 doi:10.20944/preprints201812.0156.v1

# The report of marine life genomic research

Guangyi Fan[1,2], Jianwei Chen[1], Tao Jin[1], Chengcheng Shi[1], Xiao Du[1], He Zhang[1], Yaolei Zhang[1], Hanbo Li[1], Ting Luo[1], Pengxu Yan[1], Guang Liu[1], Xiangqun Chi[1], Xiaoxuan Tan[1], Liangwei Li[1], Guilin Liu[1], Xiaochuan Liu[1], Shijie Hao[1], Kai Han[1], Xiaoyun Huang[1], Shuai Sun[1], Jing Zhou[1], Mengjun Yu[1], Lingfeng Meng[1], Yue Chang[1], Rui Zhang[1], Kaiqiang Liu[1], Mengqi Zhang[1], Yong Zhao[1], Chang Li[1], Jiao Guo[1], Xinyu Guo[1], Jiahao Wang[1], Meiqi Lv[1], Haoyang Gao[1], Yujie Liu[1], Yue Song[1], Shengjun Wang[1], Yang Deng[1], Binjie Ouyang[1], Jinzhong Lin[1], Yingjia Yu[1], lynn Fink[4], Xianwei Yang[1], Xun Xu[1,2,3], Xin Liu[1,2,3].

[1] BGI-Qingdao, BGI-Shenzhen, Qingdao, Shandong Province, 266555, China.

[2] BGI-Shenzhen, Shenzhen, 518083, China.

[3] China National GeneBank, BGI-Shenzhen, Shenzhen, 518120, China.

[4]BGI-Australia, QLD 4006 Australia.

Correspondence should be addressed to Xin Liu (liuxin@genomics.cn) and Guangyi Fan (fanguangyi@genomics.cn).


Key words: marine microorganisms, marine fungi, algae, marine plants, marine invertebrates, marine vertebrates, genome, metagenome.

 doi:10.20944/preprints201812.0156.v1

# Preface

With the continuing development of sequencing technology, genomics has been applied in a variety of biological research areas. In particular, the application of genomics to marine species, which boast a high diversity, promises great scientific and industrial potential. Significant progress has been made in marine genomics especially over the past few years. Consequently, BGI, leveraging its prominent contributions in genomics research, established BGI-Qingdao, an institute specifically aimed at exploring marine genomics. In order to accelerate marine genomics research and related applications, BGI-Qingdao initiated the International Conference on Genomics of the Ocean (ICG-Ocean) to develop international collaborations and establish a focused and coherent global research plan. Last year, the first ICG-Ocean conference was held in Qingdao, China, during which 47 scientists in marine genomics from all over the world reported on their research progress to an audience of about 300 attendees. This year, we would like to build on that success, drafting a report on marine genomics to draw global attention to marine genomics. We summarized the recent progress, proposed future directions, and we would like to enable additional profound insights on marine genomics. Similar to the annual report on plant and fungal research by Kew Gardens, and the White Paper of ethical issues on experimental animals, we hope our first report on marine genomics can provide some useful insights for researchers, funding agencies as well as industry, and that future versions will expand upon the foundation established here in both breadth and depth of knowledge.

This report summarizes the recent progress in marine genomics in six parts including: marine microorganisms, marine fungi, marine algae and plants, marine invertebrates, marine vertebrates and genomics-based applications.

 doi:10.20944/preprints201812.0156.v1

# Content

**Reports of marine life genomic research** ..... 1

**Preface** ..... 1

**1 Overview** ..... 5

**1.1 Current status of marine genomics** ..... 5

**1.2 Summary of marine organism genomes** ..... 7

**1.3 Sequencing technology** ..... 9

**1.4 Large-scale genome projects** ..... 10

**2 Genomics of marine microorganisms** ..... 12

**2.1 Genomes of bacteria and archaea** ..... 12

**2.2 Marine metagenomics** ..... 16

**2.3 Genomics of marine viruses** ..... 22

**3 Genomics of marine fungi** ..... 24

**3.1 Basic introduction of marine fungi** ..... 24

**3.2 High-throughput sequencing for marine fungi** ..... 26

**4 Genomics of marine algae and plants** ..... 30

**4.1 Algae genome** ..... 30

**4.2 Genomics of marine plants** ..... 34

**5 Marine Invertebrates** ..... 36

**5.1 Global diversity and phylogeny** ..... 36

**5.2 Genomics of marine invertebrates** ..... 37

**6 Fish genomes** ..... 44

**6.1 Brief introduction of fish** ..... 44

**6.2 Research focuses of fish genomics** ..... 46

6.2.1 Viviparity in teleost ..... 46

6.2.2 Water-to-land transition ..... 47

6.2.3 Adaptation to extreme cold ..... 47

6.2.4 Convergent evolution toward adaptation to darkness ..... 48

6.2.5 Fish disease and immunity ..... 48

6.2.6 Sex determination ..... 49

6.2.7 Metamorphosis ..... 50

**7 Genomics of marine tetrapods** ..... 52

**7.1 Brief introduction and genomes** ..... 52

**7.2 Current status of marine tetrapod genomes** ..... 52

**7.3 Conservation of marine tetrapods using genomics** ..... 54

 doi:10.20944/preprints201812.0156.v1

**8 Applications of genomic data.....57**

**8.1 Genetic engineering..... 57**

**8.2 Marine natural products ..... 60**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

# 1 Overview

## 1.1 Current status of marine genomics

The ocean, comprising the majority of our planet's hydrosphere, is the cradle of life. After evolving for billions of years, more than two million species inhabit the ocean, of which only 230,000 species are documented. The high biodiversity in the ocean provides unprecedented opportunity to explore various scientific questions, including the origin and evolution of life, adaptation to different environments, chemo- and photosynthesis, ecology, etc. Marine life can also serve as a crucial food resource for the future development of human society, providing sustainable protein, peptides and metabolites. Despite the importance and potential of marine life exploration and research, current biological research is relatively limited, especially compared to exploration of ocean resources, the development of marine equipment, and biological research of land plants and animals (for example, humans - ourselves). Thanks to the development of biotechnology, research in marine biology has made great progress in the past decade, especially with the recent developments in sequencing technology and genomics. Even marine life without a clear evolutionary background can be studied in more efficiently. Subsequently, marine genomics, which uses cutting-edge sequencing technologies to produce genomic data supported by bioinformatics analysis of the data, has significantly facilitated improvements in marine biology and industrial applications in recent years (Fig. 1.1).

Subsequent to the publication of the first fish genome (*Fugu rubripes*) in 2002 (see a list of first genomes from different clades of marine species in Table 1.1), 453 marine species now have a published reference genome, and more than 130 Tb of sequenced data, including 107 Tb metagenomics data, are publicly available. Despite the progress of marine genomics, there are still challenges ahead. These include discrepancies in data distribution due to biased sampling, difficulties in sample preparation and genome complexity. However, recent developments in sequencing technology have vastly

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

accelerated data generation and extended read lengths, while simultaneously reducing costs, thus creating opportunities for future research into marine species without reference genomes as well as populations with reference genomes, making marine genomics more scientifically rigorous and applicable to conservation and industrial applications.

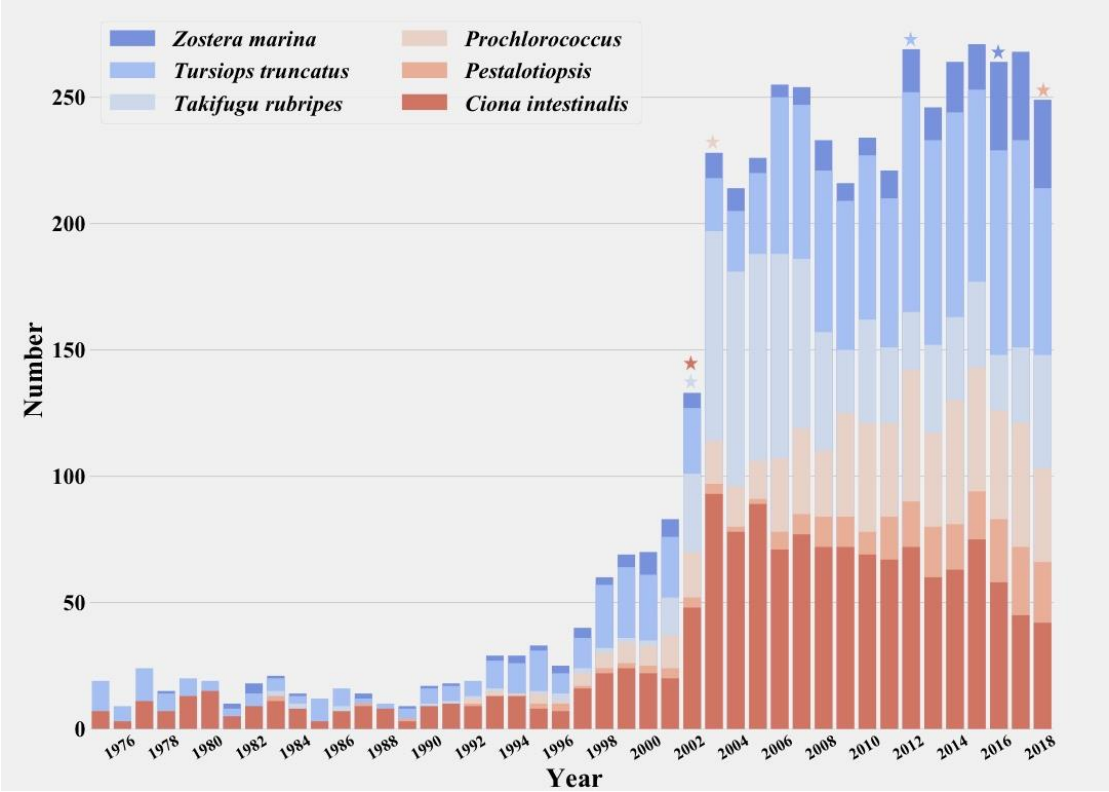


**Fig. 1.1 Publications describing marine species increased after publication of reference genomes.**

**Table 1.1 Representative publication of the first marine reference genomes in different research areas.**

| Title | Research area | Milestone | Journal & Time |
|---|---|---|---|
| The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea | Marine flowering plant | The first published marine flowering plant genome[1] | Nature, 2016 |
| Genome sequence of | Algae | The first published complete | Nature, |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| | | | |
|---|---|---|---|
| the ultrasmall unicellular red alga Cyan idioschyzon merolae 10D | | algal genome[2] | 2004 |
| Comparative genomics reveals insights into avian genome evolution and adaptation | Marine vertebrate | Twelve Marine bids published in the special issue of bird genome paper[3] | Science, 2014 |
| Structure and function of the global ocean microbiome | Marine microbe | The first comprehensive meta-genome reference of marine environment using NGS technology[4] | Science, 2015 |
| The oyster genome reveals stress adaptation and complexity of shell formation | Marine invertebrate | The first published high-quality mollusk genome using NGS technology[5] | Nature, 2012 |
| The Draft Genome of *Ciona intestinalis*: Insights into Chordate and Vertebrate Origins | Marine invertebrate | The first published invertebrate genome [6] | Science, 2002 |
| The genome sequence of Atlantic cod reveals a unique immune system | Fish | The first published fish genome using NGS technology[7] | Nature, 2011 |
| Whole-Genome Shotgun Assembly and Analysis of the Genome of *Fugu rubripes* | Fish | The first published fish genome[8] | Science, 2002 |
| Growth enhancement in transgenic Atlantic salmon by the use of an "all fish" chimeric growth hormone gene construct. | Genetic engineering | The first genetically engineered commercial fish to enter the market[9] | Nat Biotechnology, 1992 |
| Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation | Marine microbe | The first ocean bacteria sequenced[10] | Nature, 2003 |

## 1.2 Summary of marine organism genomes

We summarized the basic characteristics of published genomes of marine eukaryotic organisms including assembled genome size, GC content ratio, contig N50, scaffold N50, and BUSCO - one of important indicators for assessing genome integrity (Fig. 1.2).

**1) Genome size.** Fish and fungi genomes (the majority of which are 627-940M and ~25M-40Mb, respectively) have the most consistent genome sizes while tetrapod genome sizes can be clustered into two groups: ~1.2Gb for seabirds and ~2.5Gb for mammals. Relatively speaking, algae and invertebrates contain more species, more

 doi:10.20944/preprints201812.0156.v1

complex genomes, and their genome sizes also vary more than other classifications.

**2) GC content ratio.** GC content differs amongst the five clades. Algae genomes have the highest GC ratio (~50%-62%) while invertebrates have the lowest (~34-39%). Tetrapod genomes exhibit the most consistent GC content, ~41%.

**3) Contig N50.** Fungal genomes have a notably higher contig N50 value (~67-456Kb) than other clades; fish, tetrapod and algae genomes are similar. In contrast, invertebrate genomes generally exhibit a smaller contig N50 value (most less than 25Kb).

**4) Scaffold N50.** Tetrapods exhibit the highest scaffold N50 value, reaching ~64Mb, followed by fish and fungi, and then by algae and invertebrates (most less than 1Mb).

**5) BUSCO**. According to this criterion, the fungal genome assembly is the most complete, while algae and invertebrate genomes are inferior. In summary, tetrapods (mainly seabirds and mammals) and fungi have a higher quality assembled genomes compared to other clades, likely because of their relatively simple genomes even though some of these genomes tend to be quite large (e.g., mammalian genomes). All of the indicators for fish are relatively mild, reflecting the stability of the fish genomes. The invertebrate and algae genomes are the most complex, and their genomic characteristics and assembly quality are quite different from the other clades.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

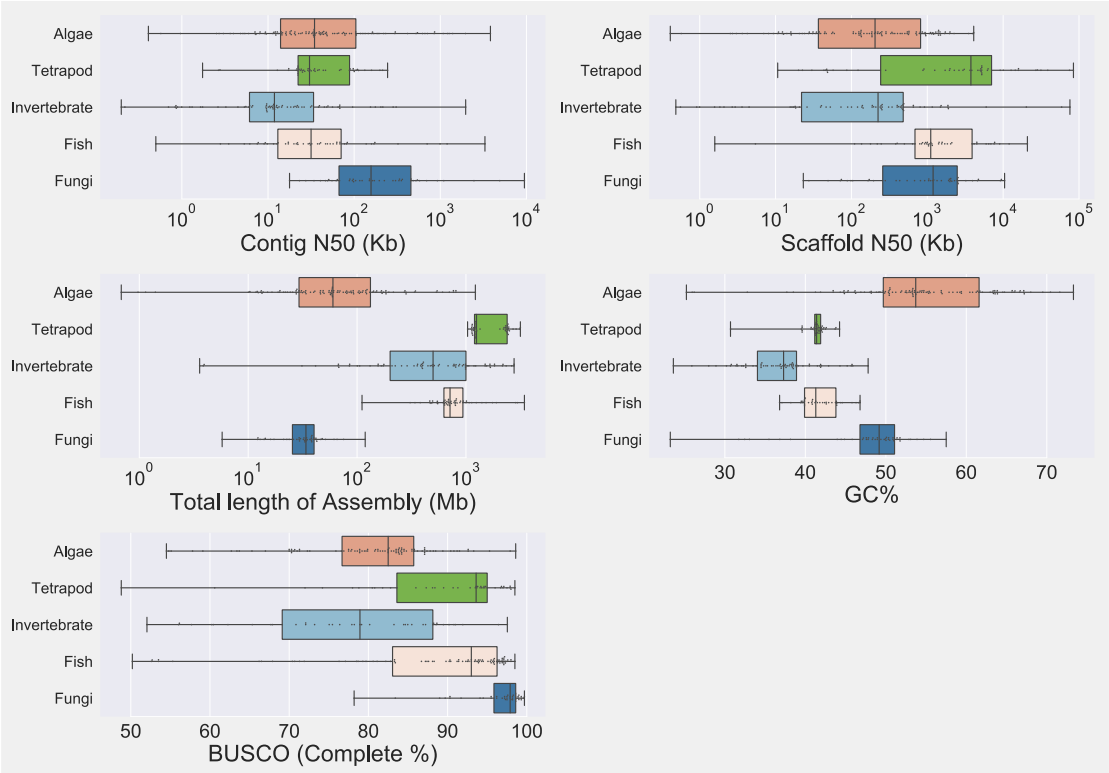


**Fig. 1.2 Summary of published marine organism genome assemblies.** Comparisons of contig N50, scaffold N50, total length, GC content and BUSCO among algae, marine tetrapod, marine invertebrate, fish and marine fungi.

## 1.3 Sequencing technology

Sequencing technology is increasing the pace of genomic research. After the invention of the ABI 370 sequencer in 1987, genomics research entered a new era of high-throughput sequencing. The first marine organism genome project, a fish genome project started in 2001 completed using Sanger sequencing technology. Subsequently, the progress of marine organism genome sequencing projects slowed until 2010 when Illumina released the Hiseq2000, their sequencing platform which became widely adopted (Fig. 1.3). As a result, the first assembled genomes of algae, fish, fungus, and tetrapods were completed in 2011 with second generation sequencing technology, heralding the explosion of marine organism genomics. By 2015-2016, third generation sequencing technology started appearing in algae, fish, fungus and invertebrate genome projects. However, no marine tetrapod genome projects have yet been completed using

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

third generation sequencing technology (Fig. 1.4).

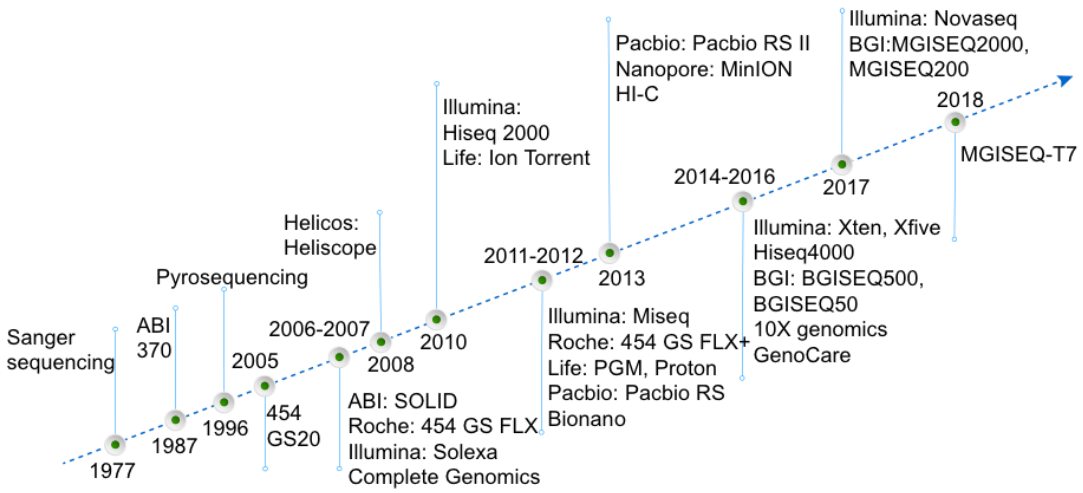


**Fig. 1.3 The development process of sequencing platforms and important associated technologies.**

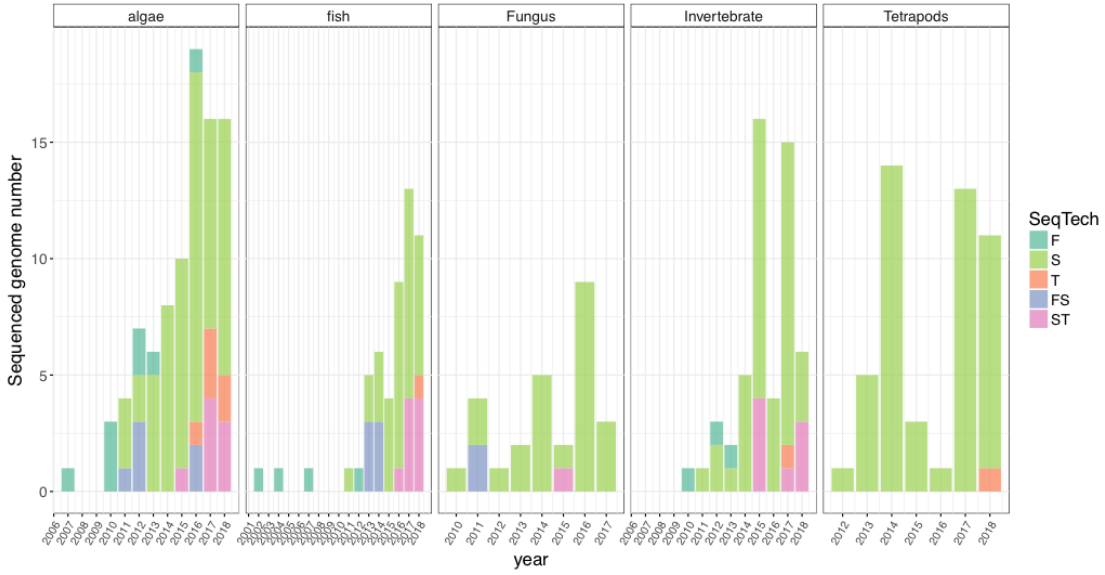


**Fig. 1.4 Timeline of published algae, fish, marine fungus, marine invertebrate and marine tetrapod genomes.** F: sanger sequencing technology; S: next-generation sequencing technology; T: single-molecular long read sequencing technology; FS: combining F and S. ST: combining S and T.

## 1.4 Large-scale genome projects

Large-scale genome projects are highly efficient in accelerating genomics research.

 doi:10.20944/preprints201812.0156.v1

Several large-scale genome projects in marine genomics have been initiated during the last decade. The Tara Oceans project led by EMBL (European Molecular Biology Laboratory), initiated in September 2009, has collected more than 30,000 ocean environmental samples from more than 200 ocean stations, and at least 243 of those samples have been sequenced, thus creating the largest environmental sequencing dataset[4]. The ongoing Transcriptomes of 1,000 Fishes (Fish-T1K) project, which aims to sequence the transcriptomes of 1,000 fishes[11], recently completed the transcriptomes of 159 fishes. The Genome 10K project (G10K), which aims to sequence the genomes of 10,000 vertebrates, includes over 4,000 fish[12]. In addition to G10K, there are other large-scale projects which plan to sequence genomes of marine species. For example, the 10KP (10 thousand plant genome project) [13] plans to sequence 4,000 algae species, and the ambitious Earth BioGenome Project (EBP) [14], which plans to sequence all known eukaryote species on earth, will also cover many marine species. However, there is yes to be established a genome sequencing initiative which is systematically designed for marine genomics and to set the course for all future marine genomics investigations. Overall, the recent progress of marine genomics has enabled the understanding of biological diversity and evolution in the ocean, and provided insights into ecological conservation, both of which are necessary to develop a sustainable human society. This report will describe the progress of marine genomics in different clades of species.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

# 2 Genomics of marine microorganisms

Marine microorganisms are highly diversified, with single-cell organisms and simple multicellular organisms from three phylogenetic groups of bacteria, archaea and eukaryotes, as well viruses and viroid. After a billion years of evolution, the marine microbiome has adapted to complex ocean environments. In recent years, with the rapid development of high-throughput sequencing, single-cell screening, and bioinformatics, genomics research in marine microbial has developed at a similarly rapid pace. Currently, there are about 8,000 genomes representing single marine bacterium species, 47% of which are Proteobacteria and 11.2% are photosynthetic bacteria. These two types of microorganisms are the most abundant and widely distributed marine bacterial species. In addition to single bacterium genomes, there are more than 100,000 metagenomics datasets, representing mixtures of DNA in environmental samples. The majority of these metagenomics sequencing efforts (91.2%) focused on amplified fragments of marker genes or conserved sequence, while 6.1% of the projects performed whole genome metagenomics sequencing (metagenome sequencing) and 2.7% performed meta-transcriptome sequencing. These datasets have vastly improved the understanding of marine microbial physiology and ecology, and have aided in further applications of this research.

## 2.1 Genomes of bacteria and archaea

The total number of marine bacteria has been estimated to be approximately $6.6 \times 10^{29}$, comprising the majority of global microbial biomass[15]. Bacterial species in different marine habitats, including coastal surface waters, open seas, and sediments are very different from each other. Despite their high diversity and relatively simple genome content, difficulties in cultivating the majority of the marine bacteria have impeded genome sequencing efforts.

Archaea account for more than 20% of all prokaryotes in seawater, and are the most

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

important microbial group in marine subsurface sediments and most geothermal habitats[16]. Most archaea resist culturing efforts and colonies that can be cultured primarily belong to *Euryarchaeota* and *Crenarchaeota*. Recent studies have shown that archaea are divided into at least four major superphyla: *Euryarchaeota*, the TACK superphylum, the DPANN superphylum, and the Asgard superphylum[17] (Fig. 2.1). Phylogenetic analysis of genomic datasets suggests that *Lokiarchaeota* (Asgard superphylum) is the most closely related group of eukaryotes, which provides further convincing evidence that eukaryotes evolved from archaea[17] and suggests that the origin of eukaryotic cells is one of the major evolutionary innovations in the life history of our planet.

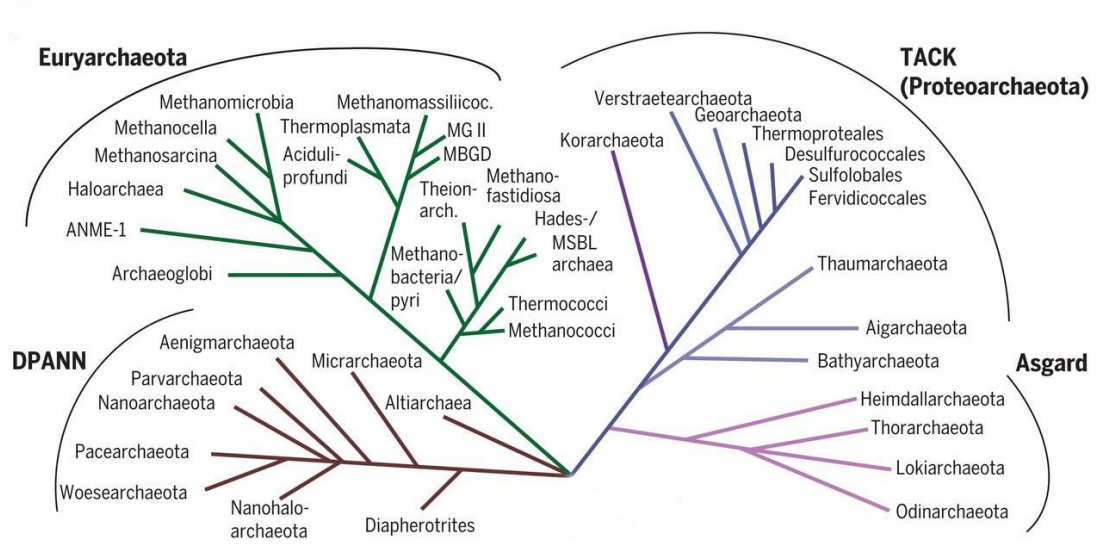


Fig. 2.1 Phylogenetic analysis indicates that the archaea includes four superphyla[17,18]

There are about 150,000 prokaryotic genomes in public databases, while only 2,694 archaea genomes are sequenced. Among them, the number of marine prokaryotic species genomes is 7,214. The distribution of species is quite biased because of abundance differences and limited methods for bacterial culture. A large proportion of of species (47%) belong to Proteobacteria, followed by Cyanobacteria/Melainabacteria and Bacteroidetes/Chlorobi groups (Fig. 2.2). At the genus level, the top 10 genera with decoded genomes are shown in Fig. 2.3, with genera of Prototheca and Vibrio at the top.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

We summarized the basic characteristics of published marine prokaryotic genomes, including assembled genome size, genomic GC ratio, and BUSCO.

