scope of protection and explicitly states the elements of the invention that the applicant seeks to secure exclusive rights to. Enzymes can be mentioned elsewhere in a patent (e.g., the background or detailed description) simply as context, prior art, or illustrative examples, without being central to what is protected. Focusing on enzymes mentioned in the claims yields patents in which the marine enzyme is likely essential to the patented method or product. This search yielded 395 patents, which we considered candidates where the functionality of a marine enzyme was likely substantial to the invention.

We then manually reviewed these candidates for additional evidence of biological functionality to confirm their relevance (see Methods: Product type prediction). This validation process resulted in 221 confirmed cases where marine enzymes were primary contributors to the biotechnological innovation (Supplementary Table 4). For 129 of these cases, a manual search of the scientific literature provided further evidence that the innovation could contribute to one or more SDGs.

By examining patent semantics and descriptions of marine enzyme functionality, we found that largest categories of product

types where marine enzymes were primary contributors included Food, feed & nutraceuticals (23.2%), Molecular biotechnology toolbox (18.81%), Green chemistry & environmental technologies (15.36%), and Bioenergy & carbon utilization (9.4%) (Figure 4A). Healthcare-related categories - Therapeutics & treatments (9.72%), Vaccines & production support (6.58%), and Clinical diagnostics (4.08%) - were also abundant. With the exception of Molecular biotechnology toolbox, most of these inventions can directly contribute to sustainable development by providing improved healthcare solutions, enriching crops and food, advancing the use of microorganisms as enzyme factories, or offering bio-based alternatives for converting feedstock into biofuel (Figure 4B; Supplementary Table 4).

The connection between an invention and its aim to contribute to sustainability often seemed explicit and clear (e.g., biofuel development, biodegradable polymers, vaccines). In other cases, the link was less direct, such as with green chemistry principles or food technologies that serve multiple purposes (e.g., nutraceuticals and animal feed). Furthermore, some inventions, particularly those in the Molecular biotechnology toolbox, were not framed in terms of sustainability at all. However, the importance of these foundational tools should not be underestimated, as they underpin a wide range of other biotechnological advances, whether sustainable or not.

