

Marine Genetic Resources in Patent Data

Paul Oldham, Stephen Hall & Colin Barnes

United Nations University & One World
Analytics

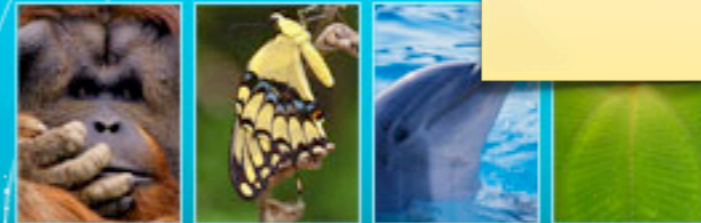
Image: Available GBIF coordinates for 3,432 marine species appearing in patent data

The Problem of Scale

- Patents provide a window into human innovations involving biodiversity but present challenges of scale;
- An estimated +1.2 million accepted species and 8.7 million predicted species (+/-1.3 million) (Mora 2011).
219,714 marine accepted species in WoRMS;
- Over 60 million patent documents in multiple languages;
- Digital data in the form of taxonomic databases (GBIF) and electronic patent data provide a means to map and monitor activity involving marine biodiversity;



```
<?xml version="1.0"
<response xmlns="http:
<header>
<source accesspoint="http:
<software name="TapirLink"
```



... free and open access to biodiversity data

GBIF is the key international data depository for taxonomic data and linked to the CoL and EOL

Search
species/country/dataset

Search



Welcome to the GBIF Data Portal

Access 396,026,747 data records (345,561,101 with coordinates) shared via the GBIF network. To learn how to use this site, please see [About](#). To tune this site for smaller displays, see [Settings](#).
Version 1.2.6 - [click here to see what is new!](#)



Explore Species

Find data for a species or other group of organisms.

Species

Information on species and other groups of plants, animals, fungi and micro-organisms, including species occurrence records, as well as classifications and scientific and common names.

Example species:

[Puma concolor \(Linnaeus, 1771\)](#)



Explore Countries

Find data on the species recorded in a particular country, territory or island.

Countries

Information on the species recorded in each country, including records shared by publishers from throughout the GBIF network.

See data for:

[United Kingdom](#)



Explore Datasets

Find data from a data publisher, dataset or data network.

Datasets

Information on the data publishers, datasets and data networks that share data through GBIF, including summary information on 10004 datasets from 464 data publishers.

Latest dataset added:

[Natural History Museum Rotterdam \(NL\) - Mollusca collection](#)



Data Network: **Ocean Biogeographic Information System**

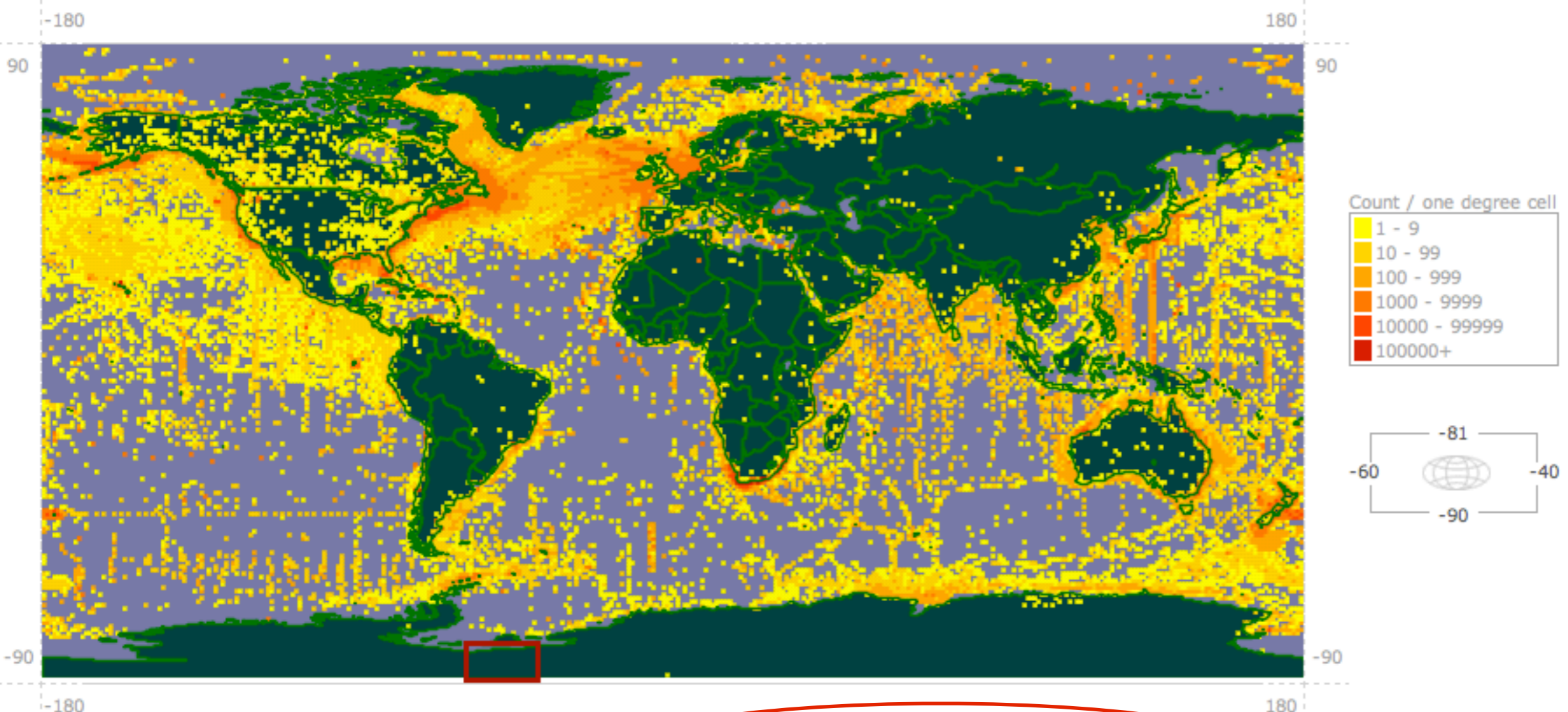
Marine data from OBIS. Provides access to geographical occurrences

Actions for Ocean Biogeographic Information System

Explore: [Occurrences](#)

Download: [One-degree cell density overlay for Google Earth](#) [Placemarks for Google Earth \(limit 10,000\)](#)

Occurrence overview



Beta: GBIF Open Geospatial Consortium services

This map only shows records with coordinates (**11,764,053** records from a total of **11,849,451** records).

Disclaimer: Maps depict density of data registered within the GBIF network index and not necessarily true species occurrence density gradients. It includes records from all datasets shared by this data publisher.