**a**) The GC ratio of most marine prokaryotes range from 25% to 40% with genome sizes of ~1M-3M. Both attributes are smaller than those of terrestrial prokaryotes which have more diverse GC ratios and genome size distributions. A possible explanation for this observation is that, under ocean oligotrophic conditions, organisms have a higher chance of survival with fewer necessary metabolic genes and less DNA replication during cell division (Fig. 2.4)[19].

**b**) All genomes with ~3000-5000 genes are nearly complete with the ~90% BUSCO marker genes found. The inferior quality genomes generally come from metagenomics data which might not contain enough material, or enough sequencing data, to fully complete the genomes of component organisms (Fig. 2.5). Most of the complete genomes were obtained by sequencing and assembling of cultured strains, thus genome integrity is quite high. Recently, some genomes have been sequenced using single cell sequencing or have been assembled from metagenome sequencing data, but the genome assemblies were incomplete (completeness between 50% and 99%) and may contain some contamination. However, those species were previously unrepresented microorganisms due to the inability to culture them, so their genomes, however incomplete, still yielded new insights into marine bacteria and their metabolic pathways.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

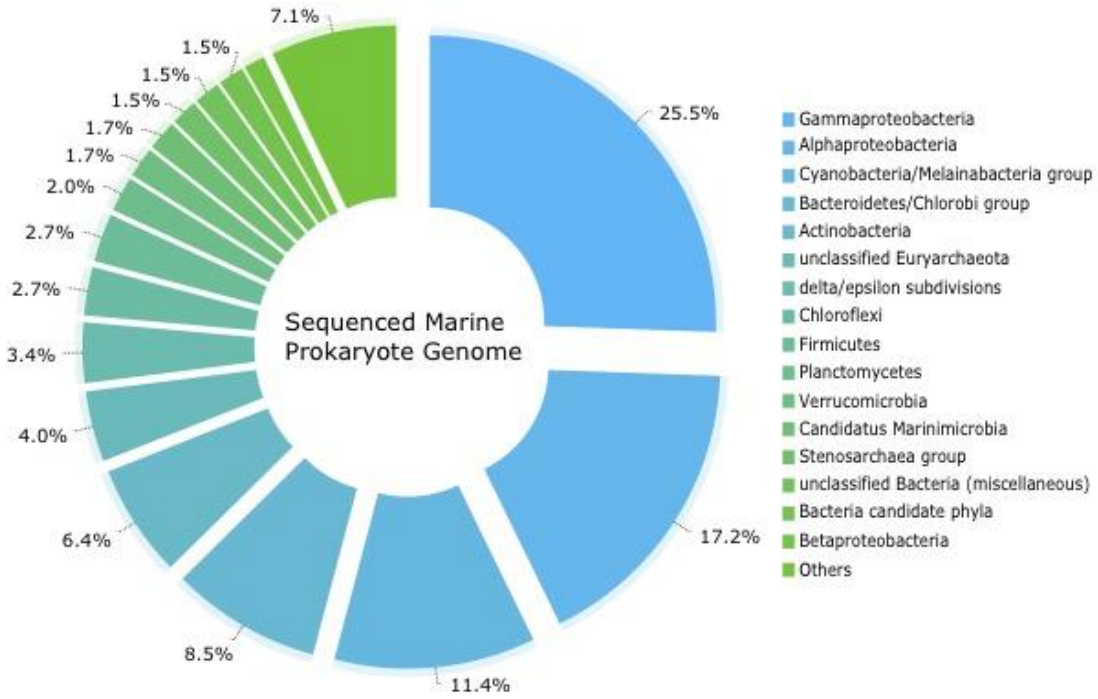


**Fig. 2.2 Categories of marine prokaryotes sequenced.**

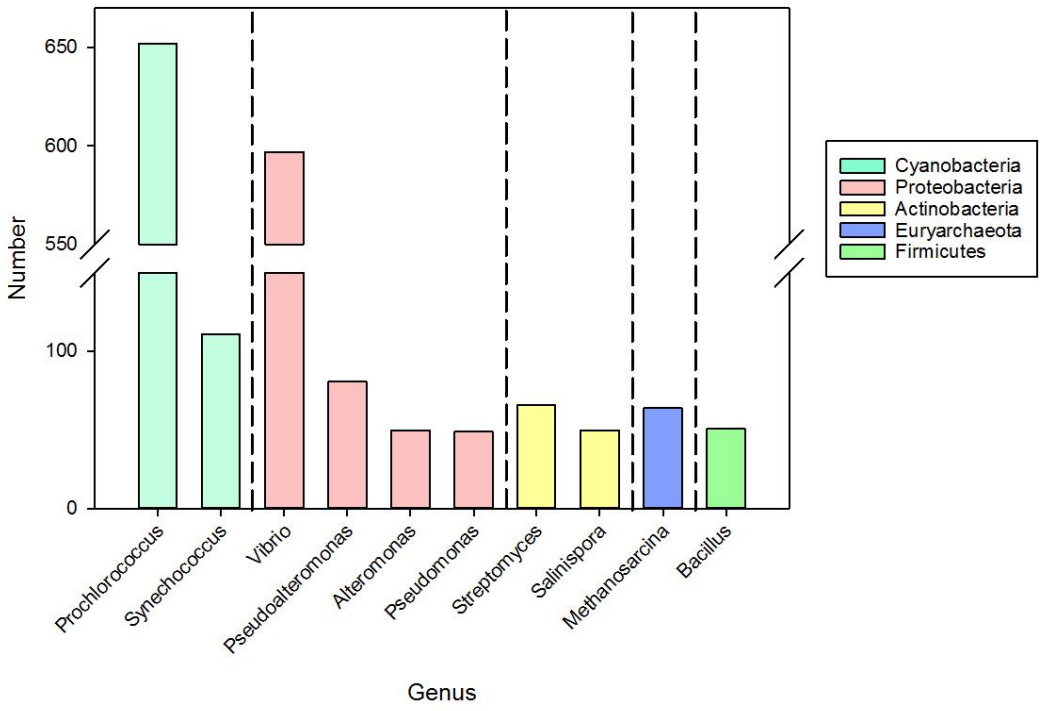


**Fig. 2.3 Sequenced marine prokaryotes in different genera.**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

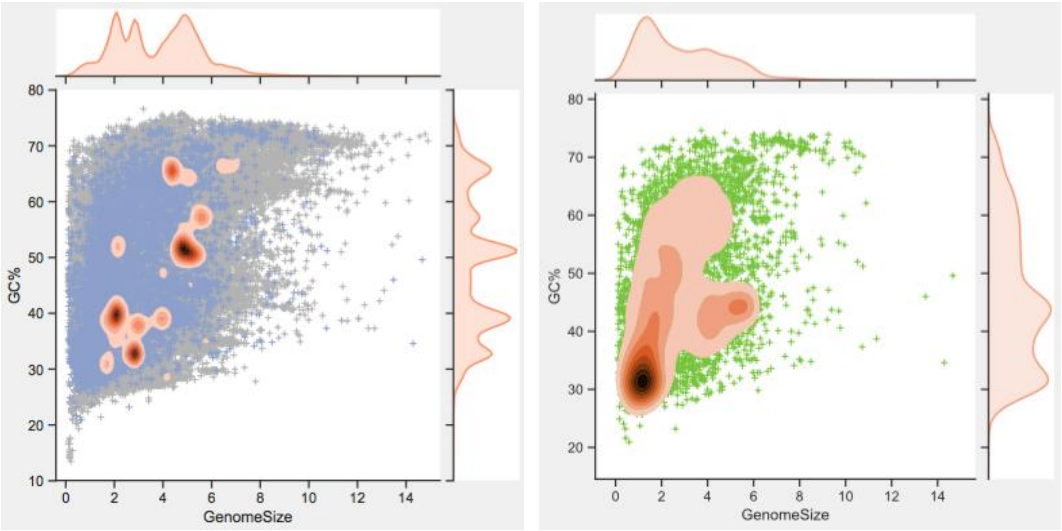


**Fig. 2.4 The distribution of GC content and genome size of all published prokaryotes (left) and marine prokaryotes (right).**

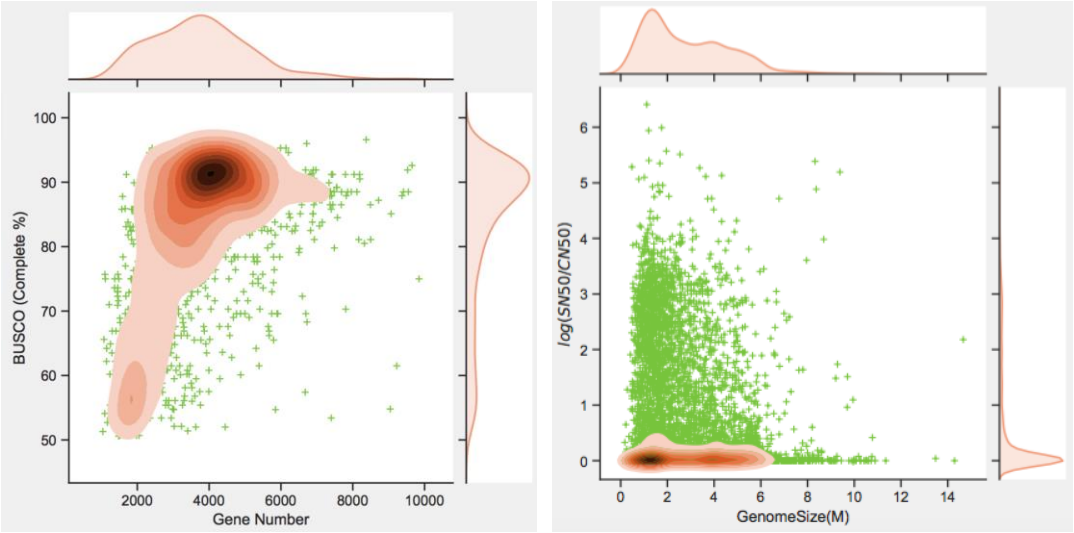


**Fig. 2.5 The distribution of BUSCO and gene number of marine prokaryotes, as well as scaffold N50/contig N50 and genome size.**

## 2.2 Marine metagenomics

**Box 2.1 The introduction to metagenomics**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

The interest in metagenomics arises from the large number of uncultivatable microorganisms that may exist in various environments. Before the concept of metagenomics existed, scientists found that 16S rRNA sequences extracted from environmental microorganisms do not belong to any known cultured microorganisms. This indicates that there are many environmental microorganisms which cannot be isolated and cultured. Studies have shown that culture-based methods capture less than 1% of microorganisms in environmental samples[20], necessitating the development of the metagenomics strategy. In 1998, Jo Handelsman first proposed the concept of metagenomics, in which DNA from the soil was extracted and directly sequenced in order to characterize component genomes[21]. In 2005, Kevin Chen and Lior Pachter further described metagenomics to be the application of modern genomics technologies to study microbial communities directly in their natural environment, without separation and culturing of individual species[22]. Unlike traditional microbial genome research, which relies on pure cultures, metagenomics research currently focuses on the genetic material of microbial community in specific environments, describing microbial composition in those environments. Rapid improvements in metagenomics have relied on the emergence of next generation high-throughput sequencing technologies. Compared to cloning-based metagenomics using Sanger sequencing, high-throughput sequencing-based metagenomics directly sequences the genetic material of all microorganisms in an environmental sample, instead of culturing by clone. This approach can inform species composition, genetic information, and functional diversity in environmental samples. It features high sensitivity, high throughput, high single-base resolution and no bias in organism representation. The rapid development of high-throughput sequencing technology has thus enabled advances in high-throughput sequencing-based metagenomics.

There are two sequencing strategies of metagenomics, including marker gene

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

sequencing and metagenome sequencing. A comparison of the two strategies is illustrated in Fig. 2.6. Overall, marker gene sequencing needs prior knowledge of the genomes to be sequenced, and the enrichment process (usually PCR amplification) can introduce bias, while whole genome or whole transcriptome sequencing approaches are not biased but require much more sequencing data resulting in high costs and difficulties in analyzing the data.

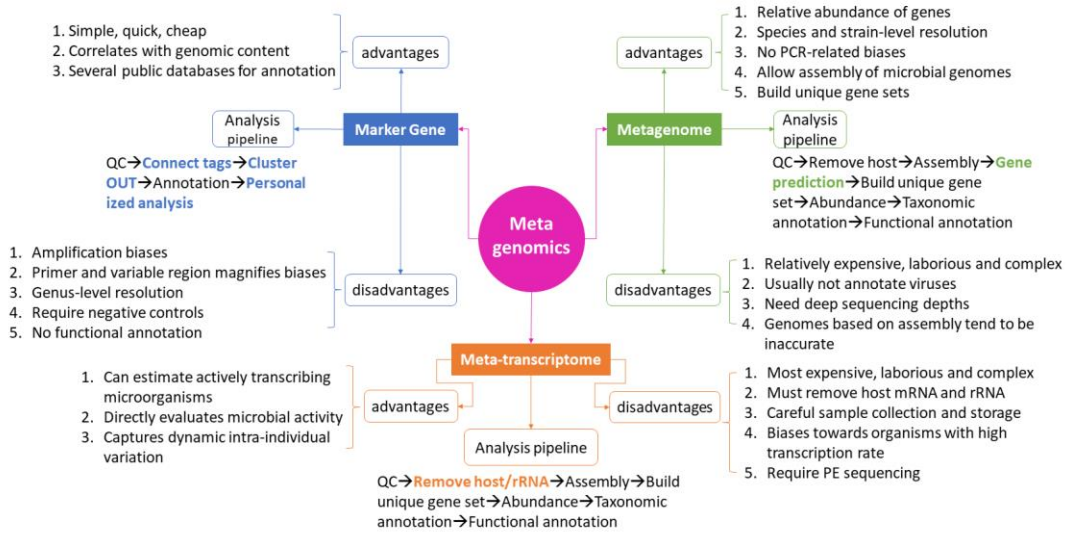


**Fig. 2.6 Comparison of marker gene-based and metagenome-based metagenomics.**

Currently, there are more than 100,000 metagenomics datasets in public databases. The sequencing strategy used has primarily been relatively low-cost amplicon sequencing (marker sequencing strategy). 91.2% of the datasets includes 16S rRNA genes for analyzing bacterial community structure, the 18S rRNA/ITS sequences for fungi, and some important functional genes for other species. The remaining datasets include around ~6,000 samples sequenced by whole genome metagenomics and ~3,000 by meta-transcriptomics (Fig. 2.7). These samples were collected from different environments in the ocean (Fig. 2.8), mainly seawater and sediment. In addition to the environmental samples, recent research has also been conducted on metagenomics of the symbiotic microorganisms of marine plants and animals (especially in corals and sponges). In order to study the community structure of symbiotic microorganisms and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

their relationship with a host, a large number of high-throughput sequencing dataset have been generated. For deep-sea research, because of the relative difficulty of sample collection and the fact that microorganisms in this environment are not easily cultured in the laboratory due to the effect of physical and chemical factors, high-throughput sequencing technology is currently the most common research method. Studies of this part of the ocean have focused on hydrothermal and cold springs.

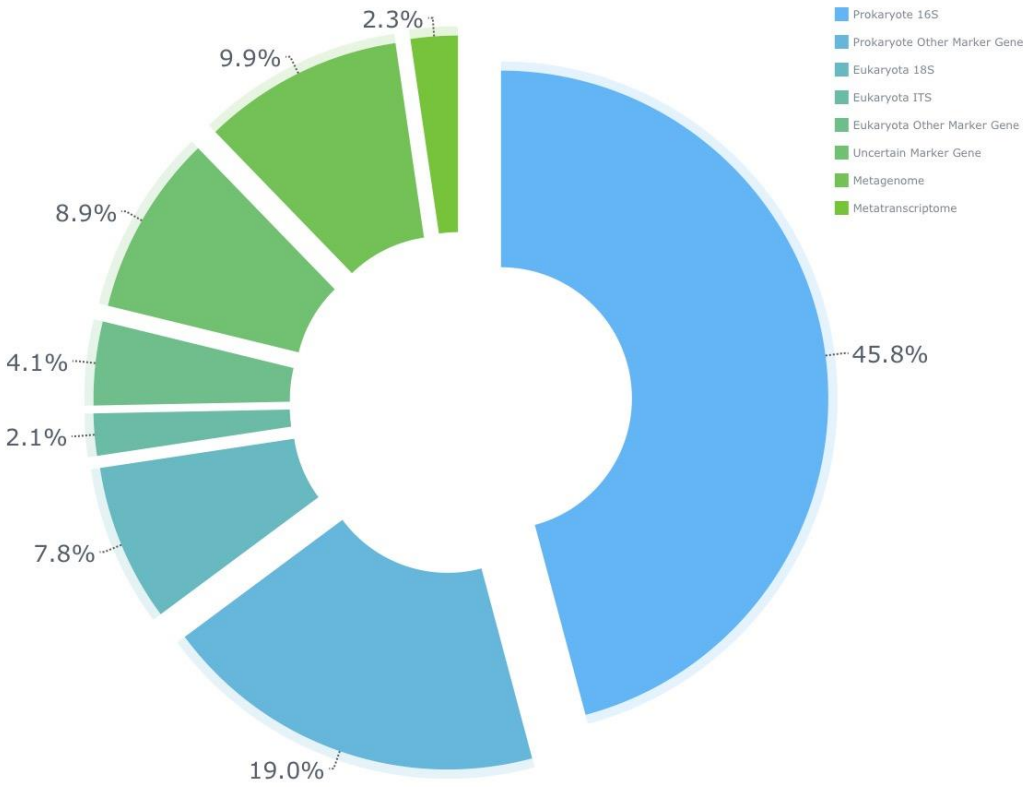


**Fig. 2.7 Proportion of metagenomics strategies of marker gene sequencing, metagenome sequencing and metatranscriptome sequencing represented in SRA.**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

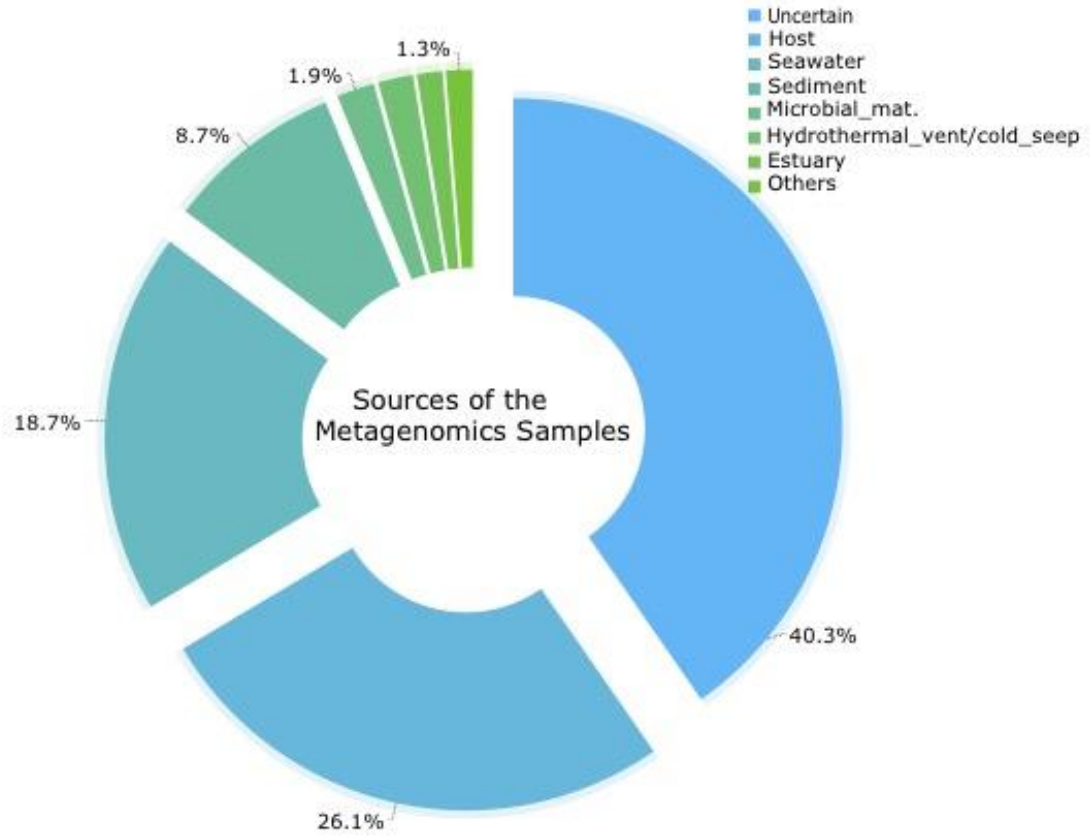


**Fig. 2.8 Sources of the metagenomics samples.**

**Box 2.2 Sequencing data required for marine metagenome sequencing**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

The amount of sequencing data is crucial for metagenomics, regardless of the sequencing strategy used. The amount of sequencing data in marker genes is usually less than 100 Mb. Determination of the amount of data for metagenome sequencing requires considering the balance between microbial diversity and sequencing cost of the sample. Most of the published data are fewer than 10 Gb per sample (Fig. 2.9). Transcriptome sequencing amounts are similar to metagenome sequencing, with data yields ranging from 1 to 20 Gb (Fig. 2.9).

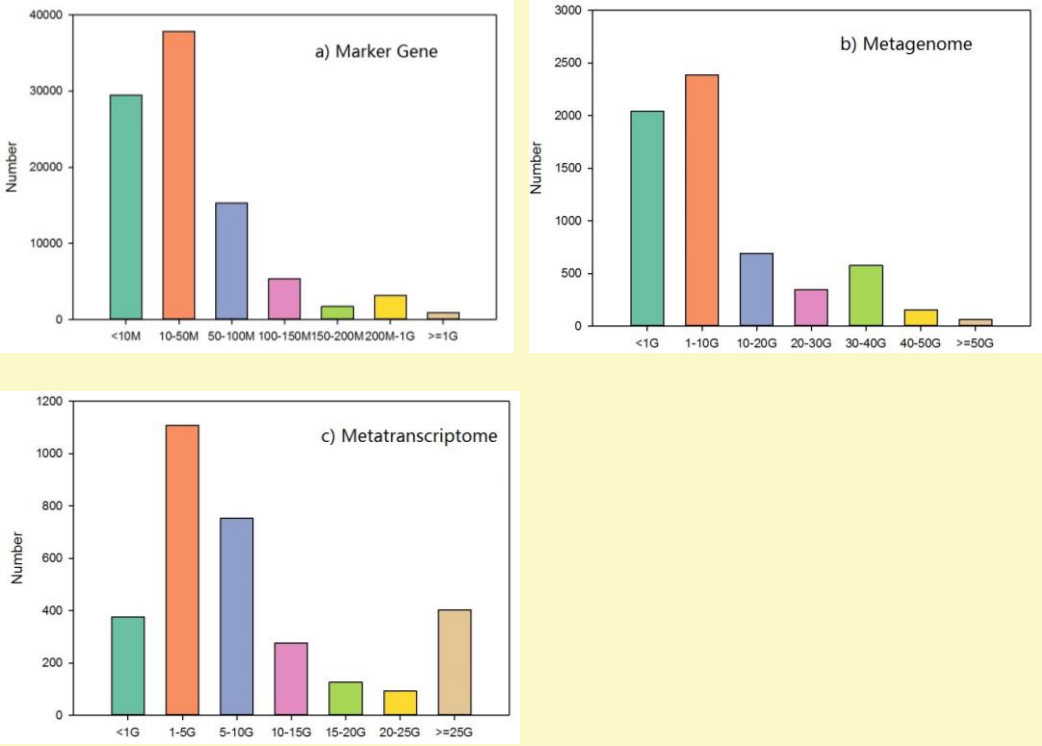


**Fig. 2.9 Amounts of sequencing data for marine metagenomics samples.**

The well-known projects of the marine metagenomics research include the Global Ocean Sampling (GOS), the Tara Ocean Expedition and the Ocean Sampling Day. Using the shotgun strategy, GOS constructed the first marine microbial gene set using Sanger sequencing[23], which is the first milestone in this area of research. It was the most important reference database prior to the widespread use of high-throughput sequencing. Tara Ocean provides the most complete collection of marine microbial genes[24] with a gene count of 40M, including data from GOS and other marine

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

metagenomic studies, as well as genes derived from some marine bacteria. This is also the first environmental microbial gene set created subsequent to human- and animal-related metagenomic gene sets. Based on the data published by Tara Ocean, scientists have done a variety of secondary analyses, such as the use of metagenomic data to reconstruct genomes of single bacteria, and thus found a large number of unknown bacteria spurring speculation about their potential functions. Ocean Sampling Day is a global collaborative project currently underway. On July 21st of each year, participating scientists collect ocean samples across the globe and sequence them, mainly using the 16S rRNA gene to construct a global marine microbial map.

### 2.3 Genomics of marine viruses

Although the first phage isolated from the marine environment was identified in 1955, the fact that marine viruses had an important impact in the ocean was not recognized until late in the 1980s. This realization was made largely as a result of achievements such as transmission electron microscopy, fluorescence microscopy and flow cytometry, which enabled the identification of viral particles directly in samples[25]. Despite their small size (~100 nm, 10-200 fg), marine viruses compose the second largest biomass in the ocean[26], with an average of $10^7$ virus-like particles per milliliter of surface seawater and the total estimated number of $10^{30}$ in the ocean[27].