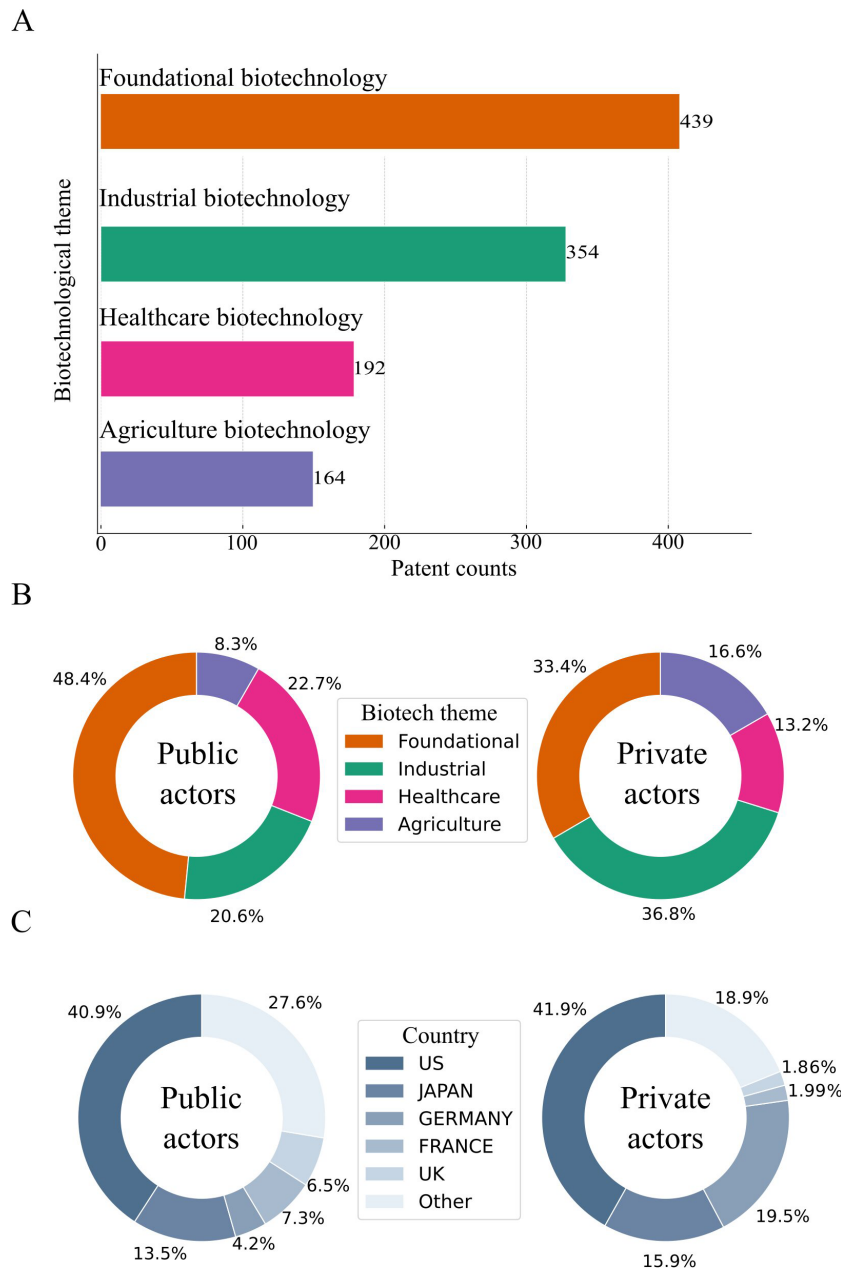
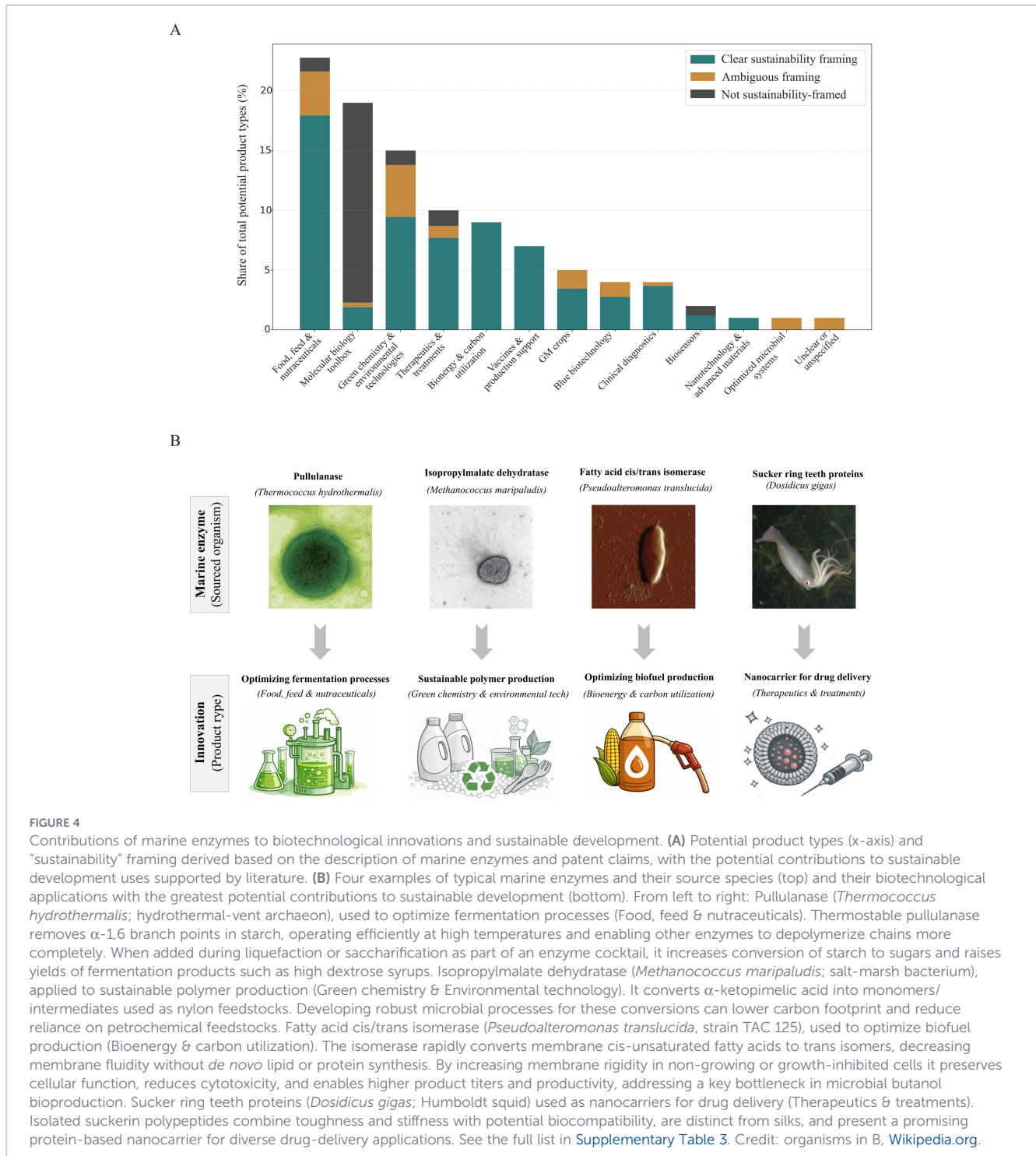