ABSPAT INDEX

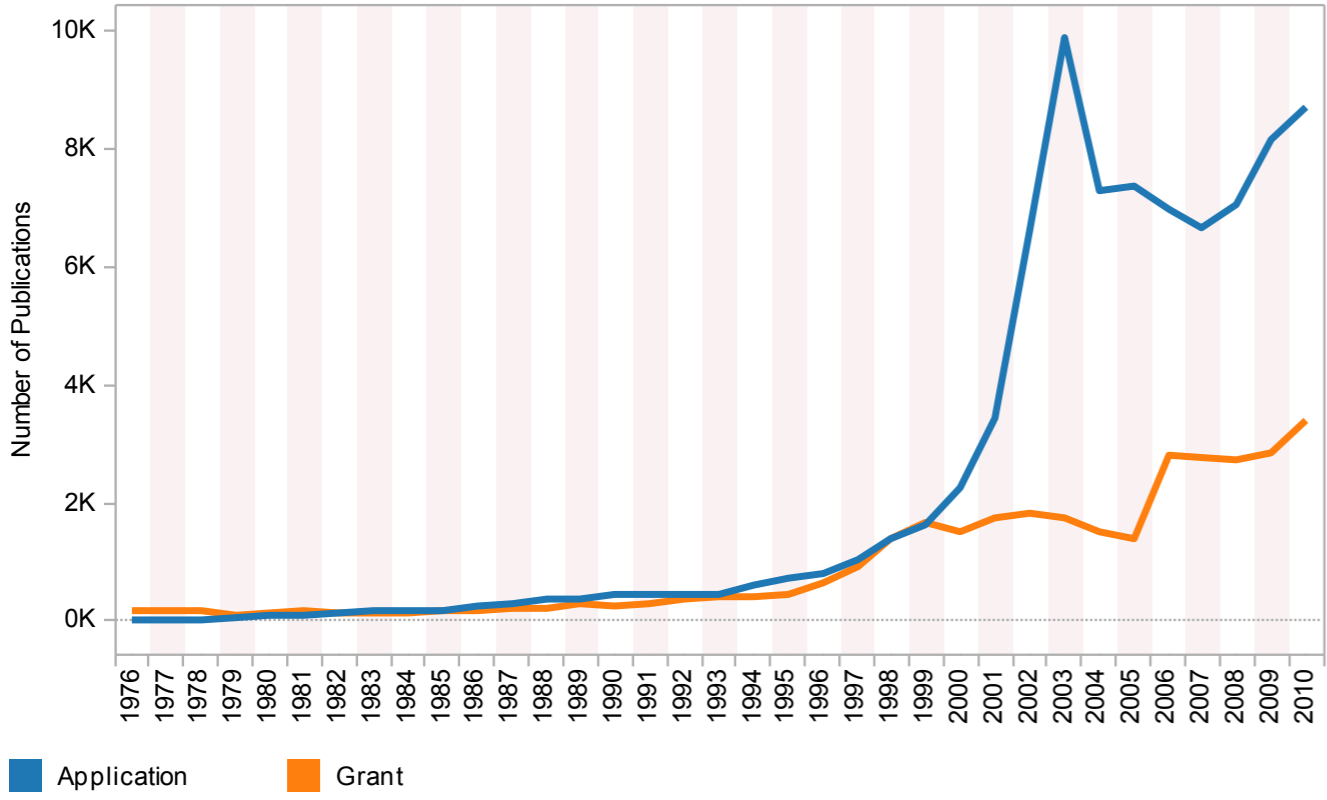
candida utilis:WO1979000421A1//19790712:description:4346,436
candida tropicalis:WO1979000421A1//19790712:description:440
saccharomyces carlsbergensis:WO1979000421A1//19790712:de
saccharomyces uvarum:WO1979000421A1//19790712:descriptio
saccharomyces carlsbergensis:WO1979000421A1//19790712:de
saccharomyces uvarum:WO1979000421A1//19790712:descriptio
saccharomyces cerevisiae:WO1979000421A1//19790712:descrip
saccharomyces cerevisiae:WO1979000421A1//19790712:descrip
saccharomyces carlsbergensis:WO1979000421A1//19790712:description:28313,28341:
saccharomyces uvarum:WO1979000421A1//19790712:description:28358,28378:
torulopsis utilis:WO1979000421A1//19790712:description:4367,4384:
alopecurus myosuroides:WO1979000427A1//19790712:description:12620,12642:
avena fatua:WO1979000427A1//19790712:description:12648,12659:
agropyron repens:WO1979000427A1//19790712:description:12665,12681:
setaria viridis:WO1979000427A1//19790712:description:12599,12614:
streptococcus mutans:WO1979000455A1//19790726:description:402,422:
escherichia coli:WO1979000467A1//19790726:description:15027,15043:
escherichia coli:WO1979000467A1//19790726:description:22144,22160:
escherichia coli:WO1979000467A1//19790726:description:22497,22513:
escherichia coli:WO1979000467A1//19790726:description:23300,23316:
staphylococcus aureus:WO1979000256A1//19790517:description:29,50:
staphylococcus aureus:WO1979000256A1//19790517:description:891,912:
staphylococcus aureus:WO1979000256A1//19790517:description:5657,5678:
staphylococcus aureus:WO1979000256A1//19790517:description:6008,6029:
staphylococcus aureus:WO1979000256A1//19790517:description:7957,7978:
streptococcus pyogenes:WO1979000256A1//19790517:description:7982,8004:
staphylococcus aureus:WO1979000256A1//19790517:description:8685,8706:

Patent data can be mined using parallel computing to identify 6 million species names from the Global Name Index

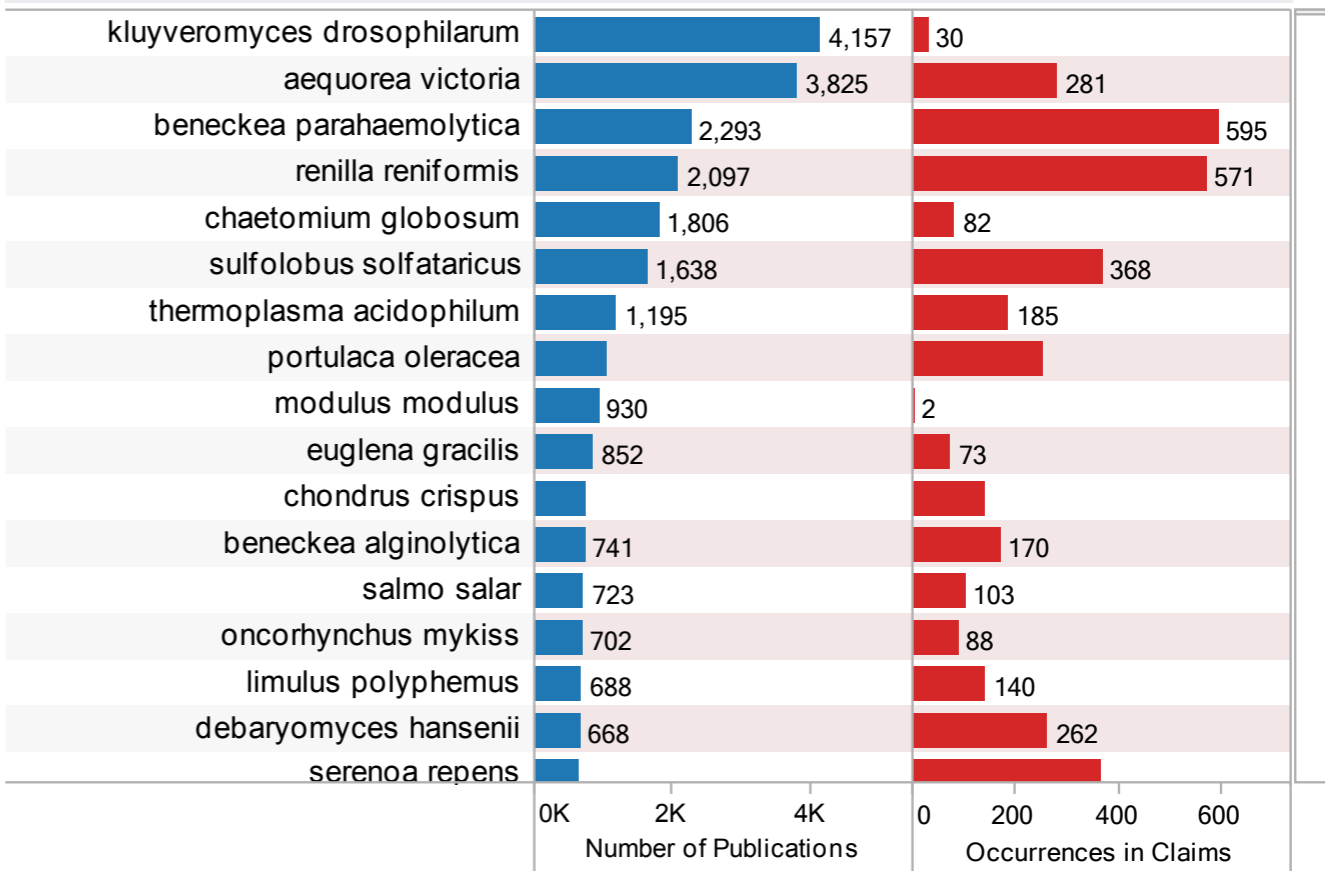
ABSPAT = Access and Benefit Sharing Patent Index. An offline index of species in patent documents.