Previous genomics research on marine viruses relied heavily on cultivation of the viruses, which is even more difficult than the cultivation of bacteria. There are not universal marker genes for viruses, but have been some development of gene markers for specific viral families. For example gp20 Portal protein of the head and gp23 Major capsid protein for T4-like have been widely used[28]. As today there are more than 250 viruses isolated from marine environments, including 16 palagibacter phages, more than 100 cyanobacterial and vibrio phages. Viral metagenomic datasets are promising for use in decoding additional viral genomes. 5,476 viral populations have been obtained by viral-fraction metagenomics from global oceans, while only 39 are

 doi:10.20944/preprints201812.0156.v1

successfully cultured[29]. In another study, over 125,000 partial DNA viral genomes are identified, including the largest phage yet identified, thus increasing the number of known viral genes by 16-fold. Combined, these results indicate that viral metagenomes will play an important role in future marine viral studies.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

# 3 Genomics of marine fungi

## 3.1 Basic introduction of marine fungi

Fungi are the second largest group of eukaryotes, after insects, and are widely distributed as parasites in animals and plants, and in the ecosystems of soil, fresh water, and the ocean. In 2017, it was estimated that there are between 2 and 3.8 million species of fungi[30], however only ~120,000 species have been identified so far[31].

Traditionally, marine fungi are classified according to habitats instead of taxonomic groups. They are classified as obligate marine fungi and facultative marine fungi[32], the obligate fungi are fungi that can only grow and form spores in oceans and estuaries, while facultative fungi are derived from freshwater or terrestrial sources. At present, high-throughput sequencing is widely used to define marine fungi. For example, a fungus that can maintain existence and metabolic activity in marine habitats through adaptation (ecological physiology), active metabolism (rRNA), gene expression (mRNA), catalytic function (proteome) or specific metabolites (metabolism) is considered to be a marine fungus[33].

**Box 3.1 True fungi and hidden fungi**

Based on recent research[34,35], we have summarized the latest taxonomic group of fungi, and divided them into 9 subkingdoms, 19 phyla, and 3 undetermined classifications. Seven true fungi groups named *Ascomycota, Basidiomycota, Chytridiomycota, Blastocladiomycota, Neocallimastigomycota, Zygomycotina,* and *Glomeromycota* are reclassified from previous studies[36]. Hidden fungi, named *Aphelidiomyceta* and *Rozellomyceta* (formerly known as *Cryptomycota*), were once considered as protozoan or protist, but currently they are considered to be fungi or sister groups to fungi[37,38]. Different from true fungi, they lack chitin cell walls at any stage of their life cycle, but possess Division 2 Chitin Synthases[39].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

**Box 3.2 The classification of marine fungi**

Since the first discovery of marine fungi in seabed wood in 1944[40], 2,369 species of marine fungi including 1,738 species from WoRMS database[41] have been collected. 83.58% of them are higher fungi, such as 1,832 species *of Ascomycota* and 148 species of *Basidiomycota*, as well as lower fungi, such as *Chytridiomycota* (45 species) and *Zygomycota* (21 species). *Microsporidia* (311 species) is parasitic in animal hosts in the ocean, causing many common diseases of crustaceans and fish[42]. Based on the phylogenetic relationship of marine fungi genera drawn from NCBI (Fig. 3.1), the yeasts *Sordariomycetes* and *Dothideomycetes* are the most important classes.

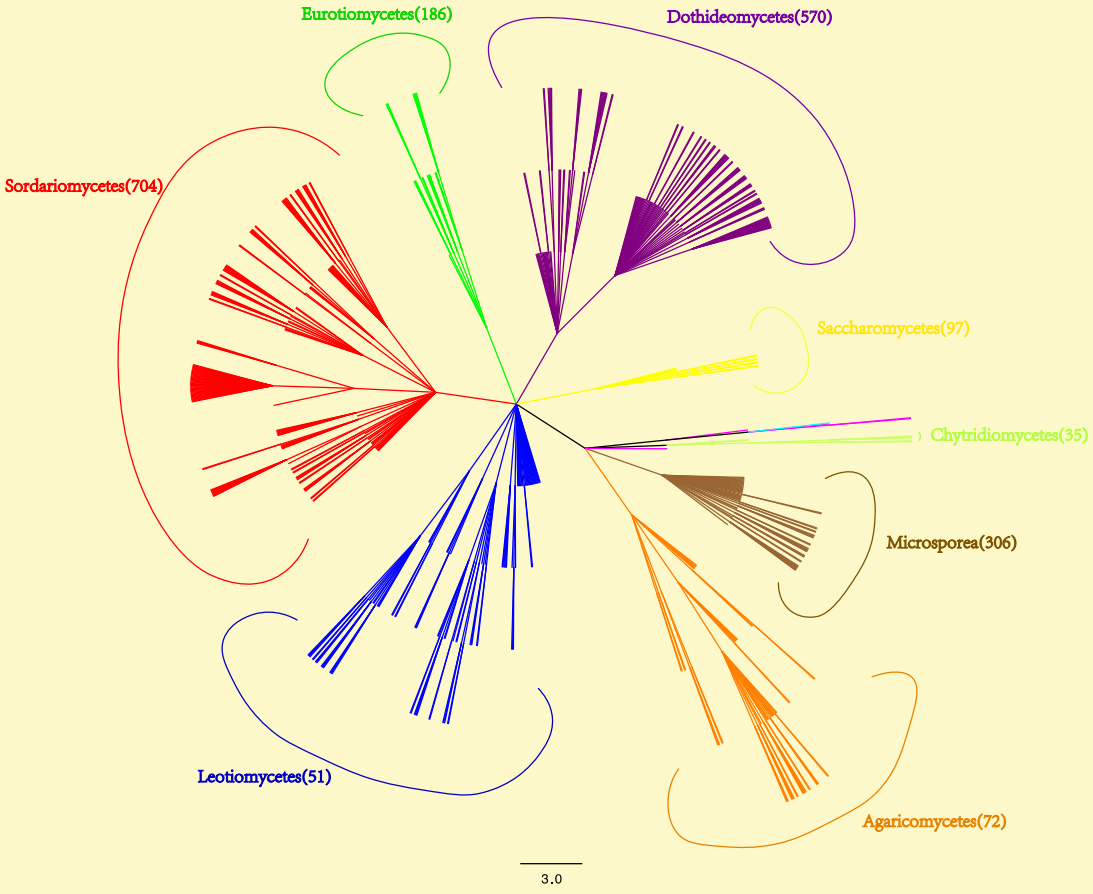


**Fig. 3.1 Phylogenetic relationship and number of marine fungi genera.**

**Box 3.3 Distribution characteristics of marine fungi**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

According to their nutrient source, marine fungi can be divided into three categories (Fig. 3.2). 1) Invertebrate symbiotic fungi are involved in the destruction of calcareous structure[43], opportunistic pathogens in corals[44,45] and encrusting sponges[46]. 2) Plant symbiotic fungi, which play an important role in the degradation of lignocellulosic fibers including 339 Manglicolous fungi[47] and 262 salt marsh plant symbiotic fungi, as well as 97 algae and seagrass symbiotic fungi. The biological interactions between fungi and algae hosts can promote growth, defense, development, and nutrient supply[48]. 3) 192 extreme ocean environmental marine fungi and mycoplankton. Fungi are highly adaptive microorganisms that can withstand high pressure, low temperature / high temperature and a high salinity environment[49].

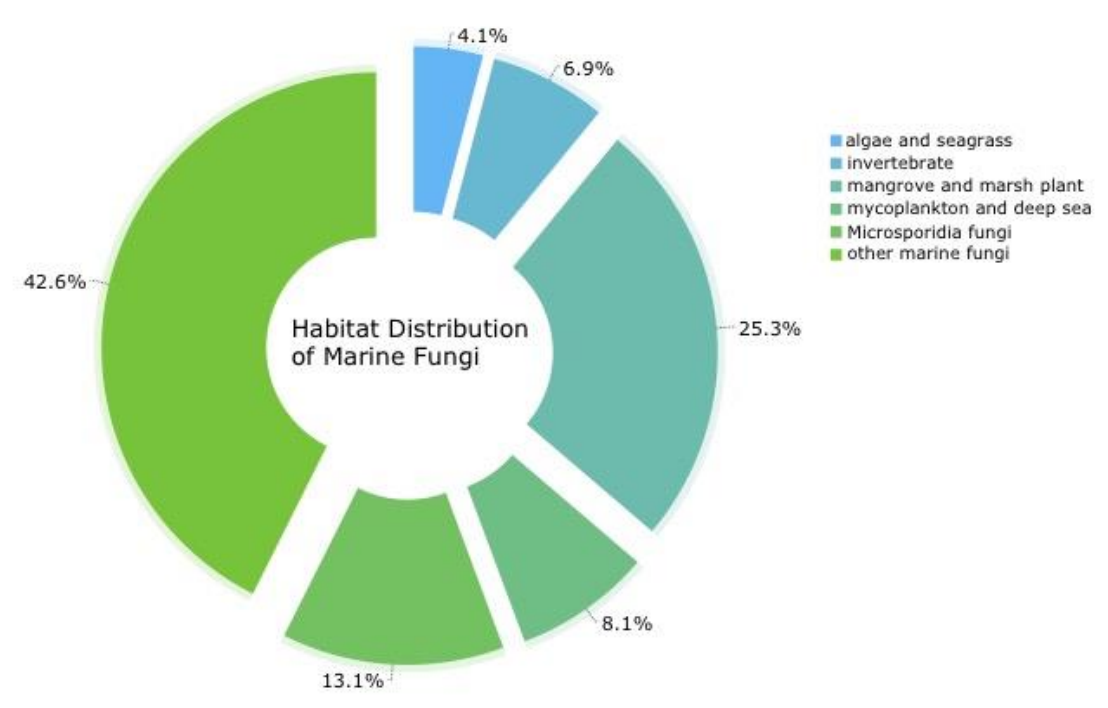


**Fig. 3.2 The source of nutrient growth of marine fungi.**

## 3.2 High-throughput sequencing for marine fungi

With the development of high-throughput sequencing technology, genome sequencing has become an important means of studying a species. In fungal genome research, there are international cooperative projects such as FungiDB[50] and "1000 Fungal Genome

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

Project"[51]. More than 1,555 fungal genome sequences have been published in NCBI by August 2018 in which 46 (~3%) are marine fungi, including 21 species of Ascomycota, 14 species of Basidiomycota, and 11 species of Microsporidia. The limitation in marine fungal genomes is caused by several reasons (Fig. 3.3).

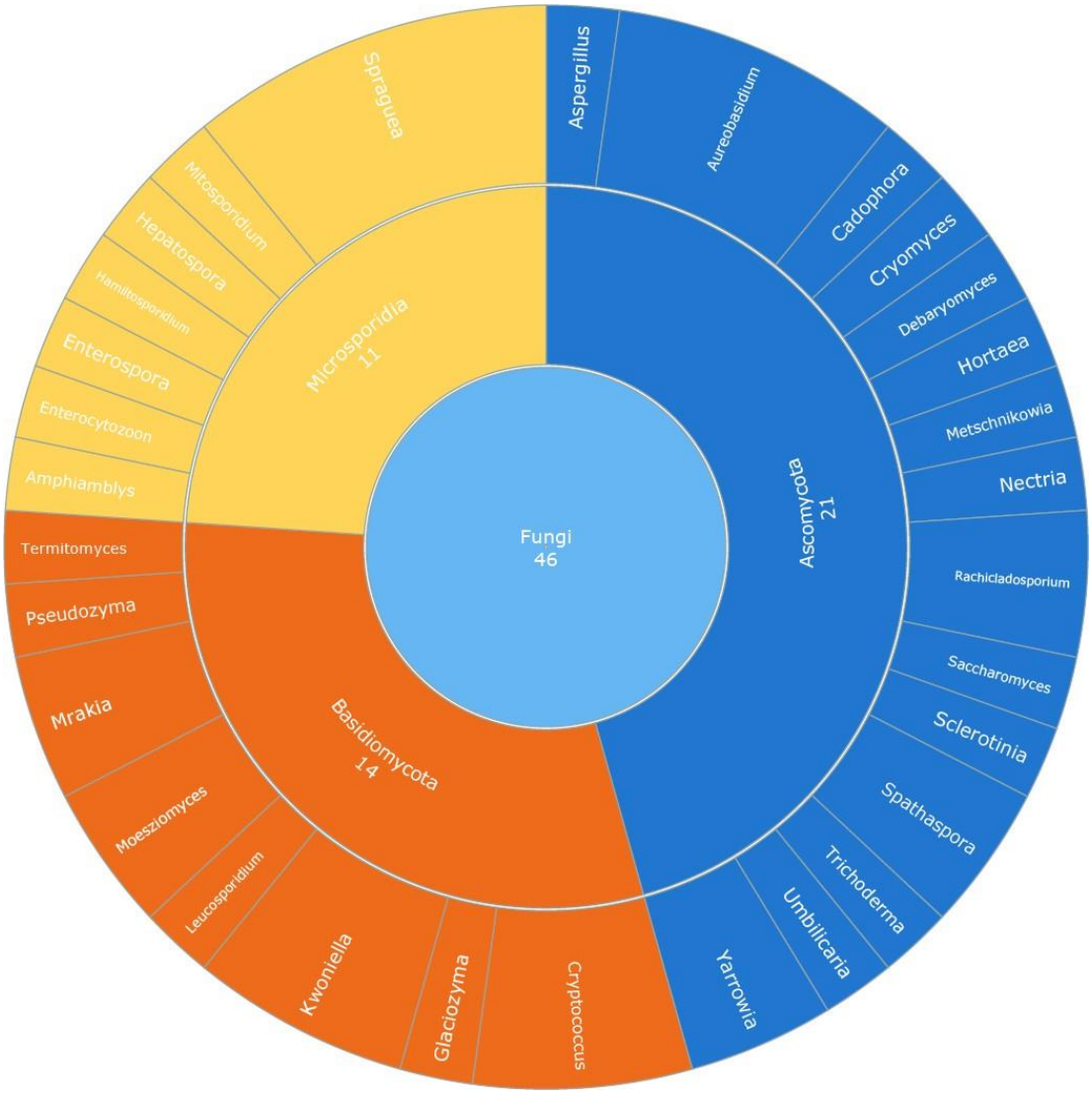


**Fig. 3.3 The classification of marine fungi.**

**Box 3.4 Difficulties of Marine genome sequence**

 doi:10.20944/preprints201812.0156.v1

There are difficulties in marine fungi isolation and genome sequence. a) The habitats of most marine fungi are difficult to microscopically separate and sample. Only about 470 ocean fungi from 244 genera can be isolated (1% of documented fungi)[52]. b) The culturing of fungal isolates from marine samples often leads to the recovery of non-fungal microbes. c) The majority of fungi harbor very high levels of cryptic diversity, making classifications based on observations of general morphological characteristics difficult and often misleading. d) DNA extracted by the metagenomics method had little fungal DNA because of the low abundance of fungal cells and the difficulty of extracting fungal DNA.

The most common application of high-throughput sequencing in marine fungal research is the identification and phylogenetic analysis of fungi using sequences of ribosomal DNA (rDNA) and internal transcribed spacers (ITS) from environmental DNA (eDNA). By August 2018, 2,399 marine fungal Sequence Read Archive (SRA) data were published in NCBI, mainly from invertebrate sources, of which 1,106 were from a coral environment (Fig. 3.4).

The development of single molecule sequencing technology has enabled the full-length sequencing of 18S rDNA and its application in the evolutionary analysis of fungi[53], but there are few studies on marine fungi. Current metagenomics and macro-transcriptomics as well as high-throughput, culture-based methods as leading-edge tools will enable comprehensive analysis to understand marine fungi more comprehensively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

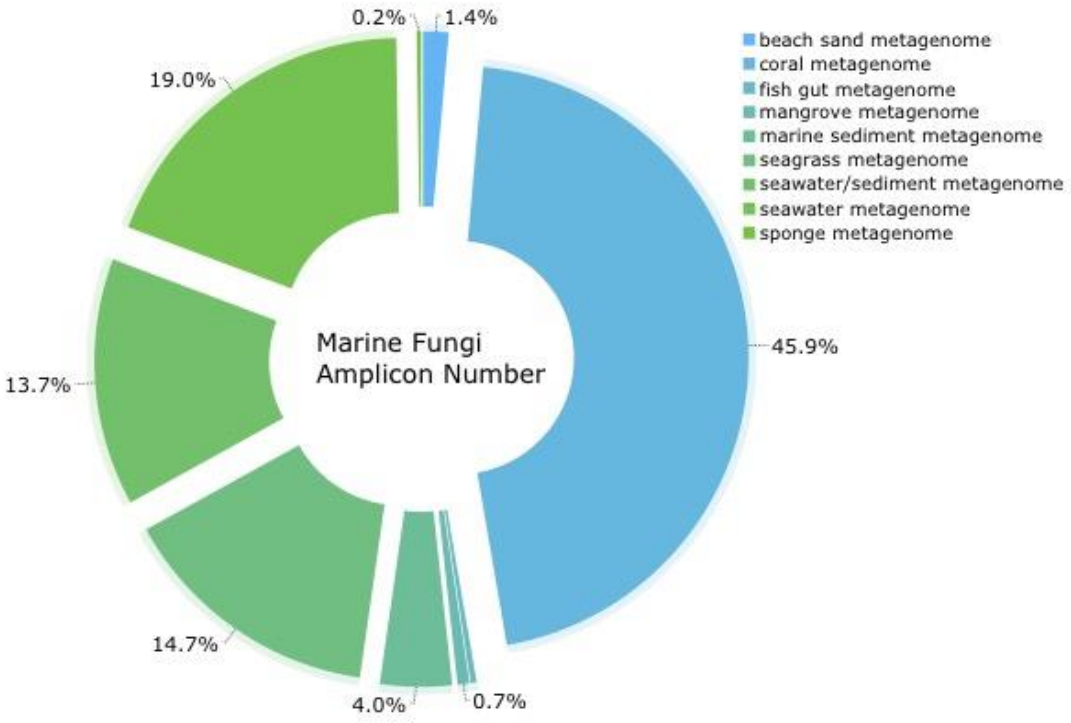


**Fig. 3.4 A summary of marine fungi in Sequence Read Archive (SRA) data**

**Box 3.4 Biological questions about fungi**

Fungal DNA detected in the marine environment spans multiple forms and lineages, including chytrids, filamentous hyphae, and multicellular, and the Dikarya yeast form appears to dominate the known marine fungal diversity. Marine fungi also play different ecological roles in marine ecosystems and are often associated with parasitic marine animals, plants and algae. The main research focuses on marine fungi include 1) the evolutionary relationship between marine fungi and terrestrial fungi; 2) genomics and proteomics studies of symbiotic mechanisms and bioactive molecule production by marine fungi and endophytic fungi in seaweed, seagrass, mangroves and marine invertebrates, especially coral and sponge-related fungi; 3) the study of the true diversity of lower fungi and fungal parasites in the marine environment; and 4) the study of the nutritional functions and evolutionary mechanisms in fungal analogs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

# 4 Genomics of marine algae and plants

Marine plants and algae are the basis of the marine ecosystem. They provide oxygen, foods and habitats for animals. There are more than 49,000 algae and 200 marine plants documented. The marine plants mainly belong to Magnoliopsida, Monocotyledoneae, Polypodiopsida and Bryopsida. Algae is a complicated group including both Eukaryotic and Prokaryotic algae. Plantae algae and Chromista algae comprise ~91% of all algae species. To date, only three genomes of marine plants and 115 genomes of algae have been reported. For algae, the sizes of sequenced genomes range from 0.56 Mb to 1,500 Mb and half of them belong to Chlorophyta.

## 4.1 Algae genome

**Box 4.1. Brief introduction for algae**

Algae are mainly plant, or plantlike, marine non-vascular organisms from several phyla, classes and families containing chloroplasts. They can be classified into green algae, red algae and brown algae based on the pigments contained in their cells, and further classified into macroalgae and microalgae according to their body size. Macroalgae are macroscopic species of great economic importance as many of them serve as foods. Microalgae have potential to be used as biofuel and to reduce water pollution with their ability to accumulate heavy metals, pesticides, organic and inorganic toxic substances and radioactive material in their cell bodies[54,55].

There are 115 algal genomes sequenced to date, with about half distributed across Chlorophyta (Fig. 4.1). The assembled genome sizes of published algal species range from 0.56 Mb (*Cryptomonas paramecium*) to 1,500 Mb (*Breviolum minutum*). Published genome research mainly focuses on algal evolution history, environmental adaption, biomass accumulation and economic and ecological roles. 30 papers focus on the terrestrial evolution, multicellular evolution and distinct genome structure of algae;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

16 papers are about environmental adaptation to ultra-cold, hot, unstable, high-salt and/or high-iron environment; and 15 papers studied the development of renewable fuel (Table 4.1).

There are ~2Tb of sequenced genomic data for algae species. However, current genome research concerning algae still faces many challenges. An obvious problem is the completeness of assemblies. The shortest scaffold N50 of algal genome assemblies is 409bp (*Euglena gracilis*) resulting in fractional gene models. Additionally, half of these assemblies have a completeness less than 80% based on BUSCO evaluation of the available algal assemblies. **Box 4.2** describes the current challenges in algal genome assembly. New technologies such as optical mapping, 10X genomics, single molecule real-time sequencing and Hi-C might be used to address these challenges and in enable higher quality genome assemblies.

**Box 4.2. Challenges for algae genomic research**

1. Some algae have symbiotic relationships with other organisms, such as fungi, in order to form lichen or live in marine coral cells. The symbiont brings many difficulties to isolate, culture and acquire the algae samples because of the sample contamination.

2. Some algae have challenging physical characteristics which make DNA extraction difficult. For example, the body of coral algae is highly calcified and the amount of DNA extracted from such tissues is very small or, when extracted, the quality is very poor, making it difficult to perform long molecular sequencing.

3. Algae genomes usually contain a high proportion of repetitive sequences. Current technology cannot successfully overcome the assembly challenges caused by high repeat contents or gigantic genome sizes. For example, it is reported that Dinoflagellates contain large amount of DNA with an estimated genome size ranging from 3 G to 215 Gb[56].

**Table 4.1 Summary of the research focus of the published algae genomes.**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| Research focus | Species |
|---|---|
| Evolution | *Cyanidioschyzon merolae* 10D*; Picoeukaryotes Micromonas; Auxenochlorella protothecoides; Prototheca wickerhamii; Cymbomonas tetramitiformis; Cyanophora paradoxa; Chlamydomonas reinhardtii; Ostreococcus lucimarinus; Micromonas; Bathycoccus prasinos; Porphyridium purpureuml; Volvox carteri; Klebsormidium flaccidum; Lotharella oceanica; Hemiselmis andersenii; Chroomonas mesostigmatica; Chlorella vulgaris; Raphidocelis subcapitata; Tetrabaena socialis; Coccomyxa sp.* C-169*;Euglena gracilis;Gonium pectorale;Phaeodactylum tricornutum;Symbiodinium kawagutii; Chlorella variabilis* NC64A*;Chondrus crispus;Chondrus crispus; Ectocarpus siliculosus; Gonium pectorale* |
| Environmental adaptation | *Dunaliella salina* Strain CCAP19/18*; Thalassiosira Pseudonana*， *Chlamydomonas eustigma; Fragilariopsis cylindrus; Galdieria sulphuraria; Picochlorum sp.; Chrysochromulina tobin: Ostreococcus tauri; Micromonas; Heterococcus sp.* DN1*; Symbiodinium goreaui; Chlorella variabilis; Pyropia yezoensis; Coccomyxa subellipsoidea; Thalassiosira Pseudonana; Thalassiosira oceanica* CCMP1005*; Picochlorum* SENEW3*;* |
| Ecological role | *Aureococcus anophagefferens; Pseudo-nitzschia multiseries; Symbiodinium minutum;* |
| Biomass accumulation | *Tetradesmus obliquus; Nannochloropsis gaditana; Tetradesmus; Dunaliella salina obliquus* UTEX 393*; Botryococcus braunii; Scenedesmus* |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| | |
|---|---|
| | *obliquus Strain* DOE0152z*; Chlorella vulgaris; Micractinium conductrix; Parachlorella kessleri; Nannochloropsis; Picochlorum soloecismus; Chlorella protothecoides; Monoraphidium neglectum; Tetradesmus obliquus* UTEX 393*; Botryococcus braunii;* |
| Economic value | *Porphyra umbilicalis; Chlorella sorokiniana; Cladosiphon okamuranus;Haematococcus pluvialis;Saccharina japonica* |

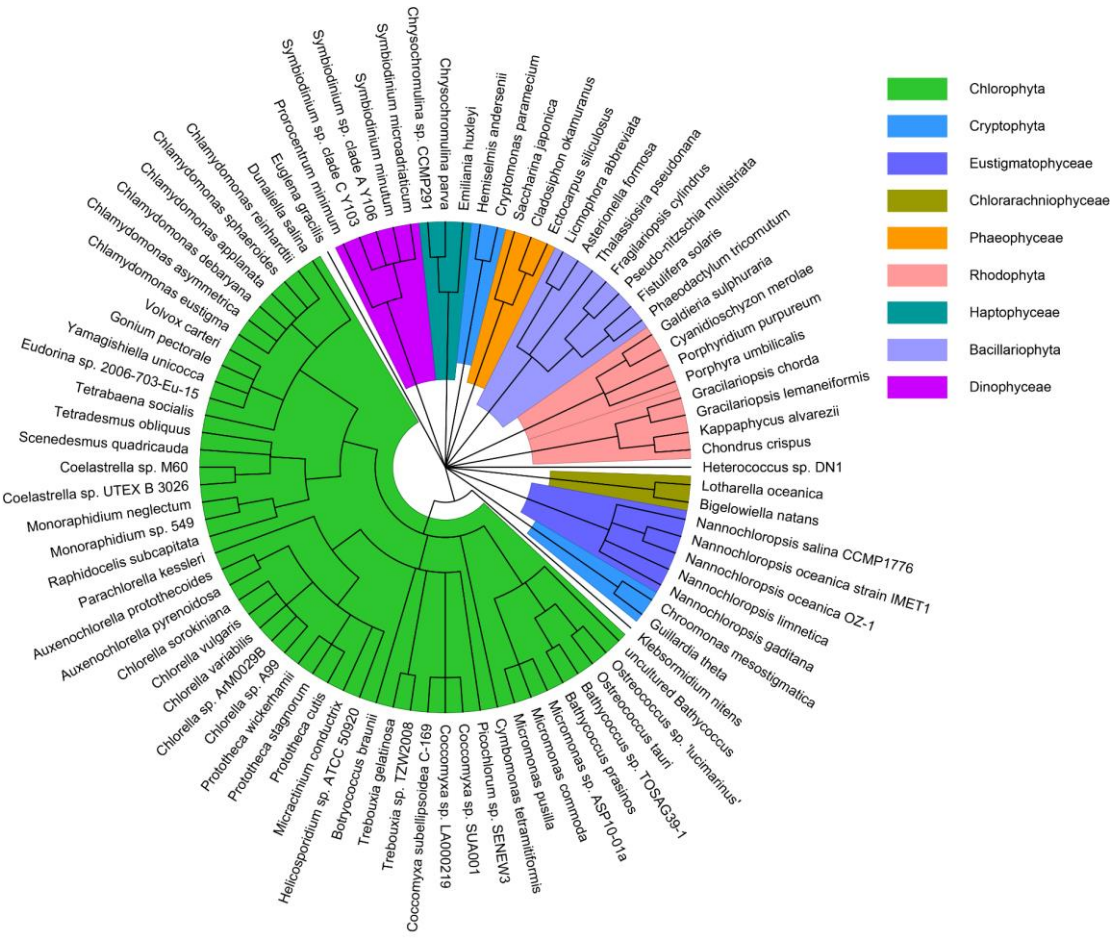


**Fig. 4.1 Phylogenetic structure of sequenced algal genomes.** The tree is constructed from the NCBI common tree. Highlights with same color are distributed in same phylogenetic clade. The largest highlight in green is chlorophyta.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

## 4.2 Genomics of marine plants

Marine plants are usually referred to as groups of embryophytes colonizing the oceans[57] or intertidal areas, which evolved from their terrestrial ancestors to adapt to the ocean environment. Marine plants include seagrasses, mangroves, mosses and ferns. Seagrasses are the only flowering plants that can live underwater. They are a kind of polyphyletic assemblage of monocots, mainly from *alismatales*[57]. Compared to sea grasses, mangroves usually colonize intertidal areas, and most mangroves belong to 3 orders of the *magnoliopsida* class (*malpighiales*: red mangroves, *lamiales*: black mangroves and *myrtales*: white mangroves) (Fig. 4.2). Mangroves have adapted to environments with high salinity, strong UV light, hypoxia and anoxic conditions of waterlogged muds[58], as well as tides which affect salt equilibrium regulation[59,60]. Thus, genomics of mangroves are especially useful for exploring mechanisms of adaptation[61-63].