FIGURE 3 Mapping the contribution of marine gene patents to different biotechnology types and actor groups. **(A)** Patent counts aggregated by biotechnological theme. **(B)** Relative contribution of public actors (government agencies, universities, and research centers) versus private actors (companies) across biotechnological themes. **(C)** Country-level contributions for public and private actors.

4 Discussion

Our semantic analysis of the MABPAT patent corpus demonstrates that a systematic examination of patent data provides broad insight into the specific applications of MGR that extends beyond anecdotal examples. By combining topic models with LLM-assisted categorization, we identified recurrent technology themes and their relationships with different applicant types. The generalizability of this methodological approach and its automated features means that it has broad applicability in capturing historical and emerging trends in biotechnology development through the assessment of patent filings and claims. As demonstrated here with specific focus on deep-sea

source organisms and sustainability framings of innovations, the approach can also be adapted to provide insight into topics or themes of particular relevance to specific policy or innovation processes. There is also valuable explanatory potential in providing an evidence-based foundation of the biotechnological sectors in which marine genetic resources appear, and how and by whom they are used in potential products, to policymakers, who frequently contend with normative or selective narratives surrounding the past contributions and future potential of biotechnology. Timely venues for patent-based analytical evidence include the forthcoming first Conference of Parties to the BBNJ Agreement, and ongoing efforts to specify and implement the Cali Fund under the Convention on Biological Diversity.



4.1 Recognizing and supporting public research as an engine of marine biotechnology

Our results underscore that private and public sector actors play distinct and perhaps complementary roles in advancing marine biotechnology. Notably, public research organizations place significant emphasis on patents focused on foundational biotechnology. These are innovations that generate platform technologies, enabling tools, or methods upon which many downstream applications build - for

instance, thermostable polymerases and ligases from thermophilic organisms that underpin PCR and molecular cloning workflows, or luciferase reporter systems cloned from marine zooplankton that serve as universal tools for imaging and assay development (Nikoomanzar et al., 2020; Markova et al., 2004). In this way, our analysis extends previous research on how public science provides benefits for follow-up academic and corporate research, including for private-sector product development that uses MGR (Dunshirn and Zhivkoplis, 2024).

This observation carries clear policy and funding implications. In the United States, which has largely dominated the MGR patent

landscape (Figure 3C), reductions in public research funding (for instance, cuts to biomedical research budgets - <https://www.biopharmadive.com/news/biotech-nih-funding-research-trump-cuts-impact/740802/> - will likely have a global dampening effect on advancements in biotechnology and, in turn, the spillover effects to the private sector that benefits from those discoveries (Becker et al., 2023; Bloom et al., 2013). Sustained public investment - both in targeted biotechnological development and in marine biodiversity and observation programs - therefore remains crucial for maintaining the long-term pipeline of innovations that can advance healthcare and sustainable development.

4.2 Transparency gaps: linking collection, sequence, funding and patents

Uneven norms of voluntary disclosure related to MGR provenance, or the associated enabling funding represent a major challenge in accurately attributing the technological value of MGR (Blasiak et al., 2019). The WIPO Treaty on Intellectual Property, Genetic Resources and Associated Traditional Knowledge (World Intellectual Property Organization (WIPO), 2024a) requires patent applicants to disclose the country of origin (or, if unknown/inapplicable, the source) for genetic resources and to identify the Indigenous people or local community (or, if unknown/inapplicable, the source) for any traditional knowledge. This should allow more direct mapping back to sampling events in practice by linking patented DNA sequences to INSDC records under WIPO Standard ST.26 (World Intellectual Property Organization (WIPO), 2023) and, given INSDC's strengthened spatiotemporal metadata requirements, should make the location of the sampling underlying patented sequences more accessible. These mandatory disclosure expectations, together with diverse national regimes that already condition patent rights on genetic-data provenance (for example, Brazil, China, and some European countries (World Intellectual Property Organization (WIPO), 2024b)) create new legal and administrative pressures to tighten links between sequence records, accession identifiers, expedition metadata and patent claims. Patent filings, however, still rarely cite accession identifiers (e.g., BioProject/BioSample/GenBank accession numbers), and funding acknowledgements in patents are often absent or ambiguous, complicating attribution of public versus private investment across the biodiversity pipeline. Initiatives such as BBNJ Standardized Batch Identifier (BBNJ Identifier) and proposed information systems under the BBNJ Agreement highlight the need to balance stronger provenance and funding traceability (Lawson et al., 2024) with the administrative burden on researchers so as not to hamper biodiversity research (Prathapan et al., 2018). To achieve this balance, the BBNJ Identifier should be designed in line with scientific good practice and built around core features such as persistence, discoverability, resolvability, and authority, so that it can be recognized across databases, integrated into existing systems, and used to trace later materials or data back to the original sampled material (Langlet et al., 2025). Beyond BBNJ Agreement implementation, science-policy cooperation should also prioritize developing tools to trace grant support - building where possible on existing infrastructure such as ORCID, which links researchers to funders - and to track use of public biological databases, enabling reconstruction of the