Trends

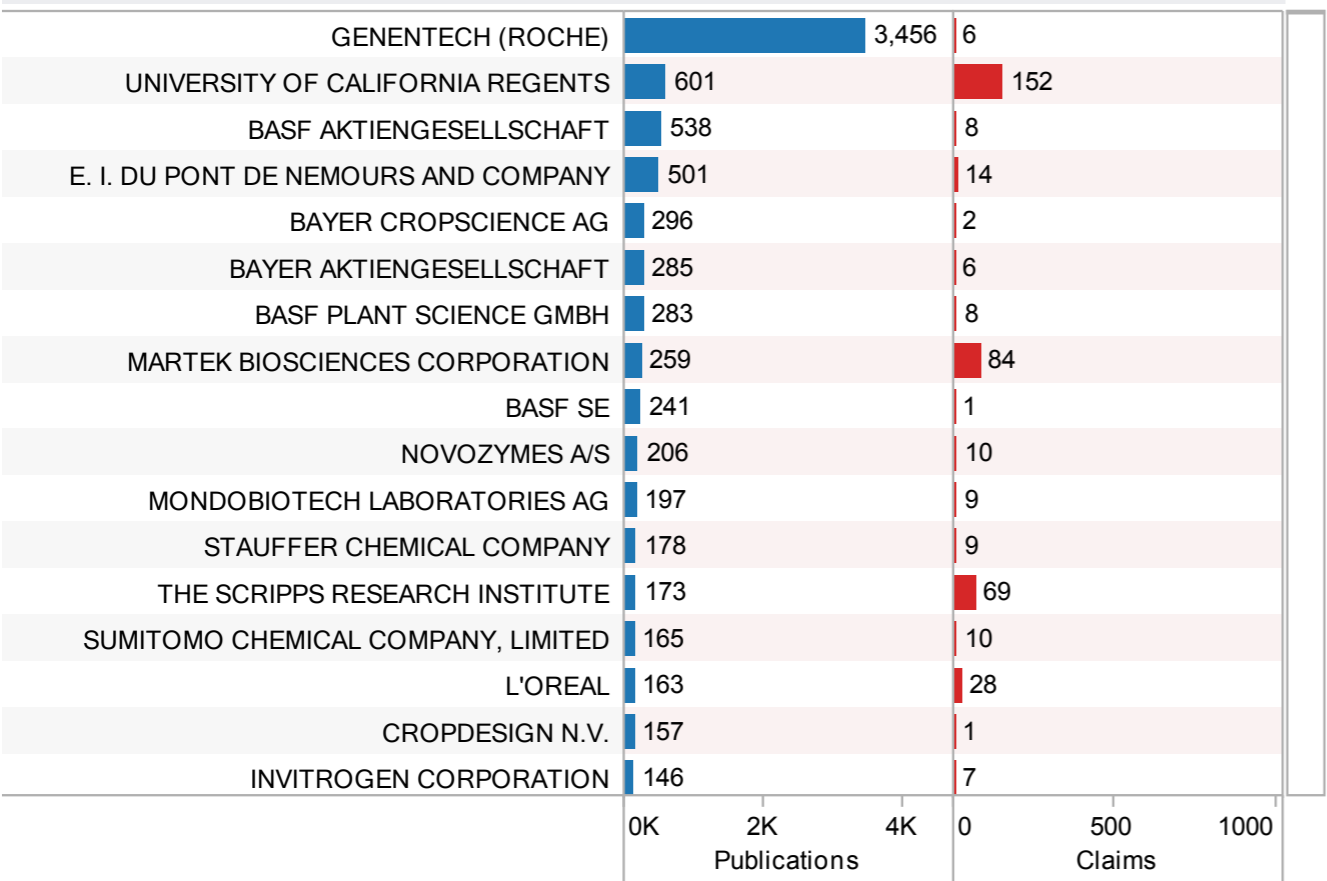
Publication Year



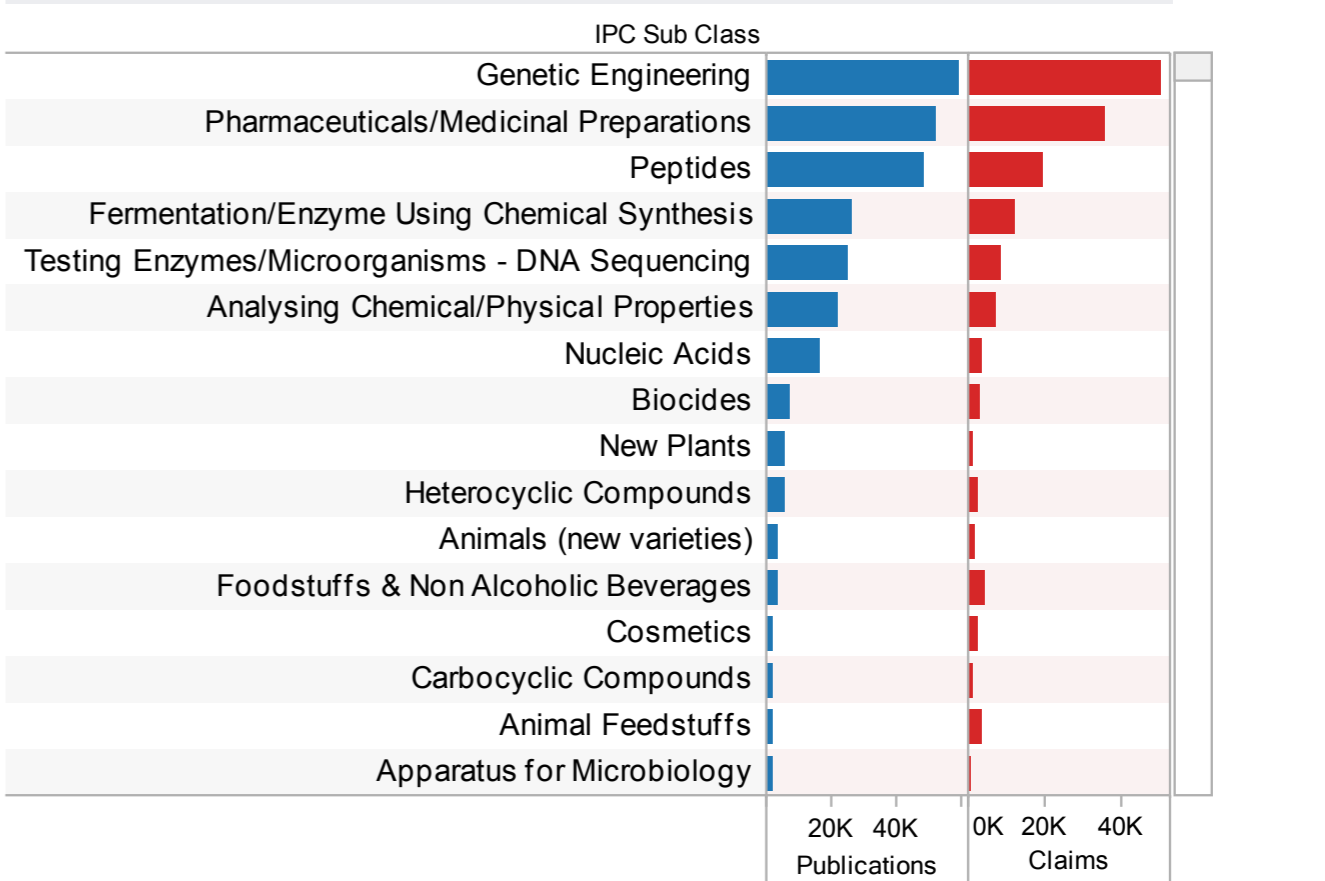
Marine Species (Ranked on Patent Publications)