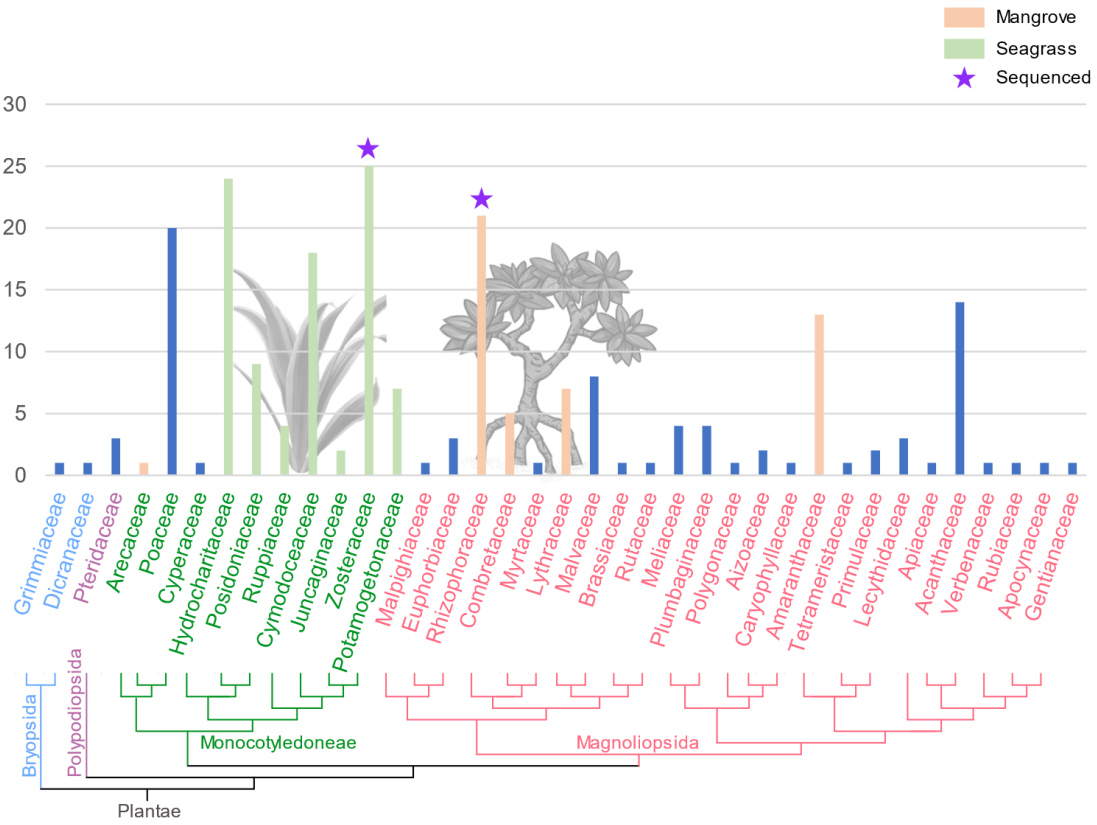


**Fig. 4.2 Numbers of marine plant species in different groups.**

 doi:10.20944/preprints201812.0156.v1

To date, only two seagrass genomes (*Zostera marina* and *Zostera muelleri*) [64,65] and one mangrove genome (*Rhizophora apiculata*)[66] have been reported. The genome of *Zostera marina*, the first marine angiosperm to be sequenced, reveals mechanisms of adaptation to the marine environment, including gene gain and loss events[65]. Similar gene gain and loss events were observed in the *Zostera muelleri* genome, suggesting that these may comprise the major genetic changes required for marine adaptation[64]. The only published mangrove genome is *Rhizophora apiculate*[66]. The combination of whole genome duplication (WGD) in *R. apiculate* and paleogeographic events – rising sea levels submerging the angiosperms living at the margins of rainforests – resulted in rapid adaption to this environment as well species diversification. Duplicated genes made it possible to develop specialized functions required for thriving in this new and challenging environment[66]. As a result, seagrasses and mangroves are the ideal model for the study of convergent evolution and the return of angiosperm plants to the marine environment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

# 5 Marine Invertebrates

## 5.1 Global diversity and phylogeny

Marine invertebrates are highly diverse, occupying 31 of 33 invertebrate phyla with as many as 172,021 accepted species. They represent over 95% of all invertebrates. The most dominant species were distributed in phyla of Arthropoda (32.25%) and Mollusca (27.52%) (Table 5.1).

**Table 5.1 The accepted species numbers of marine invertebrates***

| Phylum | # of class | # of order | # of family | # of genus | # of species | # of sequenced |
|---|---|---|---|---|---|---|
| Arthropoda | 16 | 79 | 1,114 | 8,753 | 55,472 | 17 |
| Mollusca | 8 | 53 | 564 | 4,952 | 47,345 | 15 |
| Annelida | 2 | 15 | 107 | 1,552 | 12,906 | 2 |
| Cnidaria | 7 | 25 | 349 | 1,725 | 11,601 | 13 |
| Platyhelminthes | 6 | 40 | 310 | 2,398 | 11,475 | 2 |
| Porifera | 5 | 35 | 141 | 696 | 8,747 | 1 |
| Echinodermata | 5 | 41 | 201 | 1,288 | 7,354 | 11 |
| Nematoda | 3 | 18 | 103 | 774 | 6,140 | 0 |
| Bryozoa | 2 | 3 | 199 | 887 | 6,131 | 0 |
| Nemertea | 4 | 2 | 42 | 301 | 1,320 | 1 |
| Gastrotricha | 1 | 2 | 14 | 46 | 497 | 0 |
| Acanthocephala | 2 | 4 | 16 | 91 | 492 | 0 |
| Xenacoelomorpha | 1 | 2 | 19 | 113 | 451 | 0 |
| Brachiopoda | 3 | 5 | 30 | 122 | 420 | 1 |
| Tardigrada | 2 | 3 | 12 | 51 | 207 | 0 |
| Ctenophora | 2 | 9 | 31 | 53 | 204 | 2 |
| Entoprocta | 1 | 1 | 4 | 12 | 190 | 0 |
| Kinorhyncha | 1 | 2 | 10 | 21 | 188 | 0 |
| Sipuncula | 2 | 3 | 6 | 18 | 156 | 0 |
| Rotifera | 2 | 3 | 17 | 33 | 147 | 0 |
| Chaetognatha | 1 | 2 | 9 | 26 | 131 | 0 |
| Hemichordata | 2 | 3 | 6 | 24 | 130 | 2 |
| Rhombozoa | 2 | 2 | 3 | 9 | 122 | 0 |
| Gnathostomulida | 1 | 2 | 12 | 27 | 101 | 0 |
| Loricifera | 1 | 1 | 2 | 8 | 28 | 0 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| Orthonectida | 1 | 1 | 2 | 5 | 25 | 1 |
| Priapula | 1 | 4 | 5 | 7 | 22 | 1 |
| Phoronida | 1 | 1 | 1 | 2 | 11 | 1 |
| Nematomorpha | 1 | 1 | 1 | 1 | 5 | 0 |
| Cycliophora | 1 | 1 | 1 | 1 | 2 | 0 |
| Placozoa | 1 | 1 | 1 | 1 | 1 | 1 |
| Micrognathozoa | 0 | 0 | 0 | 0 | 0 | 0 |
| Onychophora | 0 | 0 | 0 | 0 | 0 | 0 |
| **Total** | **88** | **364** | **3,332** | **23,997** | **172,021** | **71** |

* the classification of marine invertebrates was adapted from a previous study[67]; the species numbers were adapted from the WoRMS database (up to 2018.9.1); the numbers of sequenced species were collected from the NCBI taxonomy database.

## 5.2 Genomics of marine invertebrates

Only 0.041% (71 of 172,021) species of all marine invertebrates have been sequenced, mainly distributed in phyla of Arthropoda (17), Mollusca (15), Cnidaria (13), and Echinodermata (11). Marine invertebrates selected for whole genome sequencing are often prioritized according to: their economic value as seafood, such as the shrimps *Marsupenaeus japonicus* and *Penaeus monodon*[68]; their potential value in medicine and biomaterials; and their critical ecological value, like the reef-building coral *Acropora digitifera*[69]. However, there are some major challenges that present barriers to additional marine invertebrate genome sequencing (**Box 5.1**).

Genomic studies in marine invertebrates have focused on investigating a variety of evolutionary, biological, and ecological questions. Specifically, due to the fundamental phylogenetic role that this large group plays, evolutionary questions including the origin of multicellularity and early-animal evolution, bilateral emergence, and nervous and immune system development, have received lots of interest from investigators. Similarly, biological issues such as adaptation to extreme environments, biorhythms, shell formation, and longevity, alongside ecological balance, breeding and improvement, have been characterized in diverse phyla in marine invertebrates. The published studies are summarized in Table 5.2. Overall, although advances have been

 doi:10.20944/preprints201812.0156.v1

made, more accessible sequenced genomes and other omics data for marine invertebrates are required for more comprehensive studies to be performed.

**Box 5.1 Barriers in marine invertebrate genome sequencing**

i) Ecological niche: Many marine invertebrates live far from land, or live at great depths, leading to sampling difficulties and degradation of DNA.

ii) Symbiosis: organisms that live closely together make it difficult to isolate single-organism samples resulting in DNA contamination. Some examples are sponges (which contain microorganisms comprising up to 35% of the total biomass)[70] and coral reefs (symbiosis between coral and algae)[71].

iii) Heterozygosity levels: many invertebrates of have genomes with high heterozygosity, such as oyster (*Crassostrea gigas*) 2.3%[72] and sea urchin (*Strongylocentrotus purpuratus*) 4~5%[73], which increase the complexity of *de novo* genome assembly.

iv) Abstraction from human life and perceived economic value: excluding shrimp, crab, and shellfish most marine invertebrates cannot be used as a food source, such as Annelida and Platyhelminthes, so the value of their genomic data is not immediately obvious.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

**Table 5.2 Hotspots and major advances in marine invertebrate research.**

| Fields/Hotspots | Advances |
|---|---|
| **Evolution of animal development** | |
| Evolution of early-animal development mechanisms | As the oldest surviving metazoan phyletic lineage, sponges share key adhesion and signaling genes with the 'true' animals or eumetazoans[74-76]. |
| | The first, and still the only, available complete genome of sponge, *Amphimedon queenslandica*, reveals remarkable similarity to eumetazoan genomes, suggesting most gene families of true animals were already present in the last common ancestor of all animals[77]. |
| Evolution of the Bilateria | The sea anemone *Nematostella vectensis* genome displays high complexity with a gene repertoire, exon-intron structure, and large-scale gene linkage more similar to vertebrates than to some bilaterians such as flies, suggesting that the genome of the eumetazoan ancestor was similarly complex[78]. |
| | The marine Mollusca, *Lottia gigantean*, genome displays more similarities to some invertebrate deuterostome genomes than to other protostome genomes, contributing novel genes to the bilaterian ancestor background and revealing lineage-specific genome evolution[79]. |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| | |
|---|---|
| Evolution of the nervous system | The genome of the demosponge, *Tethya wilhelma,* has been sequenced and the protein repertoire - in the context of genes mediating neural-like functions - was examined. Although the comprehensive analysis is still pending, those data will shed light on the evolution of nervous system development in metazoans[80]. |
| | Despite the morphological similarity of neuromuscular junctions in bilaterians and hydra, several of the key genes required for this junction in bilaterians are absent from the hydra genome[81]. |
| | Genetic programs that are homologous to three vertebrate signaling centers - the anterior neural ridge, zona limitans intrathalamica and isthmic organizer - are reported in the hemichordate (acorn worm), *Saccoglossus kowalevskii*[82]. |
| | Massive expansions in two gene families previously thought to be uniquely enlarged in vertebrates - the protocadherins that regulate neuronal development and the C2H2 superfamily - are reported in the *Octopus bimaculoides* genome, corresponding to the octopus' complex nervous system[83]. |
| Evolution of Immunological Function | Due to the integral role of antimicrobial peptides (AMPs) in the innate immune system, a variety of marine invertebrate genomic studies have focused on the discovery and characterization of AMPs, such as novel AMPs reported in the green sea urchin, *Strongylocentrotus droebachiensis*[84], and in the oyster, *Crassostrea gigas*[85]. |
| | Although commonalities of innate defenses have been emphasized in invertebrates and vertebrates, ample evidence from complete genome studies suggests that novel immune capabilities exist among different phyla[86-88]. For instance, comparison of the genomes of the two Diptera, *Anopheles gambiae* and *Drosophila melanogaster*, which diverged about 250 million years ago, reveals surprisingly large differences in immunity-related genes[88]. |

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| | |
|---|---|
| | Molecular studies of sponges[89], cnidarians[90], shrimp[91], and sea urchins[92] have identified surprisingly diversified immune molecules. |
| | Symbionts may play more of a role in marine invertebrates' internal defense than generally appreciated[93-95]. |
| **Biological process** | |
| Adaptation to environment (deep sea, tidal zones, hydrothermal vents) | Deep sea scale worms adopted two strategies of adaptation to hypoxia in habitats: rapid evolution of tetra-domain hemoglobin in Branchipolynoe and high expression of single-domain hemoglobin in Lepidonotopodium sp.[96]. |
| | The adaptability to environment by marine molluscs, which can be obviously divided into characteristics specific to physical environment[97-99] and features for feeding strategies[100-104], arises from the expansion of specific gene families, organ-specific proteins, or ministrant bacteria communities[97,104]. |
| Biological clock | Modulation of alternative splicing is a mechanism for natural adaptation in circadian timing[105]. |
| Shell formation mechanism | The pearl oyster was sequenced and studied to uncover the molecular mechanisms that underlie the formation of shells[106]. |
| | The process of shell formation involves attribution of cells and exosomes as well as frequent duplication of genes[98,99]. |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| | |
|---|---|
| Longevity | Scientists have detected amino acid residues specific for a longevity group in sea urchin based on whole genome sequencing[107]. |
| **Ecological environment** | |
| Ecological environment | The diverse communities of symbiotic organisms extend sponges' metabolic capabilities by mediating processes such as photosynthesis, carbon, and nitrogen cycling[70,108-112]. |
| | The Acropora genome provides crucial insights into the molecular basis of coral symbiosis and responses to environmental changes[69]. The innate complex immune repertoire of corals allows them to better cope with environmental stress and pathogens[113]. However, it is hard to determine how well the stony coral, Acropora, genome reflects general coral traits or to what extent it diverged from other coral genomes[113]. |
| | Comparison of the coral *Stylophora pistillata* genome to the coral *Acropora digitifera* genome reveals that the core set of conserved proteins is enriched in functions relating to cnidarian-dinoflagellate symbiosis. Independent, uneven expansions of genes involved in algal symbiosis, innate immunity, and stress response are identified in both species, demonstrating strikingly disparate coral genomes[114]. |
| | Researchers sequenced the whole genome of *Acanthaster planci* species from Australia and Okinawa[115] and revealed key genes and a biological network regulation model in species-specific communication factors that are associated with their activity of aggregation on corals. |

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1

| **Molecular breeding and improvement** | |
|---|---|
| Molecular breeding and improvement | The genome information of two economically valuable penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*, was used to identify key genes that are important to their body plans, providing valuable resources for the study of selective breeding and some plastic biological characteristics of penaeid shrimps, including molting, lobstering, brooding eggs and sensitization in humans[68]. |

 doi:10.20944/preprints201812.0156.v1



# 6 Fish genomes

## 6.1 Brief introduction of fish

Fish comprise more than half of all vertebrate species and have been adapted to a variety of marine and freshwater habitats. Their genome evolution and diversification are important subjects for the understanding of vertebrate evolution. With the development of sequencing technology, the number of fish species with assembled, draft whole genome sequences that are available online is rising rapidly. Currently, there are more than 170 fish species with genomes in public databases, with 50 fish genome papers published in recent years (Fig. 6.1). In addition, about 2,618 fish mitochondrial genomes have been sequenced and deposited in the Mitofish website[116], and about 563 fishes as well as 37,118 transcriptomes have been sequenced. The sequenced fish species come from different orders; Gadiformes (28 species), Perciformes (22 species), and Cyprinodontiformes (15 species) are the top 3 orders with sequenced species including many major food fishes.

Presently, research on fish genomes is focused on a wide variety of topics (Fig. 6.2). Because of their wide distribution and potential to be an evolutionary model for vertebrates, studies on fish evolution, and especially evolutionary adaptation, have been abundant. Common topics includes viviparity, adaptive transition to land, adaptation to extreme environments, and convergent adaptation to their ecological niches. Meanwhile, commercial interests and the desire to understand fish themselves have motivated studies within fish genome, such as immunity and sex determination. Here, we will summarize the highlights of this research.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



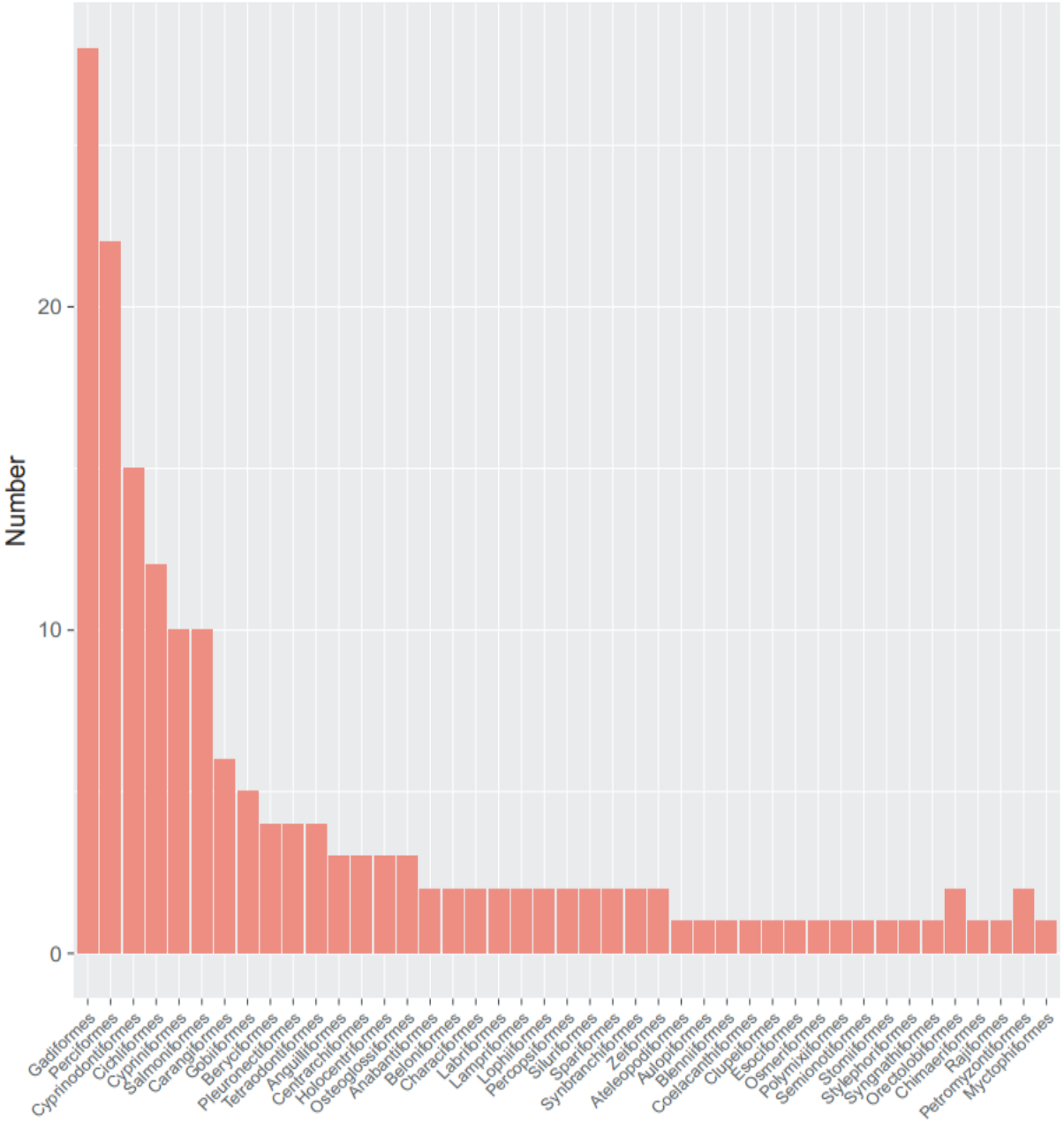


**Fig. 6.1 The numbers of sequenced fish species in different orders.**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1




**Fig. 6.2 The research focuses of published fish genomes.**

## 6.2 Research focuses of fish genomics

### 6.2.1 Viviparity in teleost

Some of the adaptive traits in fish have attracted the attention of researchers for years. Viviparity, which means retention and growth of the fertilized egg within maternal body, is one of them. It is believed that viviparity is an adaptation to a wide variety of environments. The family of Poeciliidae in Cyprinodontiformes, which contains many viviparous species whose reproduction mode is characterized as continuously complex in maternal provisioning[117], has been studied for their morphological and phenotypic adaptions[118]. Though the first species in Poeciliidae (*Xiphophorus maculatus*) with a sequenced whole genome revealed the stability of percomorph karyotypes and positive selection on viviparity-related genes in Poeciliidae[119], genomic data from viviparous teleosts

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



are almost nonexistent.

### 6.2.2 Water-to-land transition

The evolution in the water-to-land transition is another interesting topic in fish. As coelacanths and lungfishes belong to the primitive fish lineages of sarcopterygians, they are considered essential to elucidate this mode of evolution. Whole genome sequencing of coelacanth enabled the interpretation of its genomic status revealing a significantly slower rate in the evolution of protein-coding genes compared to other vertebrates[120] and the traits of both fish and tetrapods in genes[120-123]. However, the whole genome sequencing of lungfish is currently impractical due to its extraordinarily large genome (40 to 130 Gb), so the previous analyses on lungfish genes are generally based on transcriptomic data[121]. Although these studies have shed light on many aspects of lungfish biology[124,125], there are still questions cannot be solved without genomic data, underscoring the need for improved sequencing and assembly techniques.

### 6.2.3 Adaptation to extreme cold

Adaptation to extreme environments is another important aspect of evolution. The condition of freezing Antarctic water is lethal to most species, whereas the ancestors of the notothenioids were able to make the dramatic change required to live in an extreme cold environment and populate empty niches. Subsequently, these fish dominate the diversity of Antarctic fish in the Southern Ocean. Due to their habitats in polar regions and adaptive radiation in extreme cold, notothenioids are considered to be ideal models for research on evolution and development in extreme environments[126]. Information about genes related to this adaptation has been acquired from cDNA libraries, and genomic information is still limited due to the difficulties in breeding and raising notothenioids. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



recent whole genome sequencing of *Notothenia coriiceps* has shed light on the adaptive evolution of notothenioids[127]. Although high-throughput sequencing techniques are advancing, which help to reduce sequencing costs and improve the quality of genome sequencing, many mysteries await us in the genome of notothenioids.

### 6.2.4 Convergent evolution toward adaptation to darkness

Convergent evolution is the independent evolution of similar features in different lineages, leading to analogous structures similar in form or function but not present in the last common ancestor of these lineages. There are about 150 cave-living fish species[128] in which common features have been observed, such as loss of eyes and pigmentation, presenting a great opportunity to study convergent evolution. The *Astyanax* genome assembly[129] has filled in a missing piece in non-cyprinid teleost genomics, and is an important step in the research on adaptive evolution on a genomic scale. Subsequent comparisons with the transcriptomic evidence from golden-line barbel (genus *Sinocyclocheilus*) reveal different paths of convergent evolution in cave phenotypes[130] between *Astyanax* and *Sinocyclocheilus*, such as the regression in retina[131,132].