full innovation chain involving genetic material and ensuring equitable benefit-sharing. Although this would require sustained coordination across research communities, data infrastructures, and governance frameworks, emerging initiatives such as the Global Biodata Coalition, which maps key biological data resources - <https://globalbiodata.org/what-we-do/global-inventory/global-inventory-list/> - provide a useful starting point for developing a more automated and transparent global system for tracking genetic data flows.

4.3 Distinguishing primary from secondary contributions of marine sequences

To address the criticism that the functional role of MGR in patented inventions is often unclear, we provide a differentiated analysis of primary contributions (i.e., marine sequences that are central to the claimed invention, typically appearing in the claims or detailed mechanistic descriptions) and secondary contributions (i.e., marine sequences listed without a specified functional mechanism, for instance as part of background information or comparative data). This analysis was done by focusing on the claims section, which legally defines the scope of an invention and where mentions of specific marine enzymes are a stronger indicator that the sequences materially contribute to the protected technology. The subset of patents identified here provides the most direct examples of MGR driving technological development. However, our search strategy based on enzyme name detection in patent claims likely undercounts primary contributions for cases in which inventors describe functional roles without naming the enzymes explicitly. This gap could be closed manually to some extent, or become less of a challenge if standardized usage of Enzyme Commission numbers in patent filings becomes a norm in the future, particularly if filings reference enzyme functional data in public databases (i.e. https://www.brenda-enzymes.org/all_enzymes.php). For large-scale analysis, future semantic models that are capable of parsing nested semantic relationships within patent text could further automate and refine this identification, offering a more comprehensive understanding of how genetic resources contribute to sustainable technologies. This distinction is valuable for better estimating the value of MGR to innovation for benefit-sharing because accurately separating primary from secondary contributions clarifies which sequences materially drive technological advances and therefore warrant prioritization in benefit-sharing calculations, particularly given the uncertainty over what group of digital sequence information will be operationalized under the BBNJ Agreement (Convention on Biological Diversity, Executive Secretary, 2020a; Convention on Biological Diversity, Executive Secretary, 2020b; Deep-Ocean Stewardship Initiative (DOSI), 2020).

4.4 Sustainability implications of primary-use inventions

Marine sequences we classified as primary contributors to patented innovations were sometimes associated with technologies aimed at delivering on sustainability objectives such as

bioremediation enzymes, salt-tolerant catalysts for low-waste chemical processes, and enzymes that enable more efficient biomass conversion. Such innovations can contribute to reduced environmental footprints and optimized industrial processes. Thermostable alginate lyases derived from marine microbes (e.g., patent application JP2016192903A (<https://patents.google.com/patent/JP2016192903A/en>), for instance, are being engineered to efficiently break down seaweed into fermentable sugars, which has the potential to use macroalgae as a viable feedstock for producing biofuels and chemicals, reducing pressure on terrestrial agriculture and fossil fuels. In another example, the genes for key enzymes from the horseshoe crab's blood-clotting system have been cloned and expressed recombinantly (e.g., patent application WO2018159771A1 (<https://patents.google.com/patent/WO2018159771A1/en>), creating an animal-free method for detecting bacterial endotoxins in pharmaceuticals. This innovation directly replaces the harvesting of a vulnerable marine species, which has been central to biomedical safety testing for decades, and could address ethical concerns associated with "bleeding" of horseshoe crabs for medical purposes (Krusberg et al., 2024). Nevertheless, not all primary-use cases yield clear sustainability benefits; many are oriented toward incremental performance improvements in existing industrial processes or new product formulations whose net social and environmental impacts require lifecycle assessment. Likewise, some improvements in efficiency have a less obvious contribution to sustainability (e.g. using biotechnology to enhance crude oil recovery), as they can be argued to perpetuate inherently unsustainable practices, favoring lock-in and diverting attention and resources from cleaner alternatives (Sen, 2008). For example, cold-active enzymes can enable lower-temperature processing that reduces energy use and associated emissions, but their wider industrial uptake can still be constrained by cost and scale-up challenges (Kuddus et al., 2024).