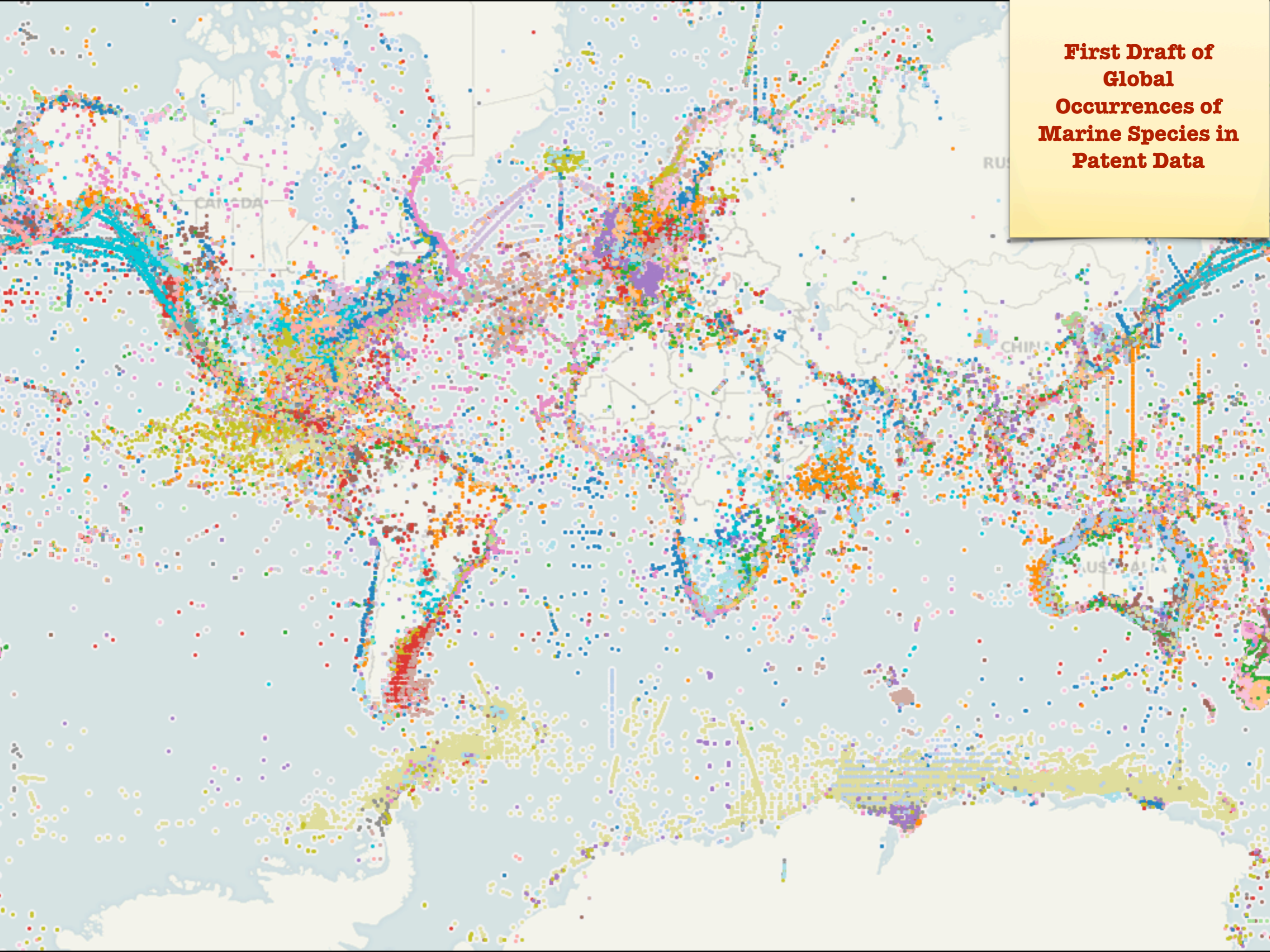
Applicants

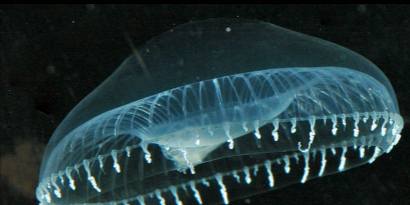



Technology Areas





**First Draft of
Global
Occurrences of
Marine Species in
Patent Data**

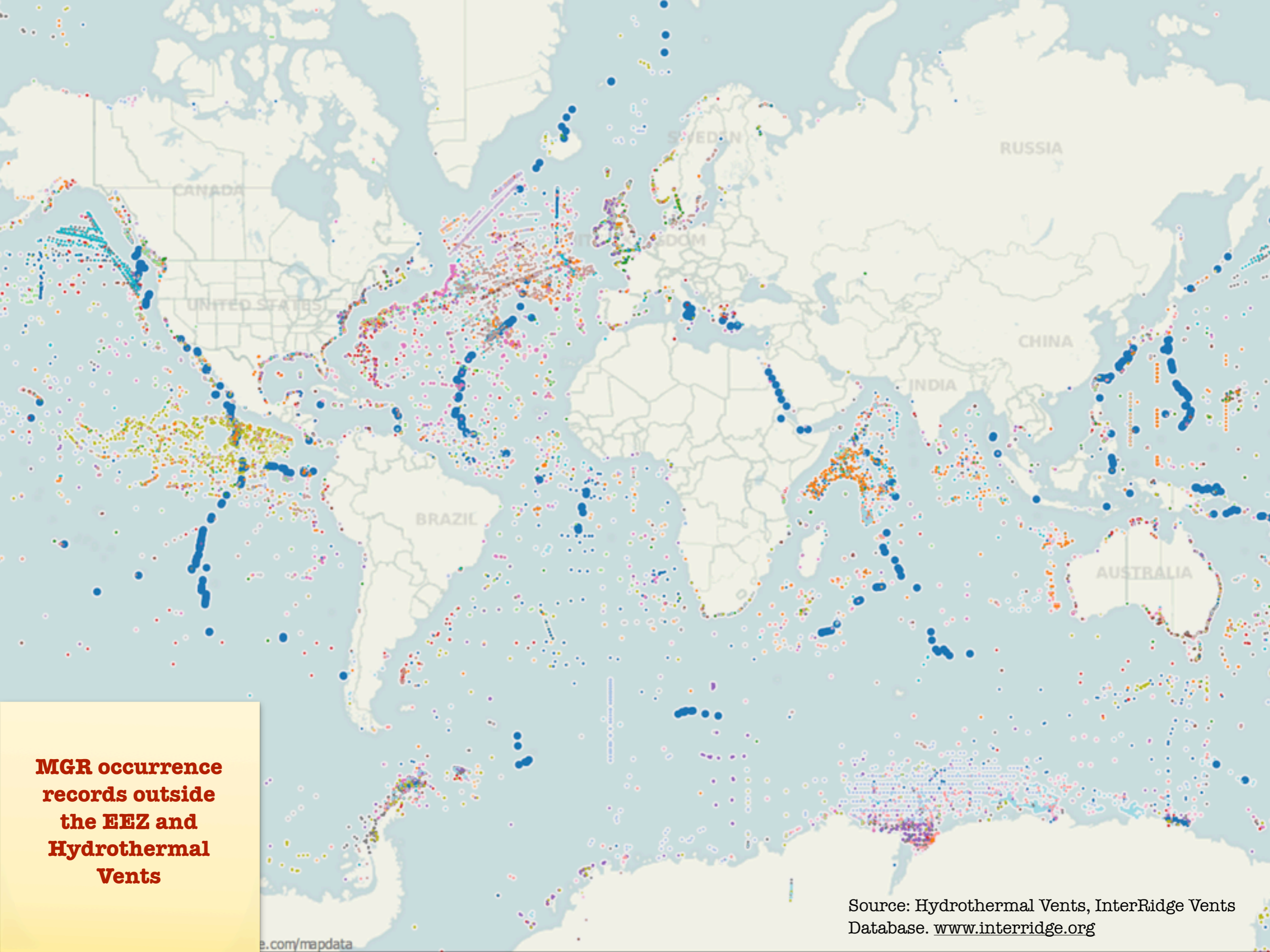


Species Name: Aequorea victoria	Kingdom: Animalia	
Brief Description of the species: A bioluminescent jellyfish		
Occurrences in Claims: 281	Number of Publications: 3825	
A bioluminescent jellyfish found along the western coast of north America including Vancouver Island. A species most noted for the production of Green Fluorescent Protein. The primary source of GFP which is used as a reporter of gene expression.		

Species Name: trichodesmium erythraeum	Kingdom: Bacteria	
Brief Description of the species: 'Sea Sawdust', a marine filamentous cyanobacteria that is nitrogen fixing.		
Occurrences in Claims: 62	Number of Publications: 119	
They are among the most important of the marine varieties of nitrogen fixing bacteria, and are being extensively studied for their role in nutrient cycling in the ocean. Trichodesmium is thought to fix nitrogen on such a scale that it accounts for almost half of the nitrogen-fixation within marine systems on a global scale. Occurs between China, Japan, South Korea and tip of western Australia.		

Species Name: Riftia pachyptila	Kingdom: Animal	
Brief Description of the species: Giant Tube Worm		
Occurrences in Claims: 4	Number of Publications: 25	
Large tubeworms which live around hydrothermal vents. With sunlight not available directly as a form of energy, the tubeworms rely on bacteria in their habitat to oxidize hydrogen sulphide using dissolved oxygen in the water for respiration. Occurs off coast of Ecuador & Galapagos, Pacific Coast of Mexico		

Species Name: Euphausia superba.	Kingdom: Animal	
Brief Description of the species: Antarctic Krill.		
Occurrences in Claims: 33 (Krill = 337)	Number of Publications: 109 (Krill = 2234)	
The keystone species for Antarctic ecosystems. The waters around the Antarctic continent harbour possibly the largest plankton assemblage in the world upon which the krill feed. The biomass of Antarctic krill is estimated to be 125 to 725 million tonnes.		



MGR occurrence records outside the EEZ and Hydrothermal Vents

Source: Hydrothermal Vents, InterRidge Vents Database. www.interridge.org



Identifying Origins

- Disclosure of origin debates involve two issues:
