Supporting Information

Arrieta et al. 10.1073/pnas.0911897107

Data Sources and Processing

Data on patented DNA sequences were obtained from GenBank (1). The patent (PAT) division of GenBank release 165 (April 2008) contains 5,338,849 sequence records divided into 38 files, which were downloaded from the National Center for Biotechnology Information's Web site (ftp://ftp.ncbi.nih.gov/genbank/). Patent information was extracted and imported into a MySQL database (2) using Perl (3) scripts written for this purpose. The PAT division of GenBank contains patent records obtained from several patent authorities around the world, mainly from US, European, Japanese, and World patent organizations. Because the same sequences are often patented under different patent authorities, preliminary analysis of the database revealed a large degree of redundancy in the reported sequences. Therefore, identical sequences reported under different patent authorities were detected and assigned a unique sequence serial number using Perl scripts to eliminate redundancy in the database.

Preliminary analysis of the taxonomic information included in the database revealed that 24% (1,329,921 sequences) of the reported patented sequences were synthetic constructs. Also, a large percentage of the sequences (37%) were tagged as unknown or unclassified, meaning that no taxonomic information was available. Sequences of unknown origin accounted for minor parts of the sequences reported by the European (<12% of reported sequences), Japanese (<17% of reported sequences), or World (<3% of reported sequences) patent organizations. None of the sequences linked to the US Patent Office contained any taxonomic information about their origin. However, about 60% of these unknown sequences belonging to the US Patent Office also appeared in the other patent databases. There may be some additional sequences originating from marine organisms, and even some additional marine species among these patented sequences of unreported origin that cannot be identified. Moreover, the information contained in the PAT division of GenBank is by no means exhaustive; therefore, it is likely that additional marine sequences have been reported under other patent authorities not covered here. Thus, the impressive number of marine species having patented genes reported in this study represents only a minimum estimate of the actual bioprospecting activity in the oceans.

The list of named species in the database was extracted, and the resulting name list was sorted and cleaned up manually to eliminate obvious spelling variants, typos, strain names or numbers, and other confounding information that could result in an overestimation of the number of named species. The list of clean unique species names was manually reviewed; marine organisms were identified and tagged in the database for subsequent analysis.

Information on the number of previously undescribed marine natural products reported per year and the corresponding phylogenetic information about the sources were compiled from several annual reports (4–9).

Because the taxonomic scheme used by GenBank (1) differs from that reported for natural products (4), the phylogenetic groups in Table S1 and in the trees in Fig. 2 and Fig. S1 have been chosen to match these two datasets and allow comparisons. Therefore, the chosen taxonomic affiliations do not represent a particular level (i.e., class, order) in a systematic hierarchy. The trees in figures 2 and fig S1 have been produced using the Interactive Tree of Life iTOL tool (10).

1. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. *Nucl Acids Res* 36:D25–D30.
2. MySQL 5.1. Available at http://dev.mysql.com/downloads.
3. Perl 5.8. Available at http://www.perl.org.
4. Blunt JW, et al. (2008) Marine natural products. *Nat Prod Rep* 25:35–94.
5. Blunt JW, et al. (2007) Marine natural products. *Nat Prod Rep* 24:31–86.
6. Blunt JW, Copp BR, Munro MHG, Northcote PT, Prinsep MR (2006) Marine natural products. *Nat Prod Rep* 23:26–78.
7. Blunt JW, Copp BR, Munro MHG, Northcote PT, Prinsep MR (2005) Marine natural products. *Nat Prod Rep* 22:15–61.
8. Blunt JW, Copp BR, Munro MH, Northcote PT, Prinsep MR (2004) Marine natural products. *Nat Prod Rep* 21:1–49.
9. Blunt JW, Copp BR, Munro MH, Northcote PT, Prinsep MR (2003) Marine natural products. *Nat Prod Rep* 20:1–48.
10. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.