Arriving at a clear answer to whether marine biotechnology is part of a "blue economy" is therefore challenging, as it is a tool rather than an outcome in and of itself. Marine-derived technologies are often pursued because marine organisms offer unique enzymes and genes adapted to extreme conditions, which can enable novel or more robust processes; and because access and development are costly, research tends to concentrate on applications with higher expected value and impact. However, these performance gains do not necessarily translate into more sustainable production cycles, and whether marine biotechnology innovations fall under the blue economy depends not only on what the tool is used for, but also how it is used - including alignment with CBD/BBNJ requirements and responsible research and innovation norms. Greater clarity about how some marine biotechnology innovations contribute to sustainability can come from coupling semantic patent analysis with additional evidence, such as product lifecycle data and substitution effects. This could help to determine whether a marine-derived technology displaces more harmful alternatives or simply creates new demand. Innovations often have both clear benefits and unintended consequences that complicate their net impact on sustainability. The case of semaglutide is illustrative: its profound benefits for patients with type 2 diabetes represent a clear societal good, yet its widespread use for weight loss

raises more ambiguous questions about systemic effects on appetite regulation, taste perception, and metabolic pathways (Brindisi et al., 2019). Some MGR-based innovations share these complex characteristics. For instance, introducing genes from algae or cyanobacteria into crops to boost photosynthetic efficiency (e.g., patent application JP2012110330A (<https://patents.google.com/patent/JP2012110330A/en>)) or cloning archaeal genes for nitrogen-transfer enzymes to improve nitrogen utilization (e.g., patent application WO2014135535A1 (<https://patents.google.com/patent/WO2014135535A1/en>)) directly addresses food security, but could also further entrench large-scale agricultural monocultures at the expense of more resilient, biodiverse farming systems. Our curated knowledge base of primary-use cases provides a starting point for such integrative assessments, enabling a more holistic evaluation of both intended benefits and potential downstream trade-offs.

Our mapping of the MGR patent landscape underscores both the significant potential and the complex challenges of reconstructing how marine biodiversity is or can be used for sustainable development. By clarifying the applications, actors, and mechanisms of innovation, this study provides an evidence base to guide the implementation of the BBNJ Agreement and support science-informed policies that incentivize the conservation and sustainable use of marine genetic diversity.

Data availability statement

The datasets analyzed for this study can be found in the figshare.com - <https://doi.org/10.6084/m9.figshare.25289404>.

Author contributions

EZ: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. PD: Data curation, Writing – review & editing, Validation. JJ: Visualization, Writing – review & editing, Validation. MJ: Validation, Writing – review & editing. RB: Funding acquisition, Writing – review & editing, Supervision, Validation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. Erik Zhivkoplis and Robert Blasiak are funded by FORMAS, project number 2020-01048. Paul Dunshirn is funded by the European Research Council, grant agreement Nr. 101124903 - TwinPolitics. Jean-Baptiste Jouffray is funded by the Knut and Alice Wallenberg Foundation (2021.0343). Marcel Jaspars is supported by UK Research and Innovation (UKRI) under the UK Government's Horizon Europe funding guarantee grant No. IFS 10057167 (University of

Aberdeen) for EU project BlueRemediomics, which is funded by the European Union under the Horizon Europe Programme, Grant Agreement No. 101082304.

Acknowledgments

We thank Juan Rocha, Ingo Fetzer, Stefan Daume, and other participants of the SRC Coding Club. Conversations about the risks and potential of using large language models inspired the initial methodological approach for this study.