- a) Technical disclosure of the origin or source of genetic resources in a patent application;
- b) The consequences of disclosure or a failure to disclose;
- In practice samples may be obtained from sources such as: commercial suppliers, public collections, genomics databases (in silico), universities, or through in situ collection.

Oldham & Burton (2010) Defusing Disclosure in Patent Applications.
UNU study. UNEP/CBD/COP/10/INF/44

Record View: US6503729B1[Add to Work File](#) | [Mark Record](#) | [Watch Record](#) | [Download](#) ▼ | [Translate](#) ▼ | [Citation Map](#) | [Highlight](#) | [Print](#)  **Full View** [Jump to: Bibliography](#) [Abstract](#) [Classes/Indexing](#) [Legal Status](#) [Family](#) [Claims](#) **[Description](#)** [Citations](#) [Other](#)**Background/ Summary ?** Collapse Background/Summary**BACKGROUND OF THE INVENTION**

1. Field of the Invention

The present application discloses the complete 1.66-megabase pair genome sequence of an autotrophic archaeon, *Methanococcus jannaschii*, and its 58- and 16-kilobase pair extrachromosomal elements. Also identified are 1738 predicted protein-coding genes.

2. Related Background Art

The view of evolution in which all cellular organisms are in the first instance either prokaryotic or eukaryotic was challenged in 1977 by the finding that on the molecular level life comprises three primary groupings (Fox, G. E., et al., *Proc. Natl. Acad. Sci. USA* 74:4537 (1977); Woese, C. R. & Fox, G. E., *Proc. Natl. Acad. Sci. USA* 74:5088 (1977); Woese, C. R., et al., *Proc. Natl. Acad. Sci. USA* 87:4576 (1990)): the eukaryotes (Eukarya) and two unrelated groups of prokaryotes, Bacteria and a new group now called the Archaea. Although Bacteria and Archaea are both prokaryotes in a cytological sense, they differ profoundly in their molecular makeup (Fox, G. E., et al., *Proc. Natl. Acad. Sci. USA* 74:4537 (1977); Woese, C. R. & Fox, G. E., *Proc. Natl. Acad. Sci. USA* 74:5088 (1977); Woese, C. R., et al., *Proc. Natl. Acad. Sci. USA* 87:4576 (1990)). Several lines of molecular evidence even suggest a specific relationship between Archaea and Eukarya (Iwabe, N., et al., *Proc. Natl. Acad. Sci. USA* 86:9355 (1989); Gogarten J. P., et al., *Proc. Natl. Acad. Sci. USA* 86:6661 (1989); Brown, J. R. and Doolittle, W. F., *Proc. Natl. Acad. Sci. USA* 92:2441 (1995)).