### 6.2.5 Fish disease and immunity

With the accumulation of reported fish diseases caused by virus and protozoan parasites, immunology in fish has been considered an important subject. As a representative population of lower vertebrates, fishes are important models in comparative immunology providing a link to the early evolution of vertebrates. Meanwhile, from the perspective of commercial interests, research on the immunity of fish will also improve the selection of disease-resistant breeds. As a result of genome-scale scanning of immune genes in zebrafish[133], numerous immune-relevant genes for both innate and adaptive

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



immunity have been in various fish genomes[134-136], thus describing an immune system in fish with only a slight difference to mammals (Fig. 6.3). With the rapid growth of genomic information and the application of new genome engineering technologies in fish, more advances in fish immunology will be achieved in the near future.

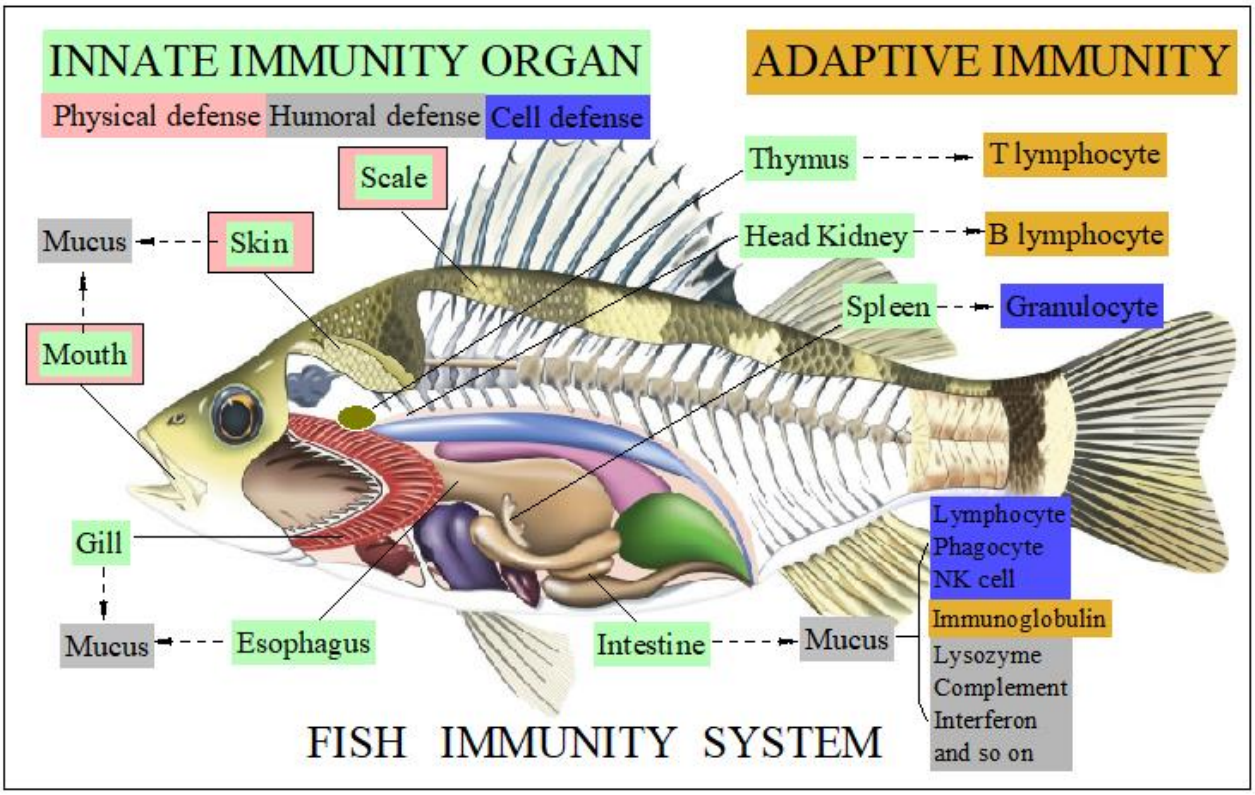


**Fig. 6.3 Innate immunity and adaptive organ of fish.**

### 6.2.6 Sex determination

Sex determination is always a focus in developmental research, and the mechanisms vary among different species (

Table 6.1). Moreover, the sex of fish can be influenced by a series of environmental elements or it may alternate in adults[137-141]. The study of fish sex determination has been facilitated by whole genome sequencing in the recent years. Furthermore, the alternation of sex-determined genes in *Acanthopagrus schlegelii* exposed its mechanism of hermaphroditism[142].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



**Table 6.1 Sex chromosome type of several representative fishes**

| Bigger size | Species | Sex chromosome type |
|---|---|---|
| Female | *Anguilla Anguilla* | -- |
| Male | *Channaargus Cantor* | XX-XY |
| -- | *Coilia brachygnathus* | ZZ-ZO |
| Female | *Cynoglossus semilaevis* | ZW-ZZ |
| Female | *Cyprinus carpio* | XX-XY |
| Female | *Dicentrarchus labrax* | XX-XY |
| Female | *Hippoglossus hippoglossus* | XX-XY |
| Male | *Ictalurus punctatus* | XX-XY |
| Male | *Lepomis macrochirus* | XX-XY |
| -- | *Loporinus elongatus* | ZZ-ZW |
| Male | *Odontobutis obscura* | -- |
| Female | *Oncorhynchus kisutch* | XX-XY |
| Female | *Oncorhynchus mykiss* | XX-XY |
| Female | *Oncorhynchus tshawytscha* | XX-XY |
| Male | *Oreochromis aureus* | ZW-ZZ |
| Male | *Oreochromis niloticus* | XX-XY |
| Female | *Paralichthys lethostigma* | XX-XY |
| Female | *Paralichthys olivaceus* | XX-XY |
| Male | *Parapercis snyderi* | -- |
| Male | *Pelteobagrus fulvidraco* | XX-XY |
| Female | *Perca flavescens* | XX-XY |
| -- | *Poecilia sphenops* | ZZ-ZW |
| Male | *Pseudobagrus ussuriensis* | XX-XY |
| -- | *Pseudatocindus tetensis* | XX-XY |
| Female | *Puntius gonionotus* | -- |
| Female | *Salmo salar* | XX-XY |
| Male | *Scarus ferrugineus* | -- |
| Female | *Scatophagus argus* | -- |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



### 6.2.7 Metamorphosis

Metamorphosis is a biological process involving a conspicuous and relatively abrupt change of the organism's body structure during its growth. Changes of body structure can be observed in most teleosts. Furthermore, in some fish, a significant relationship between morphosis and environment can be observed, for example, the change of eye symmetry during the development of flatfish. The evolution and function of genes related to metamorphosis have been researched for years[143-145], and the recent genome-wide SNP identification and the construction of genetic maps are important attempts to answer this question from a genome-level view[146,147]. What's more, the genome and transcriptome of Japanese flounder has revealed the important role of thyroid hormone and retinoic acid signaling, as well as phototransduction pathways, providing new insights into flatfish asymmetry[148].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



# 7 Genomics of marine tetrapods

## 7.1 Brief introduction and genomes

Marine tetrapods, distinguished from terrestrial tetrapods, belong to ~27 orders, consisting of ~1,130 species/subspecies (Fig. 7.1)[41] which are highly evolved, living within marine habitats and are usually at the top of the marine food web. Their adaptation to marine ecosystems occurred independently from various terrestrial or freshwater ancestors and was accompanied by major morphological transformations. They acquire most or all of their nourishment from the marine environment and spend majority of their time in the water but come back to land for mating, breeding, molting and so on.

Despite their evolutionary importance, there were only ~48 species whose whole genome has been sequenced and assembled - notably fewer than sequenced terrestrial tetrapod species (~270) - indicating that the progress of marine tetrapod genomics has lagged significantly compared to terrestrial tetrapods (Fig. 7.2). Based on available assembled genomes, marine tetrapods have relatively stable genomes with the length of ~1.2 Gb for marine birds and ~2.5 Gb for marine mammals. Although there are several recent high-quality reference genomes completed using a new sequencing strategy, for example, 10X Genomics, only a few of them have been assembled at the chromosome level. Meanwhile, in some marine tetrapod families, no species have been sequenced or only one low-quality reference genome has been made available.

## 7.2 Current status of marine tetrapod genomes

The first marine tetrapod genome, bottlenose dolphin, was sequenced in 2008 followed by polar bear in 2012[149]. In 2013, one functionally extinct species, the Yangtze River dolphin, and one extremely

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



endangered species, green sea turtle, were sequenced mainly because of the urgent need for conservation and the importance of their phylogenetic position with birds[150,151]. In 2014, research on 48 bird genomes accelerated the sequencing process of marine tetrapods, during which 12 seabird genomes were published. These studies resolved the debate about the evolution of early birds, detailing the history of bird genome evolution from the whole genome perspective; demonstrating the convergent evolution of birds both in morphology and behavior; and explored protein-coding genes and their regulatory elements for some important traits[152,153]. As for marine mammals, investigations have mainly focused on the evolution of sensory genes, marine adaptation mechanisms and dynamic population sizes. For example, olfactory receptor genes underwent an obvious decline in Yangtze River dolphin[150], minke whale[154] and Antarctic minke whale[155]. Some taste- and vision-related genes were also underrepresented, possibly non-functional or lost in these genomes. In particular, hearing and vocalization genes associated with echolocation appear to be under significant accelerated evolution compared to terrestrial mammals. A lot of genes or gene families, such as *PRDX*, *OGT*, *SLC16A1, PRDX*, *OGT*, *SLC16A1, DAG1* and *BTN1A1,* have experienced positive selection or expansion to meet the challenges of hypoxia and oxidative stress, osmotic stress, deep diving[154], and a cold environment[149]. Moreover, in the bowhead whale genome, gene gain or loss is related to DNA repair, cell-cycle regulation, cancer, and aging suggesting the affected genes might be associated with longevity[156]. Population demography is another issue that has been extensively studied[150,154,157-159]. These studies can estimate the population size of species responding to geological and climate changes or threats over time, which may help to understand the more appropriate environment for a given population and provide important information to assist with species conservation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



### 7.3 Conservation of marine tetrapods using genomics

According to the International Union for Conservation of Nature and Natural Resources – IUCN2018, ~261 species (as far as we know), are critically endangered (CR), endangered (EN), vulnerable (VU) and near threatened (NT), which should cause us concern. The most urgent action to be undertaken is increasing the awareness of species protection, including restricting anthropogenic and barbaric capture of marine species, reducing destruction of their habitats, minimizing unintentional pollution and conducting extensive field search and rescue operations. In addition to these efforts, genomic data can also provide the molecular clues necessary for effective management and conservation of marine biodiversity. From genomic data, we can investigate the variations and adaptive mechanisms involved in resilience to environment stressors (climate change) and common threats (pollution), which can direct the setting of conservation priorities and strategies for restoration. Also, genomics can uncover the genetic characteristics how organisms respond to some biological threats, such as diseases and toxins. For example, a comparison of genome-wide differences dolphins that died as a result of harmful algal blooms and those that survived showed a number of changes in allele frequencies and helped identify candidate genes for resistance to algal brevetoxin[160,161]. Furthermore, genomic data can also reveal features which do not change during selective pressure in some endangered species. For instance, loss of genetic diversity in immune-related genes and enrichment of deleterious mutations in toxin degradation genes contribute to the major genetic defects of the crested ibis and other endangered species[162].

In summary, more whole genome data, and high quality data, should be acquired because they are critical to understanding the genetic basis of disease response and adaptive changes within species. Coupled with evolutionary, transcriptional, and population genetics studies whole genome data can

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



create a strong foundation for protecting threatened and endangered species. We plan to sequence these species to in three stages: first, critically endangered (CR) and endangered (EN) species; second, vulnerable (VU) and near threatened (NT) species; and, finally, at least one species in each remaining group. After obtaining these data, we will be able to find additional genetic features governing fundamental and evolutionary processes of marine tetrapods, in particular: convergent evolution, for all or each order of marine tetrapods (which can also be investigated at a macro level), chromosome evolution, sex determination, limbs development, viviparity and oviparity.

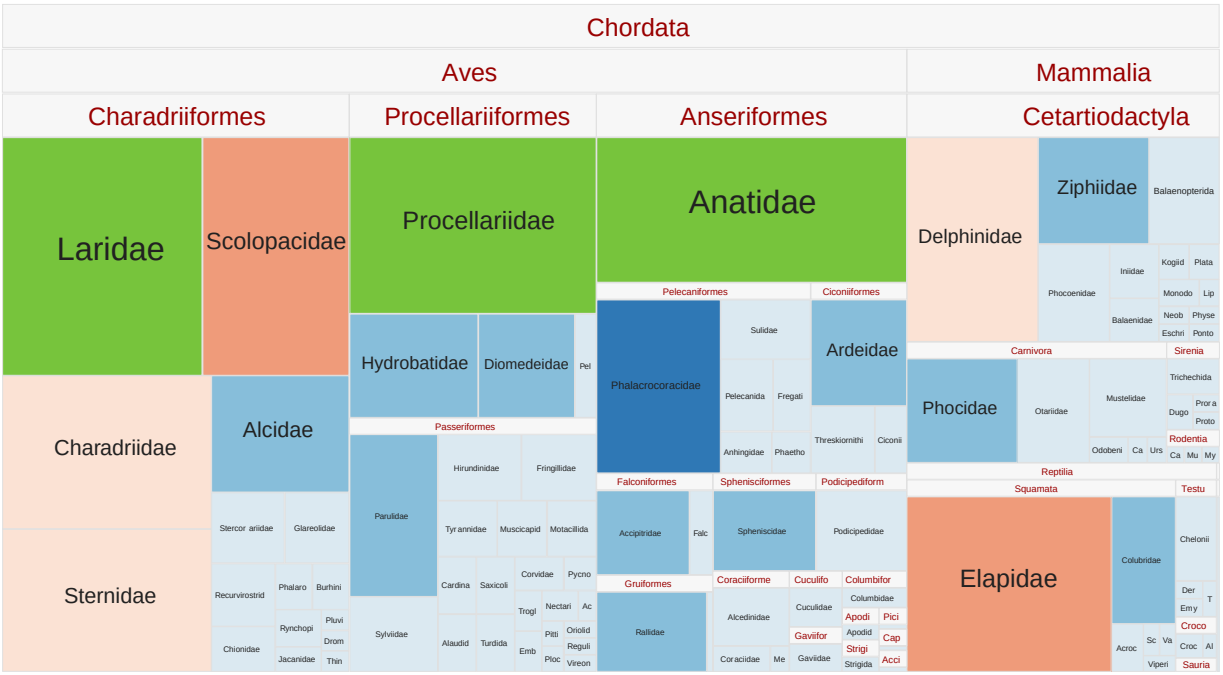


**Fig. 7.1 Species numbers of marine tetrapods in each family.** The sizes of the rectangles represent distributions of species numbers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



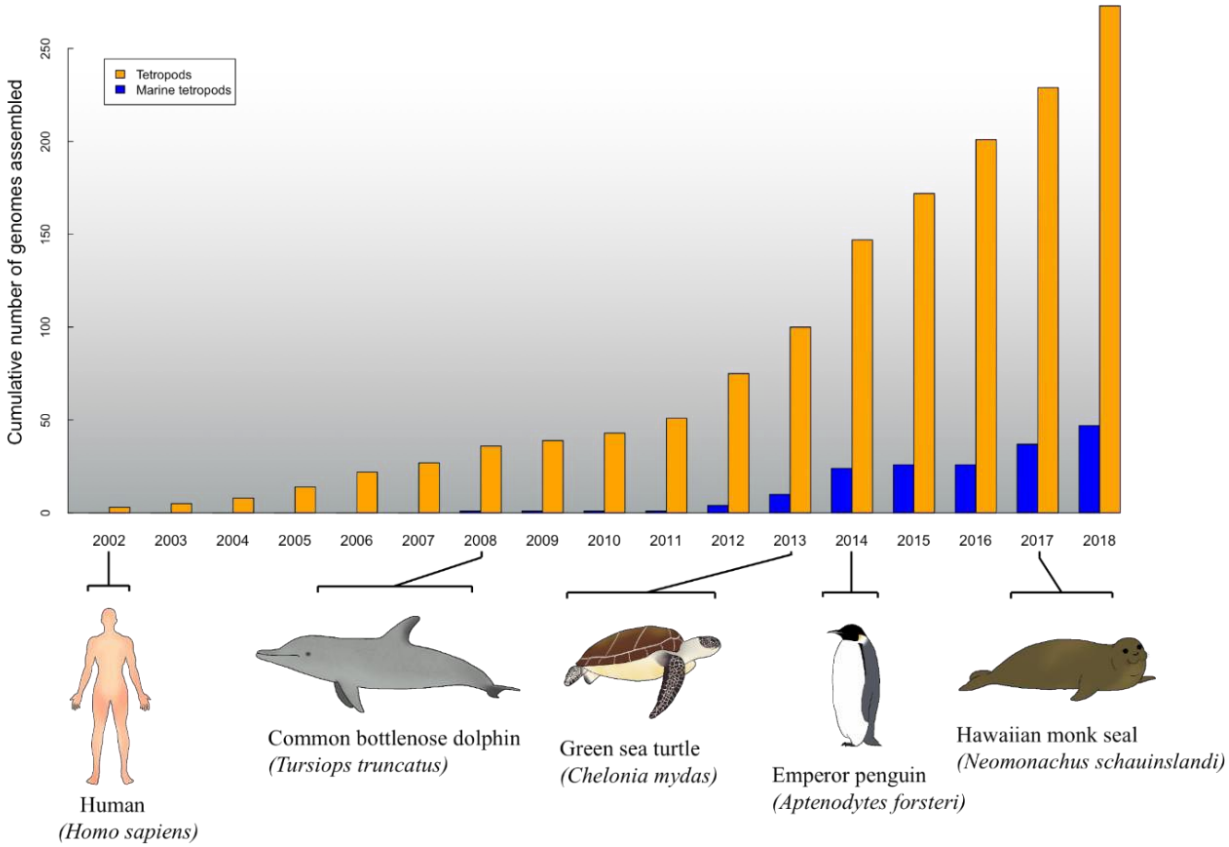


**Fig. 7.2 Numbers of assembled marine tetrapod genomes.** The cumulative generation of marine and terrestrial tetrapod genomes assembled from 2002 to 2018 is shown. The genome data are primarily obtained from NCBI and the year information is based on genome publication date.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



# 8 Applications of genomic data

## 8.1 Genetic engineering

Transgenic or genome-edited marine species are frequently produced for either scientific research or biotechnological applications. Artificially introduction and integration of a foreign gene or non-coding DNA fragment into the genomes of marine organisms, is termed transgenic modification. Organisms such as fish, crustaceans, microalgae, macroalgae, and sea urchins, with foreign transgenes integrated into their genome, are called genetically modified organisms (GMOs). New technologies such as Zinc Finger Nuclease Technology (ZFN), Transcription Activator-Like Effector Mucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), are not only efficient for transgenesis, but also make it possible to endogenously edit the genomes of marine organisms without the involvement of foreign genes.

Using transgenesis, genetically engineered marine organisms have contributed significantly to basic research areas including invertebrate and vertebrate development, the analysis of promoter/enhancer elements of genes, the dissection of signal transduction pathways, and the development of human disease models. Similarly, these organisms have also: improved biotechnological applications; and enhanced traits such as disease resistance, somatic growth, increased body color variation and stress tolerance[163-167]. Among these genetically engineered organisms, two famous applications have been commercialized: Glofish® (Fig. 8.1)[168] and Aquadvantage salmon® (Fig. 8.2)[9].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



**Fig. 8.1 Glofish® By Yorktown Technologies are sold in most pet stores in North America.**

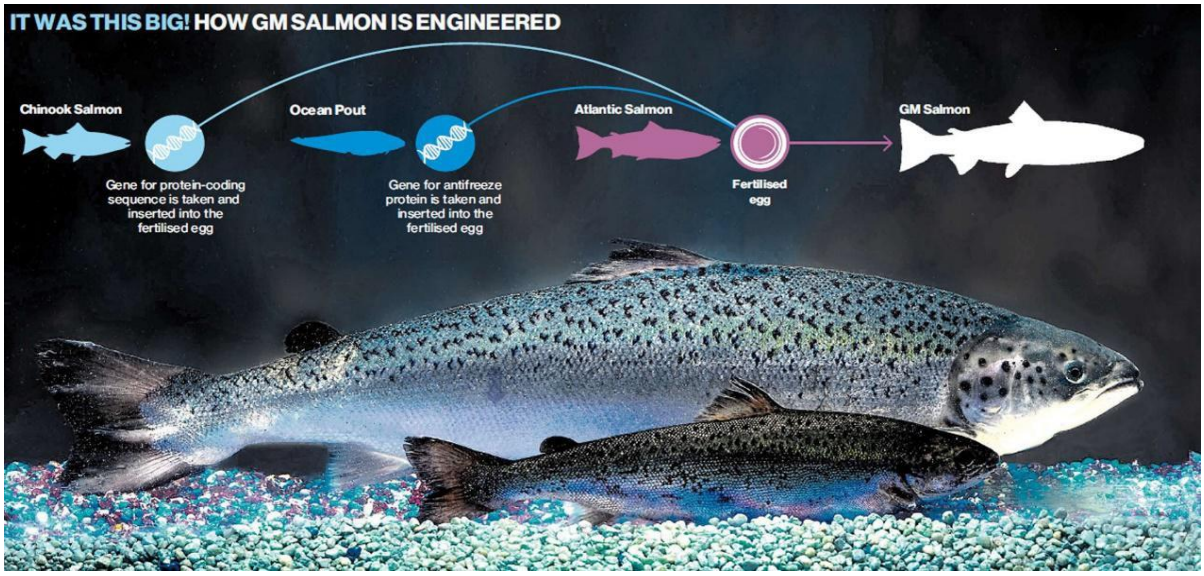


**Fig. 8.2 Aquadvantage salmon® is the first and only genetically engineered fish approved for human consumption by the FDA in North America.**

While many marine organisms have been used in transgenic research, only a limited number of species of algae[169-180], crustaceans[181,182], sea urchins[183-188] and fish[189-198] have been genome-edited by ZFN, TALENs or CRISPR. Microalgae genomes of *Chlamydomonas*[169,170,172-174,179], *Phaeodactylum*[171], *cyanobacteria*[175,177], *Synechococcus*[176] and *Nannochloropsis*[180] have been edited by targeting *COP3*, *MAA7*, *CpSRP43*, *ChlM*, *UGP*, *PEPC1*, *FKB12*, *CpFTSY*, *ZEP*, *nblA*, *glgc*, *g7988* genes or loci resulting in gene knockouts/knockins. In *Exopalaemon carinicauda*, *EcChi4* was knocked out via CRISPR/Cas9 to determine the function of the chitinase it encoded[181]. Six *hox* genes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



were knocked out by CRISPR/Cas9 to study their roles during limb development in *Parhyale hawaiensis*[182]. An increase in the primary mesenchyme cell population was observed in sea urchins embryos when HpHesC was targeted by ZFN[184]. Injection of TALEN mRNAs targeting the *HpEts* transcription factor into fertilized sea urchins' eggs resulted in the impairment of skeletogenesis[183]. However, the efficiency of these genome modification tools is far from satisfactory, although multiple attempts to knockout genes in sea urchin embryos resulted in high efficiency[185,186,188]; applying both CRISPR/Cas9 and CRISPR/Cas9d, modified sea urchins showed similar phenotypic changes, whereas genotypic changes were significantly different[187].

In fish, ZFN Technology was applied to the rainbow trout, resulting in mutation of *sdy* and disruption of male determination and differentiation[190]. In Atlantic salmon, *tyr* and *slc45a2* were successfully mutated by CRISPR/Cas9 and the P1 mosaic founders also showed varying degrees of pigment loss[192]. The use of genome editing with TALENs has also helped identify the *oca2* gene in cavefish that is responsible for reduced pigmentation[193]. In Atlantic killifish, homozygous *ahr2a* and *ahr2b* mutants generated by CRISPR/Cas9 may be useful tools for monitoring AHRs in marine environments[194]. Recently, the CRISPR/Cas9-based genome editing technology has been successfully used in the short-lived African turquoise killifish, an increasingly popular model for aging in vertebrates[195,197]. Also, the *dnd* gene was disrupted in the Atlantic salmon by CRISPR/Cas9, and the P1 mutant fish showed complete loss of pigmentation as well as a loss of germ cells in the gonads, confirming an important role for *dnd* in germline determination[189]. Disruption of the *slc24a5* by CRISPR/Cas9 caused loss of varying level of pigmentation in the skin and retina in the P1 Northeast Chinese lamprey[196]. In the Chinese tongue sole, a recent study using CRISPR/Cas9 provided evidence that *dmrt1* functions as the sex-determining gene in this species to initiate male development[191]. To increase the growth of skeletal muscle, the *myostatin* gene of red sea bream,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



*Pagrus major*, was knocked out using CRISPR/Cas9[198]. Multiple *myostatin* knockout studies have also been carried out in freshwater aquaculture fish such as channel catfish, yellow catfish and common carp[199-201] in order to increase their maximum market value.

As new technologies, especially CRISPR, merge and commercialize, genetic engineering is becoming more executable and widely applicable (Fig. 8.3).