Conflict of interest

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer CL declared a past co-authorship with the author MJ to the handling editor.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. A locally installed LLM (DeepSeek-R1-7b) was used to extract relevant keywords from patent abstracts

References

- Arrieta, J. M. (2010). What lies underneath: Conserving the oceans' genetic resources. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18318–18324. doi: 10.1073/pnas.0911897107
- Becker, B., Roper, S., and Vanino, E. (2023). Assessing innovation spillovers from publicly funded R&D and innovation support: Evidence from the UK. *Technovation* 128, 102860. doi: 10.1016/j.technovation.2023.102860
- Bennett, N. J., Cisneros-Montemayor, A. M., Blythe, J., Silver, J. J., Singh, G., Andrews, N., et al. (2019). Towards a sustainable and equitable blue economy. *Nat. Sustain.* 2, 991–993. doi: 10.1038/s41893-019-0404-1
- Birkholz, J., Günther, J., and Shkolnykova, M. (2021). "Using topic modeling in innovation studies: The case of a small innovation system under conditions of pandemic related change," in *Bremen papers on economics & Innovation*. (Bremen: University of Bremen).
- Blasiak, R., Jouffray, J.-B., Amon, D. J., Claudet, J., Dunshirn, P., Jørgensen, P. S., et al. (2023). Making marine biotechnology work for people and nature. *Nat. Ecol. Evol.* 7, 482–485. doi: 10.1038/s41559-022-01976-9
- Blasiak, R., Jouffray, J.-B., Wabnitz, C. C. C., and Österblom, H. (2019). Scientists should disclose origin in marine gene patents. *Trends Ecol. Evol.* 34, 392–395. doi: 10.1016/j.tree.2019.02.004
- Blicharska, M., Smithers, R. J., Mikusiński, G., Rönnbäck, P., Harrison, P. A., Nilsson, M., et al. (2019). Biodiversity's contributions to sustainable development. *Nat. Sustain.* 2, 1083–1093. doi: 10.1038/s41893-019-0417-9
- Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica* 81, 1347–1393. doi: 10.3386/w13060
- Brindisi, M.-C., Brondel, L., Meillon, S., Barthet, S., Grall, S., Fenech, C., et al. (2019). Proof of concept: Effect of GLP-1 agonist on food hedonic responses and taste sensitivity in poor controlled type 2 diabetic patients. *Diabetes Metab. Syndr.* 13, 2489–2494. doi: 10.1016/j.dsx.2019.06.021
- Chen, H., Zhang, Y., and Zhu, D. (2016). "Identifying technological topic changes in patent claims using topic modeling," in *Innovation, technology and knowledge management* (Cham ZG: Springer), 187–209.
- Chiarolla, C., Kock, M. A., Oldham, P., Halewood, M., Noriega, I. L., Kindness, J., et al. (2025). *Patent landscape report on inventions based on Plant Genetic Resources for Food and Agriculture (PGRFA) and related Digital Sequence Information Genetic Sequence Data (DSI/GSD): Executive summary (ENG/FR/SP)*. (FAO Plant Treaty). doi: 10.4337/9781783478330.00023
- Cisneros-Montemayor, A. M., Moreno-Báez, M., Reygondeau, G., Cheung, W. W. L., Crosman, K. M., González-Espinosa, P. C., et al. (2021). Enabling conditions for an equitable and sustainable blue economy. *Nature* 591, 396–401. doi: 10.1038/s41586-021-03327-3
- Convention on Biological Diversity, Executive Secretary (2020a). *Digital sequence information on genetic resources: Concept, scope and current use* (Montreal: Secretariat of the Convention on Biological Diversity). CBD/DSI/AHTEG/2020/1/3.
- Convention on Biological Diversity, Executive Secretary (2020b). *Report of the ad hoc technical expert group on digital sequence information on genetic resources* (Montreal: Secretariat of the Convention on Biological Diversity). CBD/DSI/AHTEG/2020/1/7.
- Deep-Ocean Stewardship Initiative (DOSI) (2020). *Digital Sequence Information—Clarifying Concepts. Policy brief*. Available online at: <https://dosi-project.org/070-DSI-Policy-brief-V4-WEB.pdf> (Accessed November 3, 2025).
- Deep-Ocean Stewardship Initiative (DOSI) (2025). *Value of MGR policy brief*. Available online at: <https://www.dosi-project.org/wp-content/uploads/Value-of-MGR-Policy-Brief.pdf> (Accessed November 3, 2025).
- Donia, M., and Hamann, M. T. (2003). Marine natural products and their potential applications as anti-infective agents. *Lancet Infect. Dis.* 3, 338. doi: 10.1016/s1473-3099(03)00655-8
- Dunshirn, P., and Zhivkoplias, E. (2024). Conducting marine genetic research for whom? Mapping knowledge flows from science to patents. *NPJ Ocean Sustain.* 3, 50. doi: 10.21203/rs.3.rs-3545616/v1
- Francesch, A. M., Glaser, K. B., Jaspars, M., Jiménez, C., Luesch, H., Mayer, A. M. S., et al. (2024). Pharmaceuticals from marine sources: past, present and future. *Pharm. J.* 313 (7987). doi: 10.1211/PJ.2024.1.324083

and claims. These keywords were mapped to four biotechnological themes using a predefined description of each theme's potential uses for product and method development, as described in Methods. Gemini 2.5 Flash was used for grammar and syntax checks of all writing contributions. GitHub Copilot assisted with script programming to improve the visual appeal of the plots.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2026.1833479/full#supplementary-material>