The era of true comparative genomics has been ushered in by complete genome sequencing and analysis. We recently described the first two complete bacterial genome sequences, those of *Haemophilus influenzae* and *Mycoplasma genitalium* (Fleischmann, R. D., et al., *Science* 269:496 (1995); Fraser, C. M., et al., *Science* 270:397 (1995)). Large scale DNA sequencing efforts also have produced an extensive collection of sequence data from eukaryotes, including *Homo sapiens* (Adams, M. D., et al., *Nature* 377:3 (1995)) and *Saccharomyces cerevisiae* (Levy, J., *Yeast* 10:1689 (1994)).

M. jannaschii was originally isolated by J. A. Leigh from a sediment sample collected from the sea floor surface at the base of a 2600 m deep "white smoker" chimney located at 21 ° N on the East Pacific Rise (Jones, W., et al., *Arch. Microbiol.* 136:254 (1983)). *M. jannaschii* grows at pressures of up to more than 500 atm and over a temperature range of 48-94° C., with an optimum temperature near 85° C. (Jones, W., et al., *Arch. Microbiol.* 136:254 (1983)). The organism is autotrophic and a strict anaerobe; and, as the name implies, it produces methane. The dearth of archaeal nucleotide sequence data has hampered attempts to begin constructing a comprehensive comparative evolutionary framework for assessing the molecular basis of the origin and diversification of cellular life.

SUMMARY OF THE INVENTION

The present invention is based on whole-genome random sequencing of an autotrophic archaeon, *Methanococcus jannaschii*. The *M. jannaschii* genome consists of three physically distinct elements: (i) a large circular chromosome; (ii) a large circular extrachromosomal element (ECE); and (iii) a small circular extrachromosomal element (ECE). The nucleotide sequences generated, the *M. jannaschii* chromosome, the large ECE, and the small ECE, are respectively provided as SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3.

The present invention is further directed to isolated nucleic acid molecules comprising open reading frames (ORFs) encoding *M. jannaschii* proteins. The

***M. jannaschii*
genome. Univ
Illinos, TIGR, Univ
John Hopkins 2003.
Granted only for
gene fragments**

1

SELECTED POLYNUCLEOTIDE AND
POLYPEPTIDE SEQUENCES OF THE
METHANOGENIC ARCHAEON,
METHANOCOCCUS JANNASHII

CROSS REFERENCE TO RELATED
APPLICATIONS

This application claims benefit under 35 U.S.C. §119(e) to U.S. Provisional Application No. 60/024,428 filed on Aug. 22, 1996, which is herein incorporated by reference in its entirety.

STATEMENT AS TO RIGHTS TO INVENTIONS
MADE UNDER FEDERALLY-SPONSORED
RESEARCH AND DEVELOPMENT

Part of the work performed during development of this invention utilized U.S. Government funds. The U.S. Government may have certain rights in the invention—DE-FC02-95ER61962; DE-FC02-95ER61963; and NAGW 2554.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present application discloses the complete 1.66-megabase pair genome sequence of an autotrophic archaeon, *Methanococcus jannaschii*, and its 58- and 16-kilobase pair extrachromosomal elements. Also identified are 1738 predicted protein-coding genes.

2. Related Background Art

The view of evolution in which all cellular organisms are in the first instance either prokaryotic or eukaryotic was challenged in 1977 by the finding that on the molecular level life comprises three primary groupings (Fox, G. E., et al., *Proc. Natl. Acad. Sci. USA* 74:4537 (1977); Woese, C. R. & Fox, G. E., *Proc. Natl. Acad. Sci. USA* 74:5088 (1977); Woese, C. R., et al., *Proc. Natl. Acad. Sci. USA* 87:4576 (1990)): the eukaryotes (Eukarya) and two unrelated groups of prokaryotes, Bacteria and a new group now called the Archaea. Although Bacteria and Archaea are both prokaryotes in a cytological sense, they differ profoundly in their

2

ures of up to more than 500 atm and over a range of 48–94° C., with an optimum temperature of 68° C. (Jones, W., et al., *Arch. Microbiol.* 136:219–224 (1982)). The name of the organism is autotrophic and a strict anaerobe. The name implies, it produces methane. The discovery of the nucleotide sequence data has hampered attempts at constructing a comprehensive comparative framework for assessing the molecular basis of the evolution and diversification of cellular life.

SUMMARY OF THE INVENTION

The present invention is based on whole-genome random sequencing of an autotrophic archaeon, *Methanococcus jannaschii*. The *M. jannaschii* genome consists of three physically distinct elements: (i) a large circular chromosome; (ii) a large circular extrachromosomal element (ECE); and (iii) a small circular extrachromosomal element (ECE). The nucleotide sequences generated, the *M. jannaschii* chromosome, the large ECE, and the small ECE, are respectively provided as SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3.

The present invention is further directed to isolated nucleic acid molecules comprising open reading frames (ORFs) encoding *M. jannaschii* proteins. The present invention also relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of *M. jannaschii* proteins. Further embodiments include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to the nucleotide sequence of a *M. jannaschii* ORF described herein.

The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, host cells containing the recombinant vectors, as well as methods for making such vectors and host cells for *M. jannaschii* protein production by recombinant techniques.

The invention further provides isolated polypeptides encoded by the *M. jannaschii* ORFs. It will be recognized that some amino acid sequences of the polypeptides

The 1980 Bayh-Dole Act Disclosure Statement for Federal Funding Facilitates Monitoring