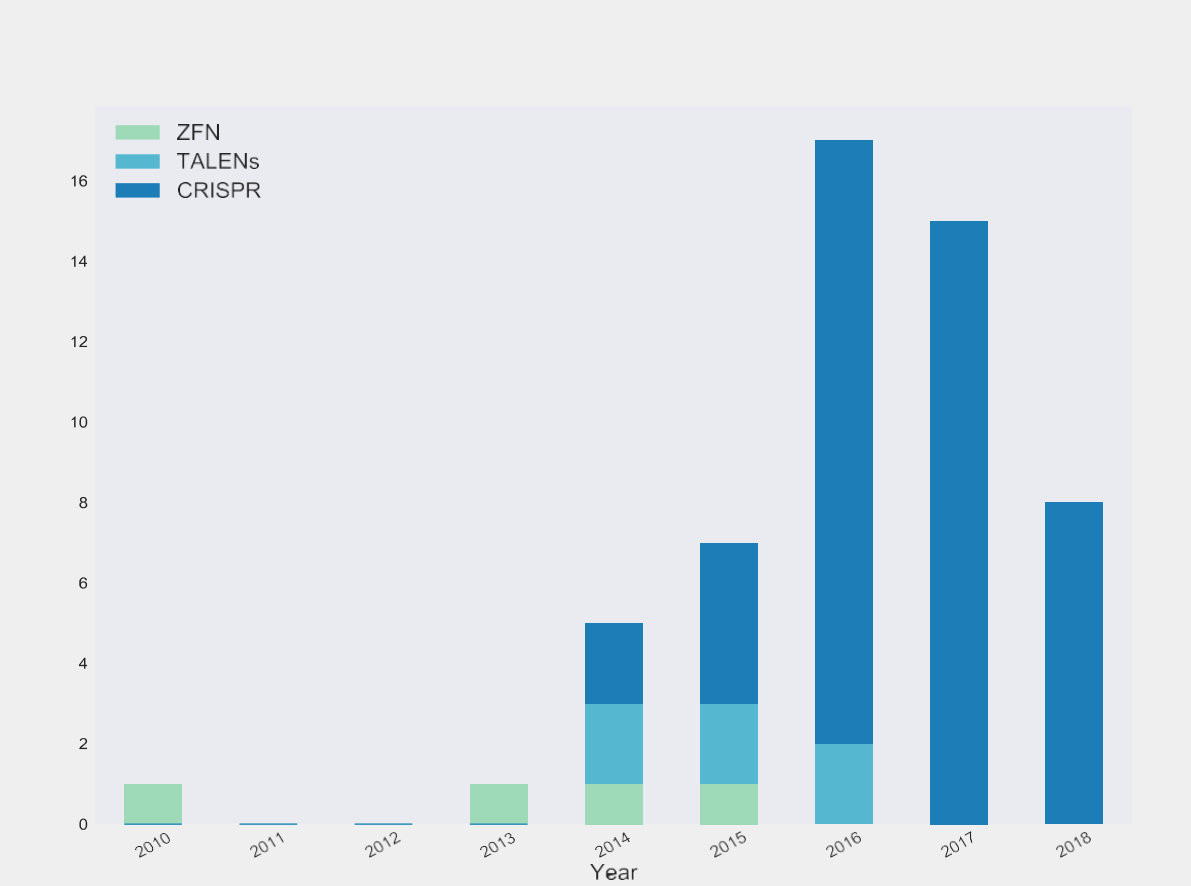


**Fig. 8.3 Genome editing in marine and aquaculture organisms.** Among the three gene editing tools available, CRISPR is most efficient and executable. Therefore, most cases of marine genome editing are CRISPR-mediated.

## 8.2 Marine natural products

Organisms living in the ocean have been subjected to intense selective pressure for millions of years resulting in many novel and unique bioactive substances required for adaptation[202]. More than 20,000 new natural marine products have been isolated over the past 50 years of which 71% have

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



not been found on land[203]. These substances are abundant, covering almost all pharmaceutical fields[204]. The basic structure of bioactive substances discovered from marine organisms is primarily peptides (or proteins), but polysaccharides (including glucoside), lipids and small molecules have also been found, with activities such as antibacterial, antitumor, antiviral, cytotoxic, anticoagulant, and antihypertensive.

Novel bioactive substances from marine organisms can also have multiple activities. For example, the unique polysaccharides, chondroitin sulfate and polypeptides of sea cucumber have been used as anti-inflammatory agents and disease-preventing food sources[205 206 207]. Four marine-derived substances, namely cytarabine, eribulin mesylate, brentuximab vedotin, and trabectidine have been approved by the FDA as drugs for the treatment of cancer[208]. To date, there are 9 drugs (developed from 8 marine compounds) in the biopharmaceutical market, mainly for the treatment of cancer, while many agents are in several different stages of the clinical pipeline[209,210].

To better study and utilize marine bioactive substances, many technologies have emerged recently for mining and identifying new bioactive substances, including using multi-omics approaches to mine new bioactive proteins/peptides (Fig. 8.4 and Table 8.1), rapidly identify natural products with the NMR and MS spectroscopic database[211] (Delp-NP platform) and the glycomics and glycogenomics strategy for screening glycans and glycosylated molecules[212 213]. However, for marine microbes, even when combining homology-based searches and phylogenetic analyses, it is still not possible to discover novel marine microbial natural products systematically on a large scale (Fig. 8.5). In the future, integrated data mining including genomics, transcriptomics, proteomics and metabolomics, as well as biosynthetic biology and structure biology, will provide alternatives to discovery approaches for marine microbial natural products (Fig. 8.6)[213-219].

Since 2000, Chinese scientists have become the main research force in marine natural products, and have discovered 46% of novel marine bioactive substances with 43% of the relevant publications. However, compared with the Western countries, most research on marine bioactive substances in China is confined to compound discovery rather than commercialization of products. Therefore, development of industrial applications of marine bioactive products is as important as discovering new products efficiently.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



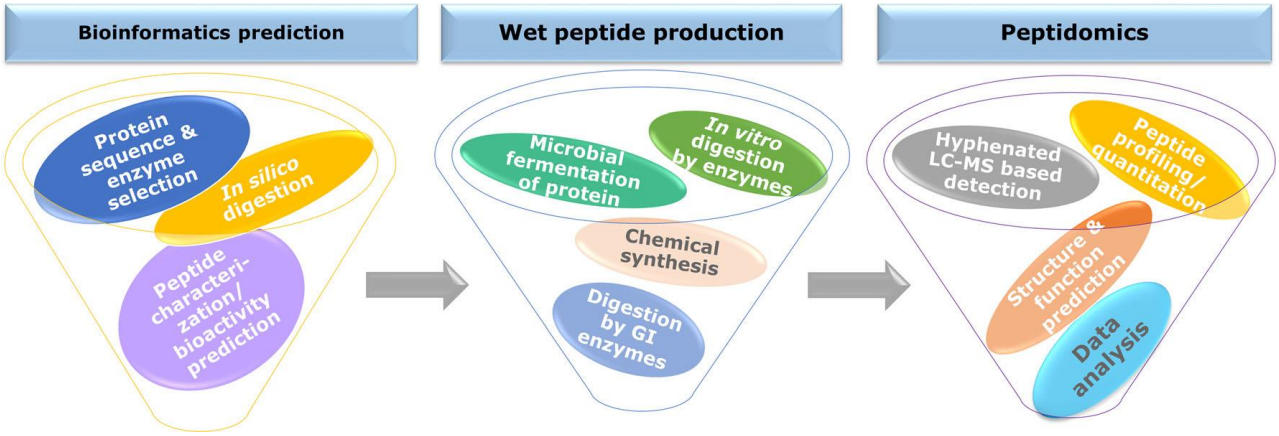


**Fig. 8.4 Strategies for mining bioactive proteins/peptides using bioinformatics analysis and peptideomics[220]**

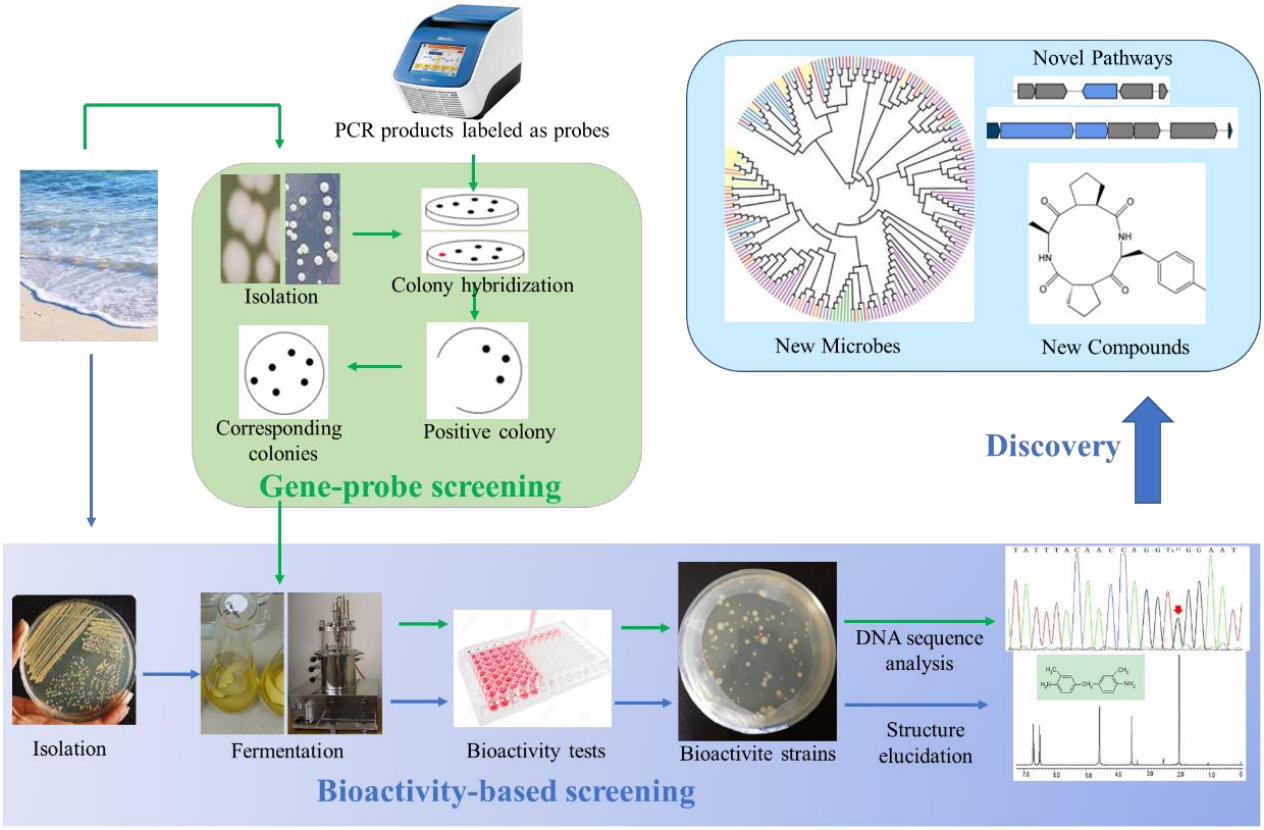


**Fig. 8.5 The combined strategy of gene-based screening and bioactivity-based screening for the discovery of marine microbial natural products (MMNPs)**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



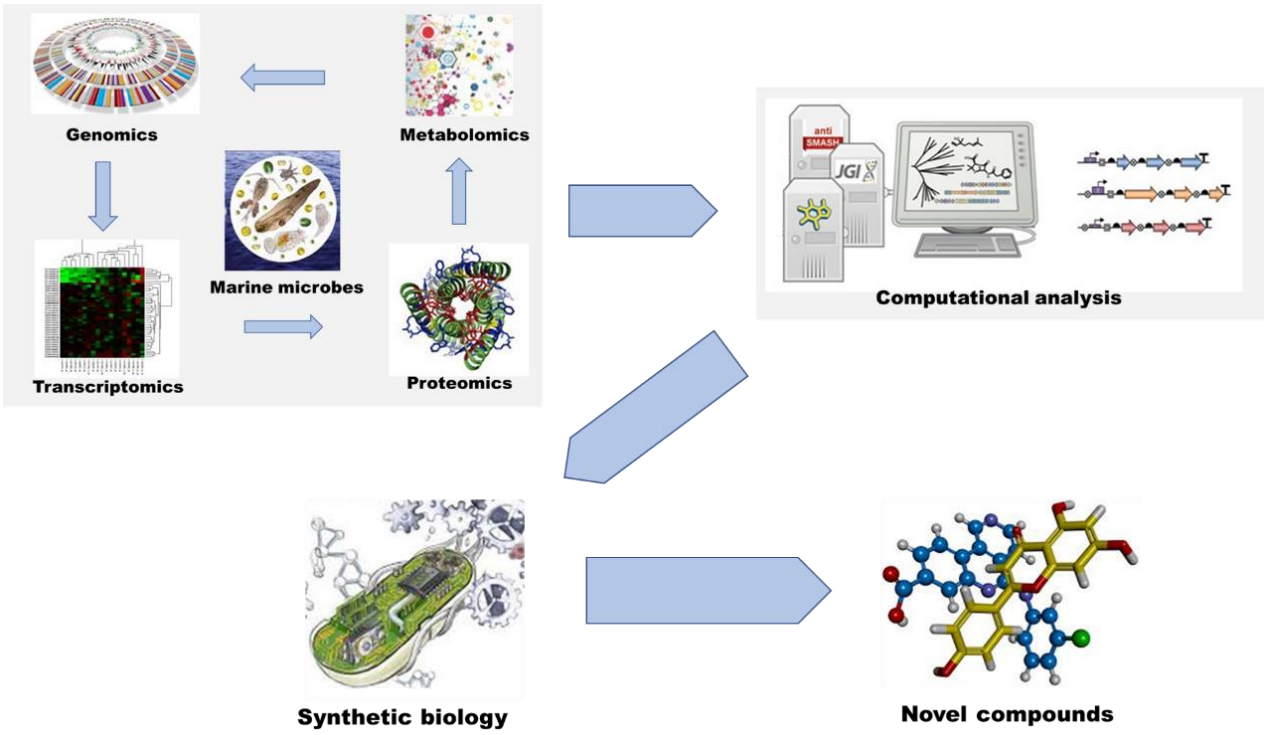


**Fig. 8.6 Flowchart describing the future of automated genome mining strategies**

**Table 8.1** A list of bioactive protein/peptide databases

| Name | Descriptions |
|---|---|
| AHTPDB | Database of antihypertensive peptides(http://crdd.osdd.net/raghava/ahtpdb/) |
| ANTISTAPHYBASE | Database containing peptides active against Staphylococcus aureus (http://www.antistaphybase.com/) |
| APD | Database of antimicrobial and anticancer peptides (http://aps.unmc.edu/AP/main.html) |
| ArachnoServer | Database of toxic peptides and proteins from spider venoms (http://www.arachnoserver.org/mainMenu.html) |
| AVPdb | Database of antiviral peptides (http://crdd.osdd.net/servers/avpdb/) |
| BaAMPs | Database of antimicrobial peptides tested against microbial biofilms (http://www.baamps.it/) |
| BACTIBASE | Database of antibacterial peptides (bacteriocins) (http://bactibase.pfba-lab-tun.org) |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | |
|---|---|
| T3DB | Comprehensively annotated database of common toxins and their targets (http://www.t3db.ca/) |
| MHCBN | Database of MHC/TAP binding peptides and T-cell epitopes (http://crdd.osdd.net/raghava/mhcbn/) |
| CAMP | Database of antimicrobial peptides and proteins (http://www.bicnirrh.res.in/antimicrobial/) |
| CancerPPD | Database of anticancer peptides (http://crdd.osdd.net/raghava/cancerppd/index.php) |
| CPPsite | Database of cell-penetrating peptides (http://crdd.osdd.net/raghava/cppsite/) |
| DADP | Database of defense peptides (http://split4.pmfst.hr/dadp/) |
| EROP-Moscow | Database of biologically active peptides (http://erop.inbi.ras.ru) |
| Hemolytik | Database of hemolytic peptides (http://crdd.osdd.net/raghava/hemolytik/) |
| HIPdb | Database of HIV inhibiting peptides (http://crdd.osdd.net/servers/hipdb/ n>) |
| Kalium | Database of toxic peptides from scorpion venom acting against potassium channels (http://kaliumdb.org/) |
| LAMP | Database of antimicrobial peptides (http://biotechlab.fudan.edu.cn/database/lamp/) |
| NeuroPedia | Database of neuropeptides including library of mass spectra (http://proteomics.ucsd.edu/Software/NeuroPedia/index.html) |
| THPdb | Database of therapeutic peptides (http://crdd.osdd.net/raghava/thpdb/index.html) |
| TumorHoPe | Database of tumor-recognizing peptides (http://crdd.osdd.net/raghava/tumorhope/) |
| WALTZ-DB | Database of amyloid hexapeptides (http://waltzdb.switchlab.org/) |

**Acknowledgements**

We appreciated all of these experts for helping us to improve this work (no ranking): Jianfang Gui, Huanming Yang, Longfei Wu, Lars Bolund, Ole Peter, Shunping He, Songlin Chen, Karsten Kristiansen, Frederick Charles Dubee, Xun Xu, Scott Edmunds, Qiang Lin, Guang Yang, Igor

 doi:10.20944/preprints201812.0156.v1



Grigoriev, Paul Jensen, Andrew Foote, Matthew Sullivan, William Richard Jeffery, Wilfred van der Donk, Hans Komen, Zhanghong Sui, Naihao Ye, Beverley Green, Chris Bowler, Oleg Simakov, Jianhai Xiang, Lynn Fink, Shi Wang, Julio Cesar Ignacio Espinoza, Wen Wang, Kun Wang, Richards Stephen, Takashi Gojobori, Senjie Lin, Piel Jörn, qing Yu, Peng Xu, Susan Brawley, Donald Anderson, Shinhan Shiu, Guangxi Wu, Ben Shen, Mark Cock, Jonas Collen, Takeshi Takeuchi, Benning Christoph, Byrappa VENKATESH, Frederic Brunet, Naoki Irie, Eiichi Shoguchi, Li Li, Dave Edwards, John Archibald, Beverley Green, Shiyu Zhao, Zenglian Wang, Xunming Zhou, Changwei Shao, Xiaodong Fang, Guojie Zhang, Lingling Zhang, Chris Bowler, Axel Meyer, Dehai Li, Ke Zhou, Wenyan Zhang, Yitao Wang.

**Author contributions**

Xin Liu and Guangyi Fan designed this work and revised the manuscript; Tao Jin, Jianwei Chen, Pengxu Yan, Guang Liu, Xiangqun Chi, Shijie Hao, Xiaochuan Liu, Xiao Du, Shuai Sun, Yue Chang, Rui Zhang, Yaolei Zhang, Hanbo Li, Ting Luo, Shengjun Wang organized and wrote the manuscript. Jiao Guo, Xiaoxuan Tan, Liangwei Li, Guilin Liu, Kai Han, Xiaoyun Huang, Le Xu, Jing Zhou, He Zhang, Mengjun Yu, Lingfeng Meng, Kaiqiang Liu, Mengqi Zhang, Yong Zhao, Chang Li, Xinyu Guo, Jiahao Wang, Meiqi Lv, Haoyang Gao, Yujie Liu, Yue Song, Yang Deng, Jinzhong Lin, Binjie Ouyang, Yinjia Yu, Jun Wang collected and analyzed the data. Lynn Fink polished the english writing of this work.

**Corresponding authors**

Correspondence to ICG-Ocean Organizing Committee (ICG-Ocean@genomics.cn), Xin Liu (liuxin@genomics.cn) or Guangyi Fan (fanguangyi@genomics.cn).

## Declaration

Due to the timeliness and the rapid iteration of marine genomics, we cannot guarantee the completeness of the reports, and the suggestions and comments are welcome.

## References


1 Olsen, J. L. *et al.* The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. **530**, 331-335 (2016).

2 Matsuzaki, M. *et al.* Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. **428**, 653-657 (2004).

3 Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. **6215**, 1311-1320 (2014).

4 Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. **348**, 1261359 (2015).

5 Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. **490**, 49-54 (2012).

6 Dehal, P. *et al.* The draft genome sequence of the ascidian Ciona intestinalis: insights into the evolutionary origins of chordates and vertebrates. **298**, 2157-2167 (2002).

7 Star, B. *et al.* The genome sequence of Atlantic cod reveals a unique immune system. **477**, 207 (2011).

8 Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. **297**, 1301-1310 (2002).

9 Du, S. J. *et al.* Growth enhancement in transgenic Atlantic salmon by the use of an "all fish" chimeric growth hormone gene construct. **10**, 176 (1992).

10 Rocap, G. *et al.* Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. **424**, 1042-1047 (2003).

11 Sun, Y. *et al.* Fish-T1K (Transcriptomes of 1,000 Fishes) Project: large-scale transcriptome data for fish evolution studies. **5**, 18 (2016).

12 Bernardi, G. *et al.* The fishes of Genome 10K. **7**, 3-6 (2012).

13 Cheng, S. *et al.* 10KP: A phylodiverse genome sequencing plan. **7**, giy013 (2018).

14 Lewin, H. A. *et al.* Earth BioGenome Project: Sequencing life for the future of life. **115**, 4325-4333 (2018).

15 Overmann, J. & Lepleux, C. Marine Bacteria and Archaea: Diversity, Adaptations, and Culturability. 21-55, doi:10.1007/978-3-319-33000-6_2 (2016).

16 Offre, P., Spang, A. & Schleper, C. Archaea in biogeochemical cycles. *Annu Rev Microbiol* **67**, 437-457, doi:10.1146/annurev-micro-092412-155614 (2013).

17 Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353-358, doi:10.1038/nature21031 (2017).

18 Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, doi:10.1126/science.aaf3883 (2017).

19 Mende, D. R. *et al.* Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol* **2**, 1367-1373, doi:10.1038/s41564-017-0008-3 (2017).

20 Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**, 4765-4774 (1998).

21 Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**, R245-249 (1998).

22 Chen, K. & Pachter, L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comput Biol* **1**, 106-112, doi:10.1371/journal.pcbi.0010024 (2005).

23 Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74, doi:10.1126/science.1093857 (2004).

24 Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359, doi:10.1126/science.1261359 (2015).

25 Bergh, O., Borsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in marine environments. *Nature* **340**, 467-468, doi:10.1038/340467a0 (1989).

26 Suttle, C. A. Viruses in the sea. *Nature* **437**, 356-361, doi:10.1038/nature04160 (2005).

27 Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in marine ecosystems. *Microbiol Mol Biol Rev* **64**, 69-114 (2000).

28 Dwivedi, B., Schmieder, R., Goldsmith, D. B., Edwards, R. A. & Breitbart, M. PhiSiGns: an online tool to identify signature genes in phages and design PCR primers for examining phage diversity. *BMC Bioinformatics* **13**, 37, doi:10.1186/1471-2105-13-37 (2012).

29 Paez-Espino, D. *et al.* Uncovering Earth's virome. *Nature* **536**, 425-430, doi:10.1038/nature19094 (2016).

30 Hawksworth, D. L. & Lucking, R. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiol Spectr* **5**, doi:10.1128/microbiolspec.FUNK-0052-2016 (2017).

31 Boddy, L. Genetics – Variation, Sexuality, and Evolution. 99-139, doi:10.1016/b978-0-12-382034-1.00004-9 (2016).

32 Kohlmeyer, J. & Kohlmeyer, E. Marine Mycology: The Higher Fungi -Introduction. 1-6, doi:10.1016/b978-0-12-418350-6.50006-1 (1979).

33 Rédou, V. *et al.* Marine Fungi. 99-153, doi:10.1007/978-3-319-33000-6_4 (2016).

34 McLaughlin, D. J. & Spatafora, J. W. The Mycota VII - Systematics and Evolution, Second Edition.

doi:10.1007/978-3-642-55318-9 (2014).

35 Tedersoo, L. *et al.* High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Diversity* **90**, 135-159, doi:10.1007/s13225-018-0401-0 (2018).

36 Kirk, P. M., Cannon, P. F., Minter, D. W. & Stalpers, J. A. Dictionary of the Fungi, 10th Edition. . *CABI Europe - UK*, [i]-xi, [1]-771 (2008).

37 Jones, M. D. *et al.* Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**, 200-203, doi:10.1038/nature09984 (2011).

38 Cavalier-Smith, T. Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *European Journal of Protistology* **49**, 115-178, doi:10.1016/j.ejop.2012.06.001 (2013).

39 Ruiz-Herrera, J. & Ortiz-Castellanos, L. Analysis of the phylogenetic relationships and evolution of the cell walls from yeasts and fungi. *FEMS Yeast Res* **10**, 225-243, doi:10.1111/j.1567-1364.2009.00589.x (2010).

40 Barghoorn, E. S. & Linder, D. H. Marine fungi: their taxonomy and biology. *Farlowia* **1**, 395-467 (1944).

41 *WoRMS*, <http://www.marinespecies.org/index.php> (2018).

42 Didier, E. S. Microsporidiosis: an emerging and opportunistic infection in humans and animals. *Acta Trop* **94**, 61-76, doi:10.1016/j.actatropica.2005.01.010 (2005).

43 Hyde, K. D. *et al.* Role of fungi in marine ecosystems. *Biodiversity & Conservation* **7**, 1147-1161, doi:10.1023/a:1008823515157 (1998).

44 Mydlarz, L. D. & Harvell, C. D. Peroxidase activity and inducibility in the sea fan coral exposed to a fungal pathogen. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **146**, 54-62, doi:10.1016/j.cbpa.2006.09.005 (2007).

45 Rypien, K. L. African dust is an unlikely source of Aspergillus sydowii, the causative agent of sea fan disease. *Marine Ecology Progress Series* **367**, 125-131, doi:10.3354/meps07600 (2008).

46 Kohlmeyer, J. & Kohlmeyer, E. Marine Mycology: The Higher Fungi - Fungi in Animal Substrates. 137-142, doi:10.1016/b978-0-12-418350-6.50023-1 (1979).

47 Sridhar, K. R. Mangrove fungi in India. *Current science* **86**, 1586-1587 (2006).

48 Dittami, S. M., Eveillard, D. & Tonon, T. A metabolic approach to study algal-bacterial interactions in changing environments. *Molecular ecology* **23**, 1656-1660, doi:10.1111/mec.12670 (2014).

49 Stock, A. *et al.* Microbial eukaryote life in the new hypersaline deep-sea basin Thetis. *Extremophiles* **16**, 21-34, doi:10.1007/s00792-011-0401-4 (2012).

50 *FungiDB* <http://fungidb.org/fungidb/> (2018).

51 *1000 Fungal Genome Project*, <http://1000.fungalgenomes.org/home/> (2018).

52 Richards, T. A., Jones, M. D. M., Leonard, G. & Bass, D. Marine Fungi: Their Ecology and Molecular Diversity.

*Annual Review of Marine Science* **4**, 495-522, doi:10.1146/annurev-marine-120710-100802 (2012).

53 Panzer, K. *et al.* Identification of Habitat-Specific Biomes of Marine Fungal Communities Using a Comprehensive Nearly Full-Length 18S rRNA Dataset Enriched with Contextual Data. *PloS one* **10**, e0134377, doi:10.1371/journal.pone.0134377 (2015).

54 Randrianarison, G. & Ashraf, M. A. Microalgae: a potential plant for energy production. *Geology, Ecology, and Landscapes* **1**, 104-120, doi:10.1080/24749508.2017.1332853 (2017).