- Heberling, J. M. (2021). Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2018093118. doi: 10.1073/pnas.2018093118
- Kolterman, O. G., Buse, J. B., Fineman, M. S., Gaines, E., Heintz, S., Bicsak, T. A., et al. (2003). Synthetic Exendin-4 (Exenatide) significantly reduces postprandial and fasting plasma glucose in subjects with type 2 diabetes. *J. Clin. Endocrinol. Metab.* 88, 3082–3089. doi: 10.1210/jc.2002-021545
- Krusberg, T., Schildt, L., Jouffray, J.-B., Zhivkoplías, E., and Blasiak, R. (2024). A review of marine genetic resource valuations. *NPJ Ocean Sustain.* 3, 46. doi: 10.1038/s44183-024-00081-7
- Kuddus, M., Roohi, Bano, N., Sheik, G. B., Joseph, B., Hamid, B., et al. (2024). Cold-active microbial enzymes and their biotechnological applications. *Microb. Biotechnol.* 17, e14467. doi: 10.4172/2168-9652.1000e132
- Langlet, A., Dunshirn, P., Jaspars, M., Humphries, F., and Kachelriess, D. (2025). “Monitoring and transparency aspects of MGR-utilization under the BBNJ agreement,” in *Decoding marine genetic resource governance under the BBNJ agreement. Sustainable development goals series*. Ed. F. Humphries (Springer, Cham). doi: 10.1007/978-3-031-72100-7_7
- Lawson, C., Humphries, F., Jaspars, M., and Rabone, M. (2024). *Data management and the “BBNJ standardized batch identifier” under the BBNJ Agreement*. Available online at: <https://papers.ssrn.com/abstract=4897429> (Accessed November 3, 2025).
- Markova, S. V., Golz, S., Frank, L. A., Kalthof, B., and Vysotski, E. S. (2004). Cloning and expression of cDNA for a luciferase from the marine copepod *Metridia longa*. *J. Biol. Chem.* 279, 3212–3217. doi: 10.1074/jbc.m309639200
- Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi: 10.1021/acs.jnatprod.9b01285
- Nikoomanzar, A., Chim, N., Yik, E. J., and Chaput, J. C. (2020). Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* 53, e8. doi: 10.1017/s0033583520000050
- Oldham, P., Hall, S., Barnes, C., Hoffman, I., and Boettcher, P. (2014). “Patent landscape report on animal genetic resources,” in *Patent Landscape Reports Project, in cooperation with the Food and Agriculture Organization of the United Nations*. (Geneva: WIPO) Available online at: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_947_3.pdf (Accessed November 3, 2025).
- Oldham, P., Kindness, J., Davidson, E., Westmoreland, A., Vanagt, T., and Jaspars, M. (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* (Luxembourg: Publications Office of the European Union). doi: 10.2926/7945806
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.48550/arxiv.1201.0490
- Pelaez, S., Verma, G., Ribeiro, B., and Shapira, P. (2024). Large-scale text analysis using generative language models: A case study in discovering public value expressions in AI patents. *Quant. Sci. Stud.* 5, 153–169. doi: 10.1162/qss_a_00285
- Prathapan, K. D., Pethiyagoda, R., Bawa, K. S., Raven, P. H., Rajan, P. D., et al. (2018). When the cure kills—CBD limits biodiversity research. *Science* 360, 1405–1406. doi: 10.1126/science.aat9844
- Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., and Gabrielli, A. (2019). Unfolding the innovation system for the development of countries: coevolution of science, technology and production. *Sci. Rep.* 9, 16440. doi: 10.1038/s41598-019-52767-5
- Roberts, M. E., Stewart, B. M., and Tingley, D. (2019). stm: An R package for structural topic models. *J. Stat. Softw.* 91, 1–40. doi: 10.18637/jss.v091.i02
- Ruiz Muller, M., and Lapeña, I. (2007). *A moving target: genetic resources and options for tracking and monitoring their international flows* (Gland: IUCN).
- Sen, R. (2008). Biotechnology in petroleum recovery: The microbial EOR. *Prog. Energy Combust. Sci.* 34, 714–724. doi: 10.1016/j.pecs.2008.05.001
- Sigwart, J. D., Blasiak, R., Jaspars, M., Jouffray, J.-B., and Tasdemir, D. (2021). Unlocking the potential of marine biodiversity. *Nat. Prod. Rep.* 38, 1235–1242. doi: 10.1039/d0np00067a
- Straccamore, M., Bruno, M., Monechi, B., and Loreto, V. (2023). Urban economic fitness and complexity from patent data. *Sci. Rep.* 13, 3655. doi: 10.1038/s41598-023-30649-1
- Törnberg, P. (2024). *Best practices for text annotation with large language models*. Available online at: <http://arxiv.org/abs/2402.05129> (Accessed November 3, 2025).
- Uglanova, I., and Gius, E. (2020). *The order of things: A study on topic modelling of literary texts*. Available online at: <https://ceur-ws.org/Vol-2723/long7.pdf> (Accessed November 3, 2025).
- UniProt Consortium (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531. doi: 10.1093/nar/gkw1099
- Weber, M., and Reichardt, M. (2023). *Evaluation is all you need: Prompting generative large language models for annotation tasks in the social sciences*. Available online at: <http://arxiv.org/abs/2401.00284> (Accessed November 3, 2025).
- World Intellectual Property Organization (WIPO) (2023). *WIPO Standard ST.26 (Sequence listings)*. Available online at: https://www.wipo.int/documents/d/standards/docs-en-03-26-01_v1_6.pdf (Accessed November 3, 2025).
- World Intellectual Property Organization (WIPO) (2024a). *WIPO treaty on intellectual property, genetic resources and associated traditional knowledge*. Available online at: <https://www.wipo.int/en/web/traditional-knowledge/wipo-treaty-on-ip-gr-and-associated-tk> (Accessed November 3, 2025).
- World Intellectual Property Organization (WIPO) (2024b). *Genetic resources disclosure document (as cited in manuscript)*. Available online at: https://www.wipo.int/documents/d/traditional-knowledge/docs-en-genetic_resources_disclosure.pdf (Accessed November 3, 2025).
- Zhivkoplías, E., Jouffray, J.-B., Dunshirn, P., Pranindita, A., and Blasiak, R. (2024). Growing prominence of deep-sea life in marine bioprospecting. *Nat. Sustain.* 7, 1027–1037. doi: 10.1038/s41893-024-01392-w