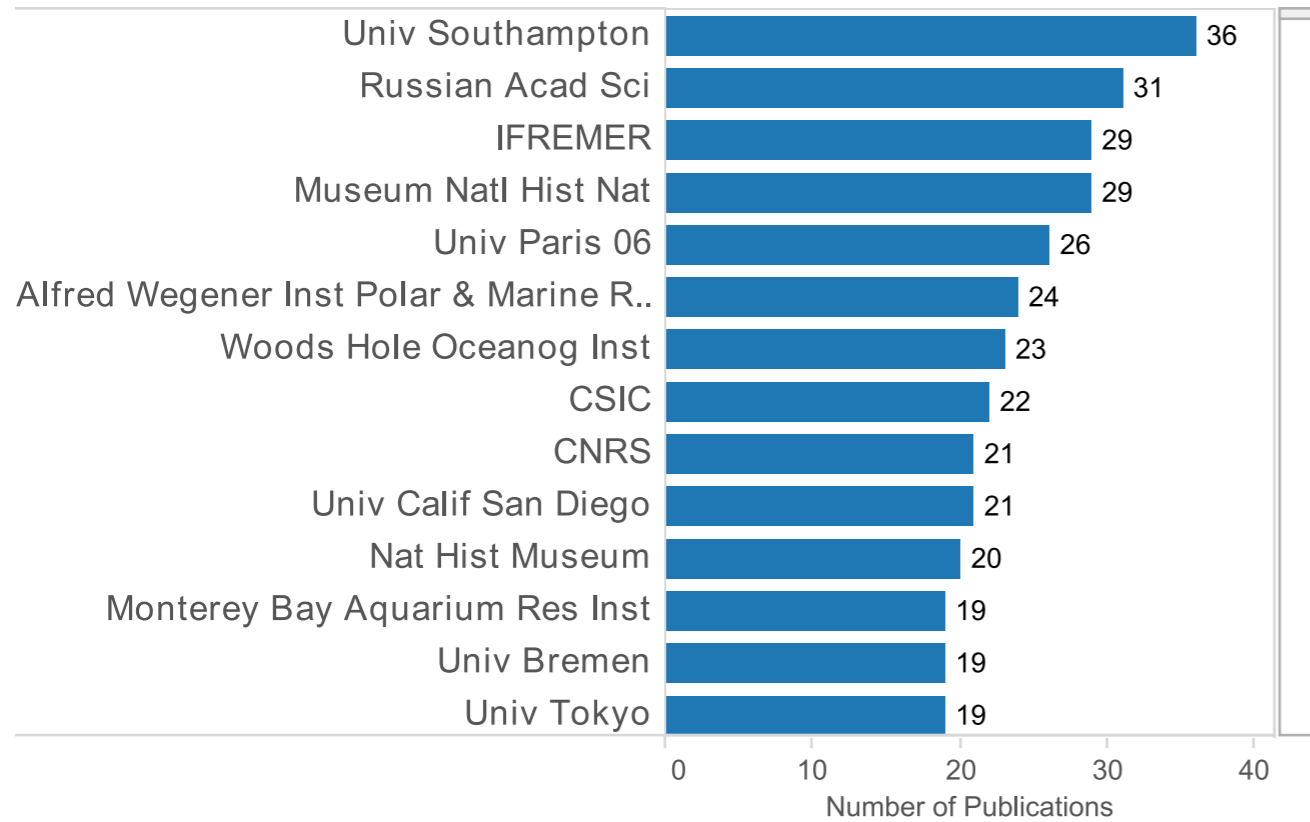


Non-Monetary Benefit-Sharing

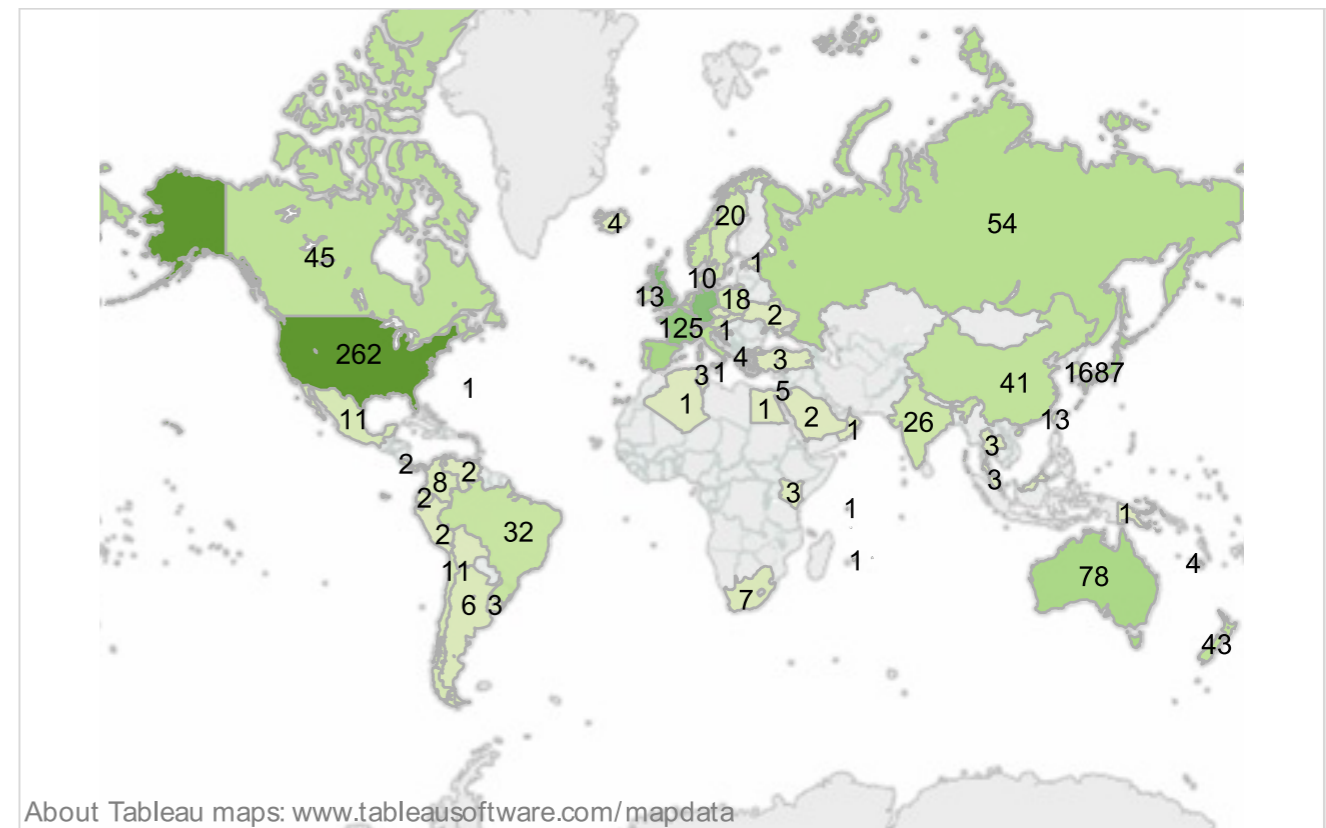
- Patents provide a good proxy indicator for intended commercial research and development;
- In debates on access and benefit-sharing considerable emphasis is also placed on non-monetary benefit sharing. However, this is difficult to quantify and to 'see';
- A way forward is provided by examining the data that accompanies scientific literature to reveal networks of collaboration, knowledge transfer and funding in scientific research on marine genetic resources.

Monitoring and Mapping the Scientific Literature for Marine Genetic Resources (Test Data)