55 Gonçalves, A. L., Pires, J. C. M. & Simões, M. A review on the use of microalgal consortia for wastewater treatment. *Algal Research* **24**, 403-415, doi:10.1016/j.algal.2016.11.008 (2017).

56 Hidalgo, O. *et al.* Is There an Upper Limit to Genome Size? *Trends Plant Sci* **22**, 567-573, doi:10.1016/j.tplants.2017.04.005 (2017).

57 Les D H, C. M. A., Waycott M. Phylogenetic studies in Alismatidae, II: evolution of marine angiosperms (seagrasses) and hydrophily. *Systematic Botany*, 443-463 (1997).

58 Giri, C. *et al.* Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography* **20**, 154-159, doi:10.1111/j.1466-8238.2010.00584.x (2011).

59 Kura-Hotta M, M. M., Tsujimura T, et al. High salt-treatment-induced Na+ extrusion and low salt-treatment-induced Na+ accumulation in suspension-cultured cells of the mangrove plant, Bruguiera sexangula. *Plant, Cell & Environment* **24**, 1105-1112 (2001).

60 Liang, S. *et al.* Transcriptional homeostasis of a mangrove species, Ceriops tagal, in saline environments, as revealed by microarray analysis. *PloS one* **7**, e36499, doi:10.1371/journal.pone.0036499 (2012).

61 B., T. P. The botany of mangroves. *Cambridge tropical biology series* (1986).

62 C, B. M. Ecophysiology of mangroves. *Trees* **2**, 129-142 (1988).

63 Parida, A. K. & Jha, B. Salt tolerance mechanisms in mangroves: a review. *Trees* **24**, 199-217, doi:10.1007/s00468-010-0417-x (2010).

64 Lee, H. *et al.* The Genome of a Southern Hemisphere Seagrass Species (Zostera muelleri). *Plant physiology* **172**, 272-283, doi:10.1104/pp.16.00868 (2016).

65 Olsen, J. L. *et al.* The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature* **530**, 331-335, doi:10.1038/nature16548 (2016).

66 Xu, S. *et al.* The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole genome sequencing. *National Science Review*, doi:10.1093/nsr/nwx065 (2017).

67 Ruggiero, M. A. *et al.* A higher level classification of all living organisms. *PLoS ONE* **10**, e0119248 (2015).

68 Yuan, J. *et al.* Genomic resources and comparative analyses of two economical penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*. *Marine Genomics* (2017).

69 Shinzato, C. *et al.* Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320 (2011).

70 Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology* **10**, 641-654 (2012).

71 Muller-Parker, G., D'Elia, C. F. & Cook, C. B. Interactions Between Corals and Their Symbiotic Algae. (2015).

72 Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49-54 (2012).

73 Consortium, S. U. G. S. The Genome of the Sea Urchin Strongylocentrotus purpuratus. *Science* **314**, 941-952 (2016).

74 Müller, W. E. G. Origin of metazoan adhesion molecules and adhesion receptors as deduced from cDNA analyses in the marine sponge *Geodia cydonium* : a review. *Cell and Tissue Research* **289**, 383-395 (1997).

75 Nichols, S. A., Dirks, W., Pearse, J. S. & King, N. Early evolution of animal cell signaling and adhesion genes. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12451-12456 (2006).

76 Suga, H., Katoh, K. & Miyata, T. Sponge homologs of vertebrate protein tyrosine kinases and frequent domain shufflings in the early evolution of animals before the parazoan-eumetazoan split. *Gene* **280**, 195-201 (2001).

77 Srivastava, M. *et al.* The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720-726 (2010).

78 Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86-94 (2007).

79 Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526 (2013).

80 Francis, W. R. *et al.* The genome of the contractile demosponge *Tethya wilhelma* and the evolution of metazoan neural signalling pathways. *bioRxiv*, 120998, doi:10.1101/120998 (2017).

81 Chapman, J. A. *et al.* The dynamic genome of Hydra. *Nature* **464**, 592-596, doi:10.1038/nature08830 (2010).

82 Pani, A. M. & Al, E. Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* **483**, 289-294 (2012).

83 Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220-224, doi:10.1038/nature14668 (2015).

84 Li, C., Haug, T., Styrvold, O. B., Jørgensen, T. Ø. & Stensvåg, K. Strongylocins, novel antimicrobial peptides from the green sea urchin, *Strongylocentrotus droebachiensis*. *Developmental and Comparative Immunology* **32**, 1430-1440 (2008).

85 Tirapé, A., Bacque, C., Brizard, R., Vandenbulcke, F. & Boulo, V. Expression of immune-related genes in the oyster *Crassostrea gigas* during ontogenesis. *Developmental and Comparative Immunology* **31**, 859-

873 (2007).

86 Christophides, G. K. *et al.* Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**, 159-165, doi:10.1126/science.1077136 (2002).

87 Loker, E. S., Adema, C. M., Zhang, S. M. & Kepler, T. B. Invertebrate immune systems-not homogeneous, not simple, not well understood. *Immunological Reviews* **198**, 10-24 (2010).

88 Zdobnov, E. M. *et al.* Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149-159 (2002).

89 Pancer, Z., Skorokhod, A., Blumbach, B. & Müller, W. E. Multiple Ig-like featuring genes divergent within and among individuals of the marine sponge *Geodia cydonium*. *Gene* **207**, 227 (1998).

90 Cadavid, L. F., Powell, A. E., Nicotra, M. L., Moreno, M. & Buss, L. W. An invertebrate histocompatibility complex. *Genetics* **167**, 357-365 (2004).

91 Cuthbertson, B. J., Shepard, E. F., Chapman, R. W. & Gross, P. S. Diversity of the penaeidin antimicrobial peptides in two shrimp species. *Immunogenetics* **54**, 442-445, doi:10.1007/s00251-002-0487-z (2002).

92 Pancer, Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13156 (2000).

93 Doino, L. J. & Mcfall-Ngai, M. J. Alterations in the proteome of the Euprymna scolopes light organ in response to symbiotic *Vibrio fischeri*. *Applied and Environmental Microbiology* **66**, 4091-4097 (2000).

94 Krueger, D. M., Gustafson, R. G. & Cavanaugh, C. M. Vertical transmission of chemoautotrophic symbionts in the bivalve *Solemya velum* (Bivalvia: Protobranchia). *Biological Bulletin* **190**, 195-202 (1996).

95 Nakayama, K., Nishijima, M. & Maruyama, T. Morula-like cells in photo-symbiotic clams harboring zooxanthellae. *Zoological Science* **15**, 339 (1998).

96 Zhang, Y. *et al.* Adaptation and evolution of deep-sea scale worms (Annelida: Polynoidae): insights from transcriptome comparison with a shallow-water species. *Scientific Reports* **7**, 46205 (2017).

97 Sun, J. *et al.* Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature Ecology and Evolution* **1**, 0121 (2017).

98 Takeuchi, T. *et al.* Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Letters* **2**, 3, doi:10.1186/s40851-016-0039-2 (2016).

99 Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49-54, doi:10.1038/nature11413 (2012).

100 Barghi, N., Concepcion, G. P., Olivera, B. M. & Lluisma, A. O. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Molecular Genetics and Genomics* **291**, 411-422, doi:10.1007/s00438-015-1119-2 (2016).

101 Du, X. *et al.* Draft genome and SNPs associated with carotenoid accumulation in adductor muscles of

bay scallop (*Argopecten irradians*). *Journal of genomics* **5**, 83-90, doi:10.7150/jgen.19146 (2017).

102 Modica, M. V., Lombardo, F., Franchini, P. & Oliverio, M. The venomous cocktail of the vampire snail *Colubraria reticulata* (Mollusca, Gastropoda). *BMC Genomics* **16**, 441, doi:10.1186/s12864-015-1648-4 (2015).

103 Murgarella, M. *et al.* A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. *PLoS One* **11**, e0151561 (2016).

104 O'Connor, R. M. *et al.* Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E5096-E5104 (2014).

105 Kaiser, T. S. *et al.* The genomic basis of circadian and circalunar timing adaptations in a midge. *Nature* **540**, 69-73, doi:10.1038/nature20151 (2016).

106 Du, X. *et al.* The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *GigaScience* **6**, gix059 (2017).

107 Sergiev, P. V., Artemov, A. A., Prokhortchouk, E. B., Dontsova, O. A. & Berezkin, G. V. Genomes of *Strongylocentrotus franciscanus* and *Lytechinus variegatus*: are there any genomic explanations for the two order of magnitude difference in the lifespan of sea urchins? *Aging (Albany NY)* **8**, 260 (2016).

108 Logue, K. *et al.* Whole - genome sequencing reveals absence of recent gene flow and separate demographic histories for *Anopheles punctulatus* mosquitoes in Papua New Guinea. *Molecular ecology* **24**, 1263-1274 (2015).

109 Reiswig, H. M. Partial carbon and energy budgets of the bacteriosponge *Verohgia fistularis* (Porifera: Demospongiae) in Barbados. *Marine Ecology* **2**, 273-293 (2010).

110 Schlappy, M. L. *et al.* Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Marine biology* **157**, 593-602, doi:10.1007/s00227-009-1344-5 (2010).

111 Taylor, M. W., Radax, R., Steger, D. & Wagner, M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews* **71**, 295-347 (2007).

112 Webster, N. S. & Taylor, M. W. Marine sponges and their microbial symbionts: love and other relationships. *Environmental Microbiology* **14**, 335-346 (2012).

113 Voolstra, C. R. *et al.* Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals. *Scientific Reports* **7** (2017).

114 Voolstra, C. *et al.* The ReFuGe 2020 Consortium—using "omics" approaches to explore the adaptability and resilience of coral holobionts to environmental change. *Frontiers in Marine Science* **2**, doi:10.3389/fmars.2015.00068 (2015).

115 Hall, M. R. *et al.* The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* **544**, 231-234, doi:10.1038/nature22033 (2017).

116 Wataru, I. *et al.* MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. **30**, 2531-2540 (2013).

117 Pollux, B., Pires, M., Banet, A. & Reznick, D. J. A. R. E. E. S. Evolution of placentas in the fish family Poeciliidae: an empirical study of macroevolution. **40**, 271-289 (2009).

118 Evans, J. P., Pilastro, A. & Schlupp, I. *Ecology and evolution of poeciliid fishes*. (University of Chicago Press, 2011).

119 Schartl, M. *et al.* The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. **45**, 567 (2013).

120 Nikaido, M. *et al.* Coelacanth genomes reveal signatures for evolutionary transition from water to land. (2013).

121 Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. **496**, 311 (2013).

122 Boudinot, P. *et al.* A tetrapod-like repertoire of innate immune receptors and effectors for coelacanths. **322**, 415-437 (2014).

123 Saha, N. R. *et al.* Genome complexity in the coelacanth is reflected in its adaptive immune system. **322**, 438-463 (2014).

124 Boisvert, C. A., Joss, J. M. & Ahlberg, P. E. J. E. Comparative pelvic development of the axolotl (Ambystoma mexicanum) and the Australian lungfish (Neoceratodus forsteri): conservation and innovation across the fish-tetrapod transition. **4**, 3 (2013).

125 King, H. M., Shubin, N. H., Coates, M. I. & Hale, M. E. J. P. o. t. N. A. o. S. Behavioral evidence for the evolution of walking and bounding before terrestriality in sarcopterygian fishes. **108**, 21146-21151 (2011).

126 Eastman, J. T. The nature of the diversity of Antarctic fishes. **28**, 93-107 (2005).

127 Shin, S. C. *et al.* The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. **15**, 468 (2014).

128 Rétaux, S. & Casane, D. J. E. Evolution of eye development in the darkness of caves: adaptation, drift, or both? **4**, 26 (2013).

129 Nelson, J. S., Grande, T. C. & Wilson, M. V. *Fishes of the World*. (John Wiley & Sons, 2016).

130 Meng, F. *et al.* Evolution of the eye transcriptome under constant darkness in Sinocyclocheilus cavefish. **30**, 1527-1543 (2013).

131 Jeffery, W. R. Regressive evolution in Astyanax cavefish. **43**, 25-47 (2009).

132 Yamamoto, Y., Stock, D. W. & Jeffery, W. R. Hedgehog signalling controls eye degeneration in blind cavefish. **431**, 844 (2004).

133 Trede, N. S., Langenau, D. M., Traver, D., Look, A. T. & Zon, L. I. J. I. The use of zebrafish to understand

immunity. **20**, 367-379 (2004).

134 Hoseinifar, S. H., Esteban, M. Ã. N., Cuesta, A. & Sun, Y. Z. J. R. i. F. S. Prebiotics and Fish Immune Response: A Review of Current Knowledge and Future Perspectives. **23**, 315-328 (2015).

135 Sugamata, R., Suetake, H., Kikuchi, K. & Suzuki, Y. J. J. o. I. Teleost B7 expressed on monocytes regulates T cell responses. **182**, 6799 (2009).

136 Zhang, Y. A. *et al.* Conservation of structural and functional features in a primordial CD80/86 molecule from rainbow trout (Oncorhynchus mykiss), a primitive teleost fish. **183**, 83-96 (2009).

137 Catherine Wilson, S. K. H., Braedan M Mccluskey, Angel Amores, Yilin Yan, Tom A Titus, Jennifer L Anderson, Peter Batzel, Michael J Carvan, Manfred Schartl, John H Postlethwait. Wild Sex in Zebrafish: Loss of the Natural Sex Determinant in Domesticated Strains. *Genetics* **198**, 1291-1308 (2014).

138 Francesc Piferrer, M. B., Laia Navarro, Alicia Gonzalez. Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (Dicentrarchus labrax L.). *General and Comparative Endocrinology* **142**, 102-110 (2005).

139 Yann Guiguen, A. F., Francesc Piferrer, Chingfong Chang. Ovarian aromatase and estrogens: A pivotal role for gonadal sex differentiation and sex change in fish. *General and Comparative Endocrinology* **165**, 352-366 (2010).

140 Jeanfrancois Baroiller, H. D., Eric Saillant. Environmental Effects on Fish Sex Determination and Differentiation. *Sexual Development* **3**, 118-135 (2009).

141 Yasuhisa Kobayashi, Y. N., Masaru Nakamura. Diversity and Plasticity of Sex Determination and Differentiation in Fishes. *Sexual Development* **7**, 115-125 (2013).

142 Zhang, Z. *et al.* Draft genome of the protandrous Chinese black porgy, Acanthopagrus schlegelii. **7**, giy012 (2018).

143 Campinho, M. A. *et al.* A thyroid hormone regulated asymmetric responsive centre is correlated with eye migration during flatfish metamorphosis. **8**, 12267 (2018).

144 Power, D. M. *et al.* The molecular and endocrine basis of flatfish metamorphosis. **16**, 95-111 (2008).

145 Savelli, I., Flamarique, I. N., Iwanicki, T. & Taylor, J. S. J. S. r. Parallel opsin switches in multiple cone types of the starry flounder retina: tuning visual pigment composition for a demersal life style. **8**, 4763 (2018).

146 Shao, C. *et al.* Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (Paralichthys olivaceus): applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis. **22**, 161-170 (2015).

147 Song, W. *et al.* Construction of high-density genetic linkage maps and mapping of growth-related quantitative trail loci in the Japanese flounder (Paralichthys olivaceus). **7**, e50404 (2012).

148 Shao, C. *et al.* The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. **49**, 119 (2017).

149 Miller, W. *et al.* Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences* **109**, E2382-E2390 (2012).

150 Zhou, X. *et al.* Baiji genomes reveal low genetic variability and new insights into secondary marine adaptations. *Nature communications* **4**, 2708 (2013).

151 Wang, Z. *et al.* The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nature genetics* **45**, 701 (2013).

152 Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320-1331 (2014).

153 Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311-1320 (2014).

154 Yim, H.-S. *et al.* Minke whale genome and marine adaptation in cetaceans. *Nature genetics* **46**, 88 (2014).

155 Kishida, T., Thewissen, J., Hayakawa, T., Imai, H. & Agata, K. Marine adaptation and the evolution of smell and taste in whales. *Zoological letters* **1**, 1 (2015).

156 Keane, M. *et al.* Insights into the evolution of longevity from the bowhead whale genome. *Cell reports* **10**, 112-122 (2015).

157 Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M. A. & Janke, A. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science advances* **4**, eaap9873 (2018).

158 Zhou, X. *et al.* Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nature communications* **9**, 1276 (2018).

159 Foote, A. D. *et al.* Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature communications* **7**, 11693 (2016).

160 Cammen, K. M., Wilcox, L. A., Rosel, P. E., Wells, R. S. & Read, A. J. From genome-wide to candidate gene: an investigation of variation at the major histocompatibility complex in common bottlenose dolphins exposed to harmful algal blooms. *Immunogenetics* **67**, 125-133 (2015).

161 Cammen, K. M., Schultz, T. F., Rosel, P. E., Wells, R. S. & Read, A. J. Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (Tursiops truncatus). *Molecular ecology* **24**, 4697-4710 (2015).

162 Li, S. *et al.* Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome biology* **15**, 557 (2014).

163 Chen, T. T. & Powers, D. A. Transgenic fish. *Trends Biotechnol* **8**, 209-215 (1990).

164 Fletcher, G. L. & Davies, P. L. *Transgenic fish for aquaculture*. (Springer US, 1991).

165 Hackett, P. B., Alvarez, M. C., Fingerman, M. & Nagabhushanam, R. J. A. M. B. The molecular genetics of transgenic fish. **4** (2000).

166 Chiou, P. Transgenic Fish: in Genomics and Genetics.

167 Chen, T. T. *et al. Transgenic Te 13. Transgenic Technology in Marine Organisms*. (2015).

168 Gong, Z. *et al.* Development of transgenic fish for ornamental and bioreactor by strong expression of fluorescent proteins in the skeletal muscle. *Biochem Biophys Res Commun* **308**, 58-63 (2003).

169 Sizova, I., Greiner, A., Awasthi, M., Kateriya, S. & Hegemann, P. J. P. J. Nuclear gene targeting in Chlamydomonas using engineered zinc-finger nucleases. **73**, 873-882 (2013).

170 Shin, S. E. *et al.* CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. *Scientific reports* **6**, 27810, doi:10.1038/srep27810 (2016).

171 Daboussi, F. *et al.* Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nat Commun* **5**, 3831, doi:10.1038/ncomms4831 (2014).

172 Kao, P. H. & Ng, I. S. CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in Chlamydomonas reinhardtii. *Bioresour Technol* **245**, 1527-1537, doi:10.1016/j.biortech.2017.04.111 (2017).

173 Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M. & Weeks, D. P. Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryotic cell* **13**, 1465-1469, doi:10.1128/EC.00213-14 (2014).

174 Baek, K. *et al.* DNA-free two-gene knockout in Chlamydomonas reinhardtii via CRISPR-Cas9 ribonucleoproteins. *Scientific reports* **6**, 30620, doi:10.1038/srep30620 (2016).

175 Ungerer, J. & Pakrasi, H. B. J. S. R. Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. **6**, 39681 (2016).

176 Wendt, K. E., Justin, U., Cobb, R. E., Zhao, H. & Pakrasi, H. B. J. M. C. F. CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacteriumSynechococcus elongatusUTEX 2973. **15**, 115 (2016).

177 Li, H. *et al.* CRISPR-Cas9 for the Genome Engineering of Cyanobacteria and Succinate Production. **38**, 293-302 (2016).

178 Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M. & Winge, P. J. S. R. A CRISPR/Cas9 system adapted for gene editing in marine algae. **6**, 24951 (2016).

179 Ferenczi, A., Pyott, D. E., Xipnitou, A. & Molnar, A. J. P. N. A. S. U. S. A. Efficient targeted DNA editing and replacement in Chlamydomonas reinhardtii using Cpf1 ribonucleoproteins and single-stranded DNA. **114**, 13567-13572 (2017).

180 Wang, Q. *et al.* Genome editing of model oleaginous microalgae Nannochloropsis spp. by CRISPR/Cas9. **88**, 1071-1081 (2016).

181 Gui, T. *et al.* CRISPR/Cas9-Mediated Genome Editing and Mutagenesis of EcChi4 in Exopalaemon carinicauda. *G3 (Bethesda)* **6**, 3757-3764, doi:10.1534/g3.116.034082 (2016).

182 Martin, A. *et al.* CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb

Specification and Evolution. *Curr Biol* **26**, 14-26, doi:10.1016/j.cub.2015.11.021 (2016).

183 Hosoi, S., Sakuma, T., Sakamoto, N., Yamamoto, T. J. D. G. & Differentiation. Targeted mutagenesis in sea urchin embryos using TALENs. **56**, 92–97 (2014).

184 Ochiai, H. *et al.* Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. **344**, 875–885 (2010).

185 Oulhen, N., Swartz, S. Z., Laird, J., Mascaro, A. & Wessel, G. M. Transient translational quiescence in primordial germ cells. *Development* **144**, 1201-1210, doi:10.1242/dev.144170 (2017).

186 Oulhen, N., Wessel, G. M. J. M. R. & Development. Albinism as a visual, in vivo guide for CRISPR/Cas9 functionality in the sea urchin embryo. **83**, 1046-1047 (2016).

187 Shevidi, S., Uchida, A., Schudrowitz, N., Wessel, G. M. & Yajima, M. J. D. D. Single nucleotide editing without DNA cleavage using CRISPR/Cas9-deaminase in the sea urchin embryo. **246**, 1036-1046 (2017).

188 Lin, C. Y. & Su, Y. H. J. D. B. Genome editing in sea urchin embryos by using a CRISPR/Cas9 system. **409**, 420-428 (2016).

189 Wargelius, A. *et al.* Dnd knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. *Scientific reports* **6**, 21284, doi:10.1038/srep21284 (2016).

190 Yano, A., Nicol, B., Jouanno, E. & Guiguen, Y. J. M. B. Heritable targeted inactivation of the rainbow trout (Oncorhynchus mykiss) master sex-determining gene using zinc-finger nucleases. **16**, 243-250 (2014).

191 Cui, Z. *et al.* Genome editing reveals dmrt1 as an essential male sex-determining gene in Chinese tongue sole (Cynoglossus semilaevis). **7**, 42213 (2017).

192 Edvardsen, R. B., Leininger, S., Kleppe, L., Skaftnesmo, K. O. & Wargelius, A. J. P. O. Targeted mutagenesis in Atlantic salmon (Salmo salar L.) using the CRISPR/Cas9 system induces complete knockout individuals in the F0 generation. **9**, e108622 (2014).

193 Ma, L., Jeffery, W. R., Essner, J. J. & Kowalko, J. E. J. P. O. Genome Editing Using TALENs in Blind Mexican Cavefish, Astyanax mexicanus. **10**, e0119370 (2015).

194 Aluru, N. *et al.* Targeted mutagenesis of aryl hydrocarbon receptor 2a and 2b genes in Atlantic killifish (Fundulus heteroclitus). *Aquat Toxicol* **158**, 192-201, doi:10.1016/j.aquatox.2014.11.016 (2015).

195 Harel, I., Valenzano, D. R. & Brunet, A. Efficient genome engineering approaches for the short-lived African turquoise killifish. *Nat Protoc* **11**, 2010-2028, doi:10.1038/nprot.2016.103 (2016).

196 Yao, Z. *et al.* Biallelic editing of a lamprey genome using the CRISPR/Cas9 system. **6**, 23496 (2016).

197 Harel *et al.* A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. **160**, 1013-1026 (2015).

198 Kishimoto, K. *et al.* Production of a breed of red sea bream Pagrus major with an increase of skeletal muscle muss and reduced body length by genome editing with CRISPR/Cas9. **495** (2018).

199 Dong, Z. *et al.* Heritable Targeted Inactivation of Myostatin Gene in Yellow Catfish(Pelteobagrus

fulvidraco) Using Engineered Zinc Finger Nucleases. **6**, e28897 (2011).

200 Khalil, K. *et al.* Generation of Myostatin Gene-Edited Channel Catfish (Ictalurus punctatus) via Zygote Injection of CRISPR/Cas9 System. **7**, 7301 (2017).

201 Zhong, Z. *et al.* Targeted disruption ofsp7andmyostatinwith CRISPR-Cas9 results in severe bone defects and more muscular cells in common carp. **6**, 22953 (2016).

202 Villa, F. A. & Gerwick, L. Marine natural product drug discovery: Leads for treatment of inflammation, cancer, infections, and neurological disorders. *Immunopharmacol Immunotoxicol* **32**, 228-237, doi:10.3109/08923970903296136 (2010).

203 Montaser, R. & Luesch, H. Marine natural products: a new wave of drugs? *Future Med Chem* **3**, 1475-1489, doi:10.4155/fmc.11.118 (2011).

204 Tao, L. *et al.* Clustered distribution of natural product leads of drugs in the chemical space as influenced by the privileged target-sites. *Sci Rep* **5**, 9325, doi:srep09325 [pii] 10.1038/srep09325 (2015).

205 Nguyen, B. C. Q., Yoshimura, K., Kumazawa, S., Tawata, S. & Maruta, H. Frondoside A from sea cucumber and nymphaeols from Okinawa propolis: Natural anti-cancer agents that selectively inhibit PAK1 in vitro. *Drug Discov Ther* **11**, 110-114, doi:10.5582/ddt.2017.01011 (2017).

206 Surayot, U., Lee, S. & You, S. Effects of sulfated fucan from the sea cucumber Stichopus japonicus on natural killer cell activation and cytotoxicity. *Int J Biol Macromol* **108**, 177-184, doi:S0141-8130(17)32684-3 [pii] 10.1016/j.ijbiomac.2017.11.102 (2018).

207 Gao, N. *et al.* Preparation and characterization of O-acylated fucosylated chondroitin sulfate from sea cucumber. *Mar Drugs* **10**, 1647-1661, doi:10.3390/md10081647 marinedrugs-10-01647 [pii] (2012).

208 Dyshlovoy, S. A. & Honecker, F. Marine Compounds and Cancer: 2017 Updates. *Mar Drugs* **16**, doi:10.3390/md16020041 (2018).

209 Mayer, A. M. *et al.* The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol Sci* **31**, 255-265, doi:10.1016/j.tips.2010.02.005 (2010).

210 Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. & Prinsep, M. R. Marine natural products. *Nat Prod Rep* **29**, 144-222, doi:10.1039/c2np00090c (2012).

211 Zani, C. L. & Carroll, A. R. Database for Rapid Dereplication of Known Natural Products Using Data from MS and Fast NMR Experiments. *J Nat Prod* **80**, 1758-1766, doi:10.1021/acs.jnatprod.6b01093 (2017).