Glossary

Marine biotechnology	As used in this paper, the term refers to the application of scientific and engineering principles to the processing or production of materials using marine biological agents, including marine genetic resources, enzymes, and associated molecular tools, for use in healthcare, agriculture, industrial processes, and foundational research. This usage is broader than the term “marine technology” as defined under the BBNJ Agreement, which refers primarily to technology transfer and capacity-building tools for ocean observation and data collection. It is also distinct from “biotechnology” as used in the BBNJ Agreement, which refers specifically to any technological application that uses biological systems, living organisms, or their derivatives to make or modify products or processes for specific use. The authors use “marine biotechnology” in a scientific and applied research context, encompassing the full spectrum of innovation pathways in which marine genetic material serves as an input, and do not intend the term to carry the specific legal meaning assigned to either “biotechnology” or “marine technology” under the BBNJ Agreement.	Sustainable development / Sustainable use	with the inclusion criteria of the World Register of Deep-Sea Species. It is important to note that this is a biological and oceanographic definition, not a legal one. Deep-sea areas may fall either within or beyond national jurisdiction. Deep-sea areas within national jurisdiction are subject to the CBD framework, while those in areas beyond national jurisdiction fall under the governance of the BBNJ Agreement. The paper does not uniformly attribute deep-sea species to either framework, and readers should interpret references to deep-sea genetic material with this jurisdictional distinction in mind. The paper uses “sustainable development” as a multidimensional concept encompassing environmental, social, and economic dimensions, as reflected in the SDGs. Where the paper engages specifically with international governance frameworks, the more legally precise concept of “sustainable use” under Article 1(13) of the BBNJ Agreement is applicable - defined as the use of components of biological diversity in a way and at a rate that does not lead to a long-term decline of biological diversity, thereby maintaining its potential to meet the needs and aspirations of present and future generations. SDG alignment as expressed in patent texts should be read as an indicative mapping of intended use rather than a definitive sustainability assessment, as set out in Methods (Section 2.4).
Deep-sea species	Following the definition used in the MABPAT database and consistent with Zhivkoplis et al. (2024) , deep-sea species in this paper are defined as those for which biological material was sampled at depths greater than 500 meters, in accordance		