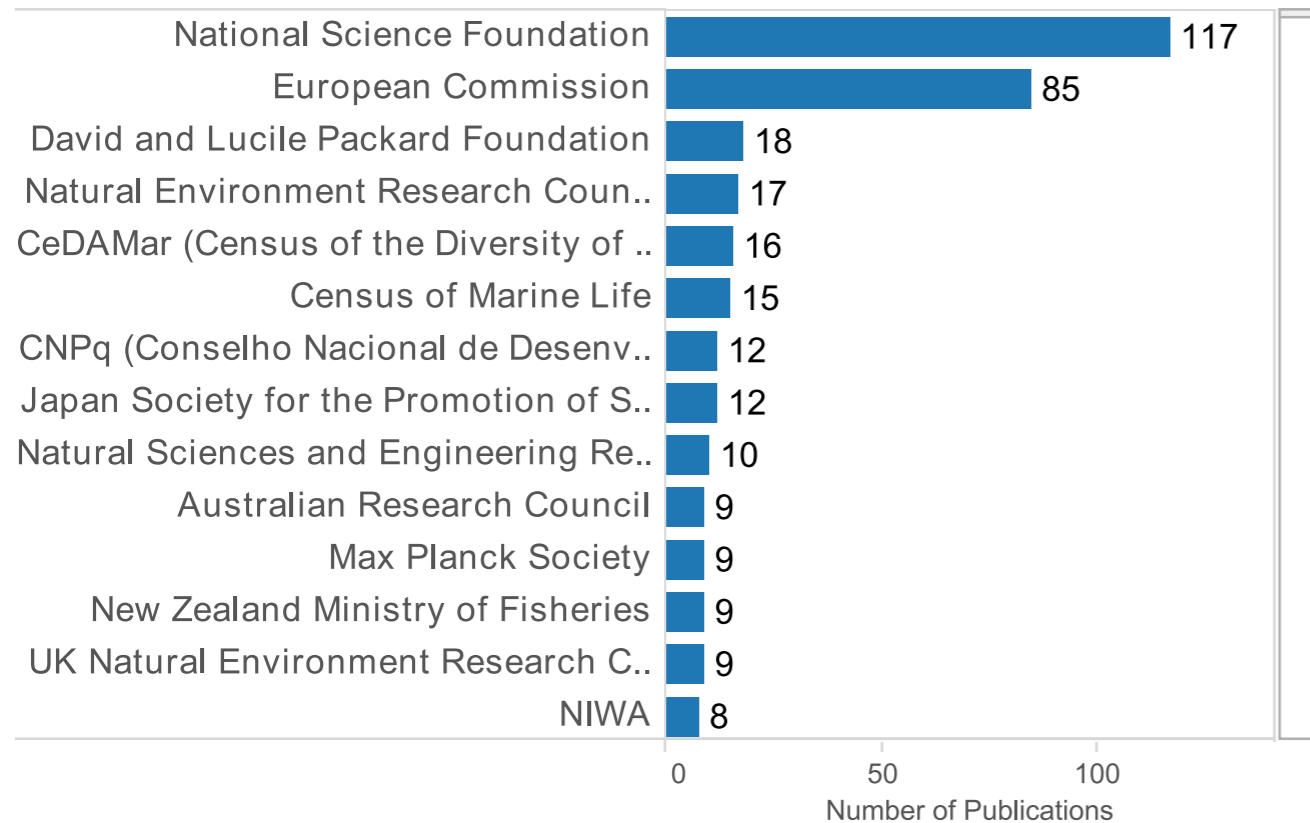
Author Affiliations



Country



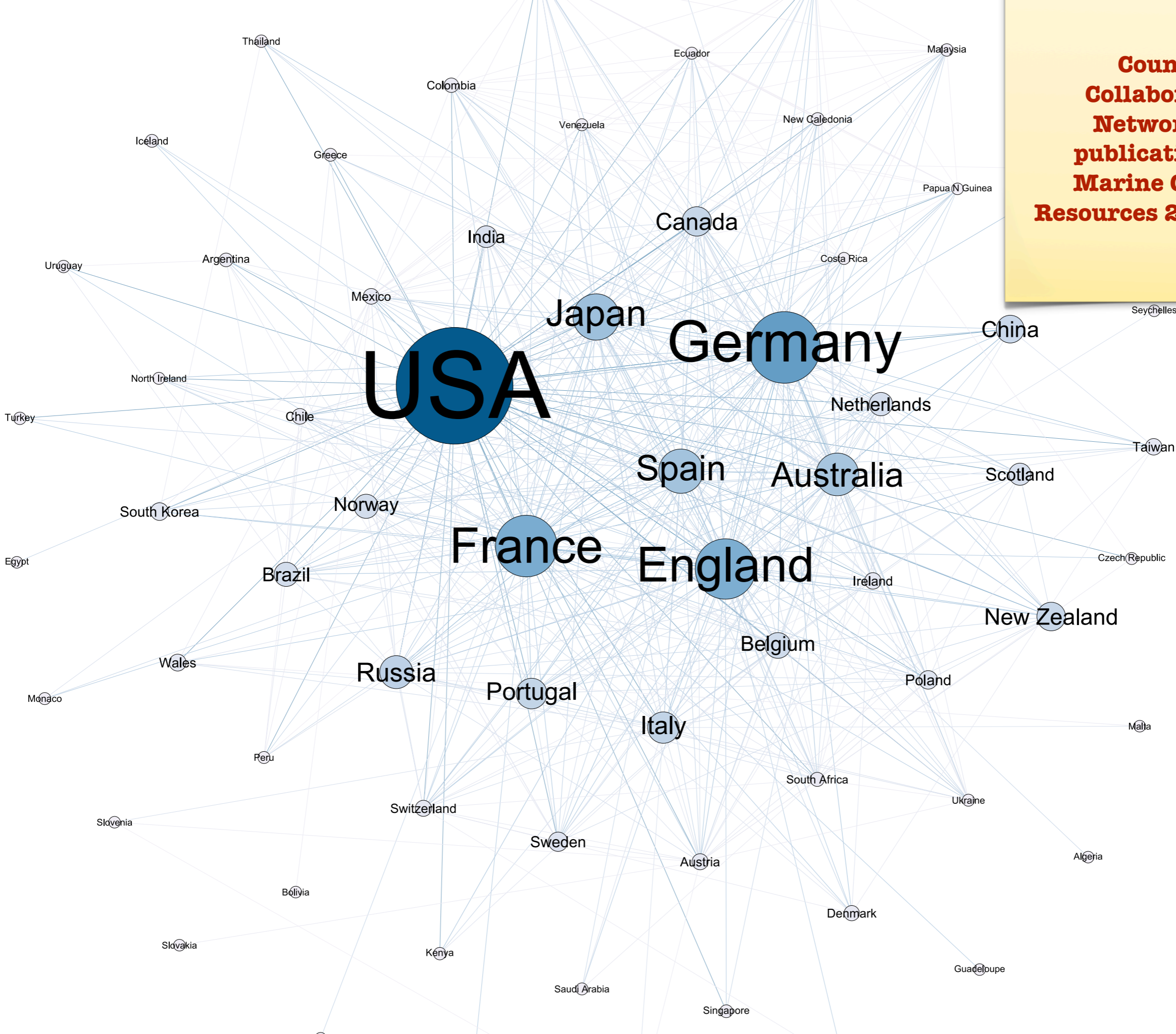
Funding Acknowledgements



Title

- 2-D difference gel electrophoresis approach to assess protein expression p
- 'Pliocardia' krylovata, a new species of vesicomid clam from cold seeps al
- 'Sponge-specific' bacteria are widespread (but rare) in diverse marine enviro
- A checklist of the extant species of ophiuroids (Echinodermata: Ophiuroide
- A comparative study on the visual adaptations of four species of moray eel
- A continuum of life histories in deep-sea demersal fishes
- A core-top calibration of B/Ca in the benthic foraminifers Nuttallides umbon
- A geographic test of species selection using planktonic foraminifera during
- A hybrid zone between Bathymodiolus mussel lineages from eastern Pacific
- A hydrothermal seep on the Costa Rica margin: middle ground in a continu
- A laboratory guide to in vitro fertilization of oceanic squids
- A large new species of the genus Ptilocrinus (Echinodermata, Crinoidea, Hy
- A latest Permian radiolarian fauna from Hushan, South China, and its geolo
- A molecular phylogenetic appraisal of the systematics of the Aglaopheniida
- A new deep-sea benthopelagic chaetognath of the genus Bathyspadella (Ch

**Country
Collaboration
Network for
publications on
Marine Genetic
Resources 2011-2013.**





Best Practices for Monitoring

- Improving coverage of marine species in analysing patent data (i.e. using WoRMS);
- Evidence based analysis of origins/sources in patents;
- Improving coverage & validating geocoding in GBIF data to remove noise;
- Using existing geocoded resources for deep sea locations;
- Using existing best practice and technical advice in patent analysis from the WIPO Patent Landscaping Group;
- Making indexed data available in an accessible and useful format;



Thank-you

The research in this presentation was made possible
with support from:

The UK Intellectual Property Office

Department for Environment, Food and Rural Affairs
(DEFRA), UK

Institute of Advanced Studies, United Nations University

Contact: poldham@mac.com

Background Image: Copyright Kate Whitley/Wellcome Trust, London

One World Analytics is a trading name of POMC Consulting Ltd., UK.