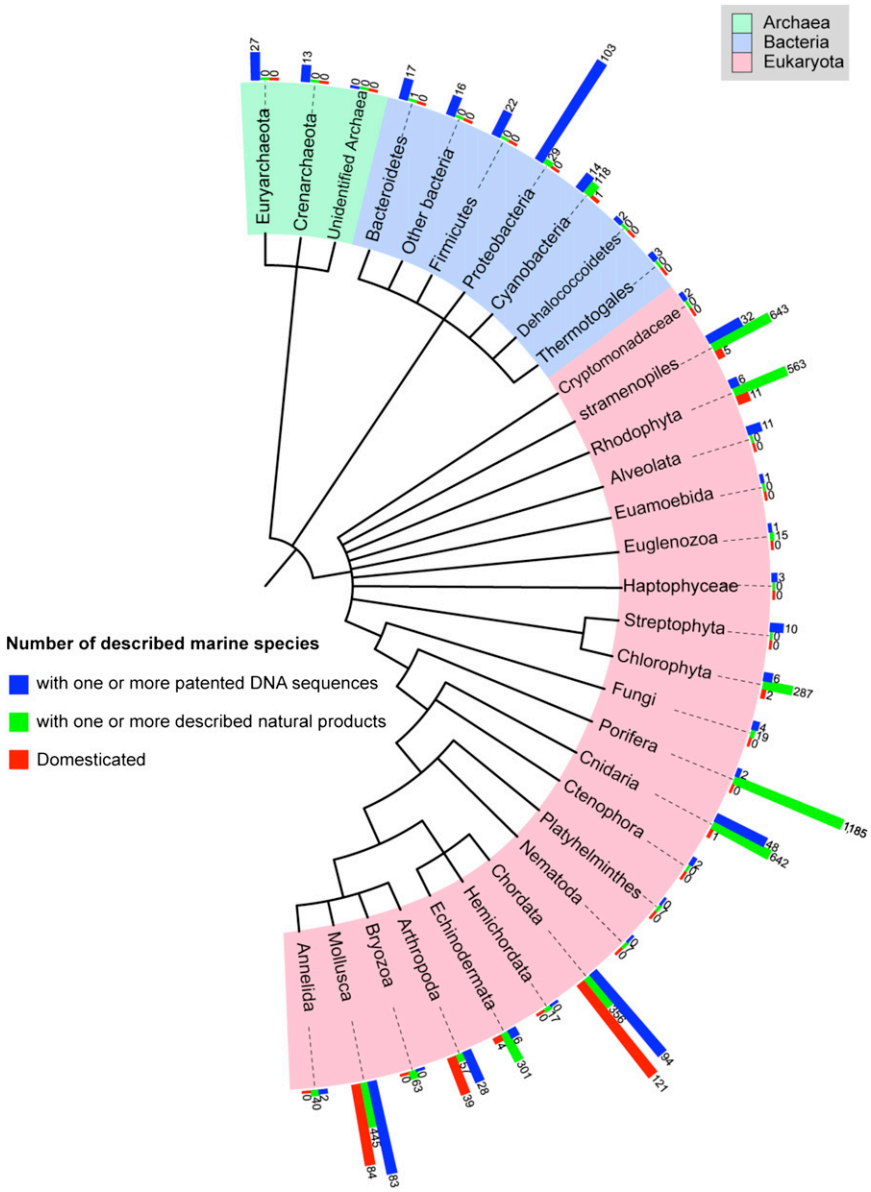

Fig. S1. Number of marine species subjected to different uses divided among different taxonomic groups.

Table S1. Number of named marine species accounted as sources of patented genes, natural products, or domesticated for human consumption according to their taxonomic affiliation

	Marine GenBank patents		Marine natural products		
Taxonomic group	Species*	Sequences*	Species†	Compounds†	Domesticatedspecies‡
Unidentified Archaea	0	18	0	0	0
Crenarchaeota	13	124	0	0	0
Euryarchaeota	27	422	0	0	0
Cyanobacteria	14	92	118	602	1
Proteobacteria	103	739	29	3	0
Firmicutes	22	240	0	0	0
Bacteroidetes	17	52	1	4	0
Thermotogales	3	161	0	0	0
Dehalococcoidetes	2	109	0	0	0
Other bacteria	16	62	0	0	0
Alveolata	11	28	0	0	0
Fungi	4	48	19	48	0
Haptophyceae	3	17	0	0	0
Euamoebida	1	3	0	0	0
Annelida	2	40	40	49	0
Arthropoda	28	124	57	23	39
Chordata	94	927	356	977	121
Cnidaria	48	498	642	3,469	1
Ctenophora	2	3	0	0	0
Echinodermata	6	17	301	1,074	4
Mollusca	83	760	445	1,179	84
Porifera	2	16	1,185	6,668	0
Rhodophyta	6	20	563	1,548	11
Chlorophyta	6	45	287	315	2
Streptophyta	10	135	0	0	0
Stramenopiles	32	223	643	1,476	5
Cryptomonadaceae	2	3	0	0	0
Euglenozoa	1	6	15	36	0
Bryozoa	0	0	63	180	0
Hemichordata	0	0	17	26	0
Nematoda	0	0	7	2	0
Platyhelminthes	0	0	7	33	0
Total	558	4,932	4,795	17,712	268

The number of unique sequences or natural products described for each taxonomic group is also reported.
*Patented genes. Unique sequences or named species (this paper).
†Natural products. Data from Blunt et al. (4).
‡Domesticated for human consumption. Data from Duarte CM, Marbá N, Holmer M (2007) Ecology. Rapid domestication of marine species. Science 316:382–383.