212 Pomin, V. H. Marine medicinal glycomics. *Front Cell Infect Microbiol* **4**, 5, doi:10.3389/fcimb.2014.00005 (2014).

213 Kersten, R. D. *et al.* Glycogenomics as a mass spectrometry-guided genome-mining method for microbial

glycosylated molecules. *Proc Natl Acad Sci U S A* **110**, E4407-4416, doi:10.1073/pnas.1315492110 (2013).

214 Jeong, Y. *et al.* The dynamic transcriptional and translational landscape of the model antibiotic producer Streptomyces coelicolor A3(2). *Nat Commun* **7**, 11605, doi:10.1038/ncomms11605 (2016).

215 Chen, Y., McClure, R. A., Zheng, Y., Thomson, R. J. & Kelleher, N. L. Proteomics guided discovery of flavopeptins: anti-proliferative aldehydes synthesized by a reductase domain-containing non-ribosomal peptide synthetase. *J Am Chem Soc* **135**, 10449-10456, doi:10.1021/ja4031193 (2013).

216 Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**, 828-837, doi:10.1038/nbt.3597 (2016).

217 Mohimani, H. *et al.* Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol* **13**, 30-37, doi:10.1038/nchembio.2219 (2017).

218 Kersten, R. D. *et al.* A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol* **7**, 794-802, doi:10.1038/nchembio.684 (2011).

219 Chen, Y. *et al.* Gobichelin A and B: Mixed-Ligand Siderophores Discovered Using Proteomics. *Medchemcomm* **4**, 233-238, doi:10.1039/C2MD20232H (2013).

220 Agyei, D., Tsopmo, A. & Udenigwe, C. C. Bioinformatics and peptidomics approaches to the discovery and analysis of food-derived bioactive peptides. *Anal Bioanal Chem* **410**, 3463-3472, doi:10.1007/s00216-018-0974-1

10.1007/s00216-018-0974-1 [pii] (2018).

## Supplementary table 1. The list of the published eukaryote species (to 2018.09)

| Species | Genome Size (Mb) | Contig N50 (Kb) | Scaffold N50 (Kb) | BUSCO (%) | Release Date | Article / Accession Number |
|---|---|---|---|---|---|---|
| **Marine fungus** | | | | | | |
| *Amphiamblys* sp. WSBS2006 | 5.62 | 10.63 | 12.04 | 61.70 | 2016/11/10 | PRJNA321520 |
| *Aspergillus* sp. Z5 | 33.81 | 195.84 | 195.84 | 98.60 | 2015/7/1 | PRJNA285783 |
| *Aureobasidium melanogenum* | 26.12 | 416.08 | 1419.38 | 99.00 | 2017/5/22 | PRJNA376057 |
| *Aureobasidium pullulans* EXF-150 | 29.62 | 353.05 | 1166.85 | 98.30 | 2014/7/8 | PRJNA207874 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Aureobasidium pullulans* P25 | 29.51 | 203.50 | 249.00 | 99.00 | 2018/9/18 | PRJNA479380 |
| *Aureobasidium subglaciale* EXF-2481 | 25.80 | 690.55 | 821.21 | 99.30 | 2014/7/8 | PRJNA161477 |
| *Cadophora malorum* | 47.80 | 331.28 | 388.31 | 99.60 | 2016/5/19 | PRJEB13389 |
| *Cryomyces antarcticus* CCFEE 534 | 24.32 | 4.76 | 4.76 | 80.30 | 2013/12/6 | PRJNA222806 |
| *Cryptococcus gattii* VGI | 17.26 | 67.73 | 71.23 | 94.10 | 2014/10/24 | PRJEB5464 |
| *Cryptococcus gattii* VGII CBS 7750 | 17.48 | 15.44 | 1125.13 | 88.30 | 2013/10/31 | PRJEB4173 |
| *Cryptococcus* sp. 05/00 | 23.80 | 150.92 | 165.60 | 89.30 | 2016/5/19 | PRJEB13495 |
| *Debaryomyces hansenii* | 11.62 | 184.38 | 184.38 | 95.50 | 2016/7/10 | PRJNA323601 |
| *Enterocytozoon hepatopenaei* | 3.25 | 125.01 | 125.01 | 38.60 | 2017/4/6 | PRJNA350317 |
| *Enterospora canceri* | 3.10 | 15.70 | 26.53 | 70.70 | 2017/4/12 | PRJNA316740 |
| *Glaciozyma antarctica* PI12 | 20.03 | 111.76 | 1263.96 | 90.00 | 2018/2/5 | PRJNA202387 |
| *Hamiltosporidium tvaerminnensis* OER-3-3 | 13.27 | 0.44 | 0.44 | 17.00 | 2009/9/30 | PRJNA39213 |
| *Hepatospora eriocheir* | 4.83 | 3.35 | 3.35 | 22.60 | 2017/4/12 | PRJNA312885 |
| *Hortaea werneckii* EXF-2000 | 51.67 | 8.19 | 9.28 | 82.70 | 2013/6/17 | PRJNA87027 |
| *Kwoniella mangroviensis* CBS 10435 | 22.65 | 716.88 | 1966.69 | 91.00 | 2016/7/14 | PRJNA202099 |
| *Kwoniella mangroviensis* CBS 8507 | 22.65 | 521.80 | 1048.16 | 91.00 | 2016/7/14 | PRJNA191223 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Kwoniella mangroviensis* CBS 8886 | 22.87 | 849.36 | 2035.78 | 90.70 | 2016/7/14 | PRJNA191224 |
| *Leucosporidium scottii* | 26.75 | 42.20 | 244.61 | 91.40 | 2018/4/17 | PRJNA378219 |
| *Metschnikowia australis* | 14.35 | 542.23 | 542.23 | 97.30 | 2017/3/29 | PRJNA374844 |
| *Mitosporidium daphniae* | 5.64 | 32.03 | 32.03 | 5.20 | 2014/9/30 | PRJNA243305 |
| *Moesziomyces antarcticus* | 18.11 | 214.27 | 701.21 | 95.90 | 2014/8/21 | PRJDB2910 |
| *Moesziomyces antarcticus* T-34 | 18.07 | 42.64 | 730.47 | 97.30 | 2013/1/25 | PRJDB53 |
| *Mrakia blollopis* | 30.48 | 1718.11 | 1718.11 | 88.20 | 2014/12/9 | PRJDB3253 |
| *Mrakia frigida* | 28.65 | 33.23 | 34.72 | 84.40 | 2015/1/7 | PRJNA268263 |
| *Nectria sp.* B-13 | 62.84 | 229.66 | 1522.10 | 98.20 | 2017/10/24 | PRJNA394176 |
| *Pseudozyma hubeiensis* SY62 | 18.44 | 203.61 | 445.58 | 96.20 | 2013/5/16 | PRJDB993 |
| *Rachicladosporium antarcticum* | 47.41 | 774.46 | 896.82 | 98.30 | 2017/4/3 | PRJNA342238 |
| *Rachicladosporium sp.* CCFEE 5018 | 44.77 | 33.69 | 1358.70 | 98.00 | 2017/4/3 | PRJNA342238 |
| *Saccharomyces jurei* | 11.94 | 738.74 | 738.74 | 96.90 | 2018/8/20 | PRJEB24816 |
| *Sclerotinia glacialis* | 41.10 | 43.85 | 943.42 | 99.70 | 2017/5/31 | PRJNA277973 |
| *Spathaspora arborariae* UFMG-19.1A | 12.87 | 63.79 | 679.21 | 91.00 | 2013/11/20 | PRJNA207280 |
| *Spathaspora boniae* | 12.30 | 104.10 | 104.10 | 98.90 | 2017/4/17 | PRJNA361130 |
| *Spraguea lophii* 42_110 | 4.98 | 4.79 | 5.95 | 80.90 | 2013/7/16 | PRJNA73605 |
| *Spraguea lophii* Celtic Deep | 5.76 | 98.46 | 98.46 | 81.90 | 2016/12/1 | PRJNA269798 |
| *Spraguea lophii* EM120 Celtic Sea | 5.76 | 98.46 | 98.46 | 81.90 | 2016/12/1 | PRJNA269798 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Spraguea lophii* North Atlantic | 5.85 | 95.04 | 95.04 | 82.80 | 2016/12/1 | PRJNA269798 |
| *Spraguea lophii* RA12034 Celtic Sea | 5.80 | 92.21 | 92.21 | 82.50 | 2016/12/1 | PRJNA269798 |
| *Termitomyces* sp. J132 | 67.30 | 41.88 | 268.51 | 93.10 | 2015/8/10 | PRJNA193471 |
| *Trichoderma virens* FT-333 | 38.63 | 167.48 | 173.92 | 97.60 | 2014/12/9 | PRJNA268050 |
| *Umbilicaria pustulata* | 39.23 | 19.94 | 104.30 | 97.60 | 2017/4/14 | PRJEB11664 |
| *Yarrowia lipolytica* NCIM 3589 | 20.47 | 184.64 | 189.64 | 99.00 | 2018/9/14 | PRJNA328405 |
| *Yarrowia lipolytica* NCIM 3590 | 20.01 | 41.52 | 73.21 | 98.20 | 2018/9/14 | PRJNA328405 |
| **Algae and marine plant genome** | | | | | | |
| *Asterionella formosa* | 68.42 | 13.33 | 15.91 | 84.10 | 2017/8/16 | GCA_002217885.1 |
| *Asterochloris* sp. Cgr/DA1pho | 55.82 | 119.01 | 784.87 | 86.10 | 2011/12/7 | https://genome.jgi.doe.gov/portal/Astpho1/download/Astpho1_genomic_scaffolds.fasta.gz |
| *Aureococcus anophagefferens* | 56.66 | 33.66 | 1405.78 | 75.60 | 2011/2/15 | GCF_000186865.1 |
| *Auxenochlorella protothecoides* | 22.92 | 35.09 | 285.54 | 87.10 | 2014/7/23 | GCA_000733215.1 |
| *Auxenochlorella protothecoides* | 22.92 | 35.09 | 285.54 | 87.10 | 2014/8/1 | GCA_000733215.1 |
| *Auxenochlorella pyrenoidosa* | 56.99 | 11.70 | 1392.76 | 71.30 | 2015/11/3 | GCA_001430745.1 |
| *Bathycoccus prasinos* | 15.07 | 663.42 | 955.65 | 78.90 | 2012/11/22 | GCA_002220235.1 |
| *Bathycoccus* sp. TOSAG39-1 | 10.06 | 12.17 | 14.08 | | 2016/12/2 | GCA_900128745.1 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Bigelowiella natans* | 91.41 | 59.46 | 59.46 | 65.10 | 2012/11/2 | GCA_000320545.1 |
| *Blastocystis hominis* | 18.82 | 296.81 | 900.60 | | 2010/7/6 | GCA_000151665.1 |
| *Botryococcus braunii* | 184.38 | 163.33 | 373.00 | 79.20 | 2017/2/22 | GCA_002005505.1 |
| *Breviolum minutum* | 609.48 | 34.31 | 125.23 | | 2013/7/17 | GCA_000507305.1 |
| *Chlamydomonas applanata* | 78.50 | 25.01 | 105.70 | 80.90 | 2016/4/26 | GCA_001662365.1 |
| *Chlamydomonas asymmetrica* | 141.92 | 22.72 | 114.16 | 83.50 | 2016/4/26 | GCA_001662385.1 |
| *Chlamydomonas debaryana* | 120.36 | 9.53 | 27.22 | 67.00 | 2016/4/26 | GCA_001662405.1 |
| *Chlamydomonas eustigma* | 66.63 | 46.21 | 465.13 | 83.80 | 2017/8/31 | GCA_002335675.1 |
| *Chlamydomonas reinhardtii* | 120.40 | 44.36 | 1695.18 | 83.80 | 2007/10/15 | GCA_000002595.3 |
| *Chlamydomonas sphaeroides* | 122.19 | 16.31 | 44.73 | 79.50 | 2016/4/26 | GCA_001662425.1 |
| *Chlorella sorokiniana* | 58.53 | 3818.10 | 4091.73 | 80.80 | 2018/5/17 | GCA_003130725.1 |
| *Chlorella* sp. A99 | 40.93 | 14.75 | 1727.42 | 82.10 | 2018/4/23 | GCA_003063905.1 |
| *Chlorella* sp. ArM0029B | 92.96 | 12.83 | 805.07 | 76.90 | 2018/1/24 | GCA_002896455.3 |
| *Chlorella* sp. NC64A | 46.16 | 27.65 | 1469.61 | 82.50 | 2010/9/16 | GCA_000147415.1 |
| *Chlorella variabilis* | 46.16 | 27.65 | 1469.61 | 82.50 | 2010/9/16 | GCA_000147415.1 |
| *Chlorella vulgaris* | 37.34 | 14.20 | 27.82 | 77.50 | 2015/6/5 | GCA_001021125.1 |
| *Chondrus crispus* | 104.98 | 77.75 | 242.69 | 82.60 | 2013/5/22 | GCF_000091205.1 |
| *Chroomonas mesostigmatica* | 0.70 | 232.70 | 232.70 | | 2012/8/9 | GCA_000286095.1 |
| *Chrysochromulina parva* | 65.76 | 16.05 | 16.05 | 83.50 | 2018/1/17 | GCA_001275005.1 |
| *Chrysochromulina* sp. CCMP291 | 59.07 | 24.05 | 24.11 | 72.60 | 2015/8/26 | GCF_000372725.1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Cladosiphon okamuranus* | 169.73 | 66.17 | 505.95 | 97.90 | 2016/8/18 | GCA_000310025.1 |
| *Coccomyxa* sp. LA000219 | 48.55 | 523.05 | 2254.07 | 88.20 | 2014/12/30 | GCA_000812005.1 |
| *Coccomyxa* sp. SUA001 | 11.75 | 0.57 | 0.57 | | 2015/8/5 | GCA_001244535.1 |
| *Coccomyxa subellipsoidea* C-169 | 48.83 | 1959.569 | 1959.57 | 88.50 | 2012/4/13 | GCA_000258705.1 |
| *Coelastrella* sp. M60 | 80.22 | 9.34 | 9.34 | 91.20 | 2016/4/25 | GCA_001630525.1 |
| *Coelastrella* sp. UTEX B 3026 | 151.55 | 7.07 | 10.71 | 70.00 | 2017/10/18 | GCA_002588565.1 |
| *Cryptomonas paramecium* | 0.49 | 160.19 | 160.19 | | 2011/3/30 | GCA_000194455.1 |
| *Cryptophyceae* sp. CCMP2293 | 534.47 | 5.12 | 439.32 | 55.10 | 2016/9/3 | https://genome.jgi.doe.gov/portal/Crypto2293_1/download/Crypto2293_1_AssemblyScaffolds.fasta.gz |
| *Cyanidioschyzon merolae* | 16.55 | 859.12 | 859.12 | 87.10 | 2007/7/11 | GCA_000341285.1 |
| *Cymbomonas tetramitiformis* | 281.27 | 4.80 | 10.93 | 59.10 | 2015/8/5 | GCA_001247695.1 |
| *Dunaliella salina* | 343.70 | 7.23 | 353.03 | 54.50 | 2017/8/31 | GCA_002284615.1 |
| *Ectocarpus siliculosus* | 195.81 | 32.34 | 3939.08 | 78.90 | 2010/6/24 | GCA_000978595.1 |
| *Emiliania huxleyi* | 167.68 | 29.72 | 404.81 | 70.30 | 2013/5/2 | GCA_002256025.1 |
| *Eudorina* sp. 2006-703-Eu-15 | 184.03 | 300.39 | 564.04 | 83.90 | 2006/7/3 | GCA_003117195.1 |
| *Euglena gracilis* | 41.20 | 0.41 | 0.41 | | 2016/5/6 | GCA_001638955.1 |
| *Fistulifera solaris* | 49.74 | 75.18 | 330.81 | 79.90 | 2017/6/26 | GCA_001750085.1 |
| *Fragilariopsis cylindrus* | 80.54 | 78.23 | 1295.60 | 81.50 | 2016/9/30 | GCA_900291995.1 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Galdieria sulphuraria* | 12.09 | 134.00 | 134.00 | | 2016/8/14 | GCA_003194525.1 |
| *Galdieria sulphuraria* 074W | 13.71 | 93.03 | 172.32 | 85.80 | 2013/1/8 | GCA_001704855.1 |
| *Gonium pectorale* | 148.81 | 16.24 | 1267.14 | 81.20 | 2016/3/9 | GCA_001584585.1 |
| *Gracilariopsis chorda* | 92.18 | 220.27 | 220.27 | 87.10 | 2018/6/6 | GCA_003194525.1 |
| *Gracilariopsis lemaneiformis* | 88.69 | 14.51 | 34.59 | 95.30 | 2018/7/31 | GCA_003346895.1 |
| *Guillardia theta* | 87.15 | 40.45 | 545.81 | 71.30 | 2012/12/5 | GCF_000315625.1 |
| *Helicosporidium* sp. ATCC 50920 | 12.37 | 3.04 | 3.04 | 62.70 | 2014/5/13 | GCA_000690575.1 |
| *Hemiselmis andersenii* | 0.57 | 184.76 | 184.76 | | 2008/4/24 | GCA_000018645.1 |
| *Heterococcus* sp. DN1 | 60.74 | 3.97 | 4.23 | | 2013/11/22 | GCA_000498555.1 |
| *Kappaphycus alvarezii* | 336.72 | 848.97 | 848.97 | 79.70 | 2018/3/9 | GCA_002205965.2 |
| *Klebsormidium nitens* | 104.21 | 72.78 | 134.93 | 91.50 | 2014/6/3 | GCA_000708835.1 |
| *Licmophora abbreviata* | 29.21 | 6.98 | 6.98 | 78.30 | 2018/4/24 | GCF_000150955.2 |
| *Lotharella oceanica* | 0.68 | 194.12 | 207.54 | | 2014/6/5 | GCA_000698435.2 |
| *Micractinium conductrix* | 61.02 | 1210.50 | 1210.50 | 83.50 | 2018/3/21 | GCA_002245815.2 |
| *Micromonas commoda* | 21.11 | 1394.11 | 1394.11 | 83.20 | 2009/4/10 | GCF_000090985.2 |
| *Micromonas pusilla* | 21.96 | 81.16 | 1183.54 | 81.80 | 2009/4/9 | GCF_000151265.2 |
| *Micromonas* sp. ASP10-01a | 19.58 | 11.39 | 22.48 | 79.90 | 2015/4/28 | GCA_001430725.1 |
| *Monoraphidium neglectum* | 69.71 | 9.15 | 15.66 | 57.80 | 2015/2/26 | GCF_000611645.1 |
| *Monoraphidium* sp. 549 | 74.66 | 105.99 | 105.99 | 84.50 | 2017/12/8 | GCA_002814315.1 |
| *Nannochloropsis gaditana* B-31 | 27.59 | 40.86 | 1065.99 | 78.60 | 2014/2/18 | GCA_002838785.1 |
| *Nannochloropsis gaditana* CCMP1894 | 30.87 | 1081.03 | 1141.55 | 83.90 | 2017/12/13 | GCA_001614215.1 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Nannochloropsis gaditana* CCMP526 | 33.99 | 20.80 | 37.64 | 70.30 | 2012/1/4 | GCA_000569095.1 |
| *Nannochloropsis gaditana* CCMP527 | 25.62 | 12.74 | 12.74 | 70.60 | 2016/4/8 | GCA_001614225.1 |
| *Nannochloropsis limnetica* | 33.51 | 2.70 | 2.70 | 54.80 | 2016/4/8 | GCA_000226695.1 |
| *Nannochloropsis oceanica* | 27.64 | 12.33 | 12.33 | 72.30 | 2011/9/29 | GCA_001614235.1 |
| *Nannochloropsis oceanica* OZ-1 | 28.02 | 39.28 | 39.28 | 77.50 | 2016/4/8 | GCA_001870945.1 |
| *Nannochloropsis oceanica* strain IMET1 | 31.50 | 39.28 | 935.20 | 77.90 | 2016/11/9 | GCA_001614245.1 |
| *Nannochloropsis salina* CCMP1776 | 24.36 | 12.19 | 12.19 | 71.00 | 2016/4/8 | GCA_002887195.1 |
| *Ostreococcus lucimarinus* | 13.20 | 708.93 | 708.93 | 84.20 | 2007/4/10 | GCF_000092065.1 |
| *Ostreococcus tauri* | 13.03 | 48.02 | 717.46 | 82.90 | 2014/10/2 | GCF_000214015.3 |
| *Parachlorella kessleri* | 59.19 | 33.74 | 33.89 | 78.90 | 2015/12/19 | GCA_001598975.1 |
| *Pavlovales* sp.CCMP2436 | 165.41 | 5.96 | 252.37 | 65.30 | 2016/8/17 | https://genome.jgi.doe.gov/portal/Pavlov2436_1/download/Pavlov2436_1_AssemblyScaffolds.fasta.gz |
| *Pelagophyceae* sp.CCMP2097 | 85.82 | 14.24 | 186.14 | 77.80 | 2016/4/8 | https://genome.jgi.doe.gov/portal/Pelago2097_1/download/Pelago2097_1_AssemblyScaffolds.fasta.gz |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Phaeodactylum tricornutum* | 27.45 | 417.21 | 945.03 | 81.20 | 2008/12/12 | GCA_900005105.1 |
| *Picochlorum* sp. SENEW3 | 13.39 | 126.22 | 126.22 | 90.80 | 2014/7/28 | GCA_000240725.1 |
| *Picochlorum* sp. soloecismus | 15.25 | 621.32 | 724.71 | 91.40 | 2017/12/8 | GCA_002818215.1 |
| *Porphyra umbilicalis* | 87.89 | 168.86 | 202.02 | 70.00 | 2017/7/27 | GCA_002049455.2 |
| *Porphyridium purpureum* | 19.45 | 20.53 | 20.53 | 87.10 | 2013/5/16 | GCA_000397085.1 |
| *Prorocentrum minimum* | 29.35 | 2.52 | 2.53 | 98.60 | 2016/5/26 | GCA_001652855.1 |
| *Prototheca cutis* | 19.97 | 56.44 | 1409.61 | 88.80 | 2018/6/27 | GCA_002897115.1 |
| *Prototheca stagnorum* | 16.90 | 33.27 | 1107.25 | 86.80 | 2017/11/16 | GCA_002794665.1 |
| *Prototheca wickerhamii* | 27.69 | 7.99 | 31.15 | 88.50 | 2018/6/19 | GCA_003255715.1 |
| *Pseudo-nitzschia multistriata* | 55.16 | 63.53 | 131.46 | 71.30 | 2017/4/25 | GCF_000149405.2 |
| *Raphidocelis subcapitata* | 51.16 | 88.43 | 341.80 | 85.50 | 2018/5/30 | GCA_003203535.1 |
| *Rhizophora apiculata* | 232.06 | 2448.77 | 5420.13 | 92.00 | 2017/9/28 | GCA_900174605.1 |
| *Saccharina japonica* | 543.43 | 36.45 | 252.01 | 70.30 | 2015/4/22 | GCF_000350225.1 |
| *Scenedesmus obliquus* | 107.72 | 104.29 | 186.62 | 84.20 | 2016/11/5 | GCA_900108755.1 |
| *Scenedesmus quadricauda* | 65.35 | 8.09 | 8.09 | | 2017/9/20 | GCA_002317545.1 |
| *Symbiochloris reticulata Africa extracted metagenome v1.0* | 58.57 | 42.27 | 46.45 | 85.50 | 2018/3/22 | https://genome.jgi.doe.gov/portal/SymretAf1/download/SymretAf1_AssemblyScaffolds.fasta.gz |
| *Symbiochloris reticulata Scotland extracted metagenome v1.0* | 59.36 | 47.50 | 55.09 | 85.50 | 2018/3/14 | https://genome.jgi.doe.gov/portal/SymretSc1/do |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | wnload/SymretSc1_Ass emblyScaffolds.fasta.gz |
| *Symbiochloris reticulata Spain extracted metagenome v1.0* | 56.84 | 30.06 | 30.73 | 82.50 | 2017/10/13 | https://genome.jgi.doe.gov/portal/SymretSp1/download/SymretSp1_AssemblyScaffolds.fasta.gz |
| *Symbiochloris reticulata Spain reference genome v1.0* | 58.32 | 34.18 | 37.32 | 84.20 | 2017/11/1 | https://genome.jgi.doe.gov/portal/Dicre1/download/Dicre1_AssemblyScaffolds.fasta.gz |
| *Symbiochloris reticulata Switzerland extracted metagenome v1.0* | 59.35 | 48.55 | 56.72 | 84.80 | 2018/3/21 | https://genome.jgi.doe.gov/portal/SymretSw1/download/SymretSw1_AssemblyScaffolds.fasta.gz |
| *Symbiodinium kawagutii* | 935.07 | 35.63 | 380.91 | | 2015/11/6 | http://web.malab.cn/symka_new/data/Symbiodinium_kawagutii.assembly.935Mb.fa.gz |
| *Symbiodinium microadriaticum* | 808.23 | 18.59 | 573.51 | | 2017/1/6 | GCA_001939145.1 |
| *Symbiodinium minutum* | 609.48 | 34.31 | 125.23 | | 2013/7/17 | GCA_000507305.1 |
| *Symbiodinium* sp. clade A Y106 | 766.66 | 18.02 | 133.47 | 92.60 | 2018/6/22 | GCA_003297005.1 |
| *Symbiodinium* sp. clade C Y103 | 703.70 | 20.31 | 249.18 | 93.30 | 2018/6/22 | GCA_003297045.1 |
| *Tetrabaena socialis* | 135.78 | 5.38 | 145.93 | | 2018/1/18 | GCA_002891735.1 |
| *Tetradesmus obliquus* | 107.72 | 104.29 | 186.62 | 84.20 | 2016/11/5 | GCA_900108755.1 |
| *Thalassiosira oceanica* | 92.19 | 3.64 | 3.64 | 63.70 | 2012/7/25 | GCA_000296195.2 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Thalassiosira pseudonana* | 32.44 | 1267.20 | 1992.43 | 76.60 | 2009/1/16 | GCA_001742925.1 |
| *Trebouxia gelatinosa* | 61.73 | 0.96 | 3512.60 | | 2015/1/16 | GCA_000818905.1 |
| *Trebouxia* sp. TZW2008 | 69.35 | 145.71 | 223.45 | 90.40 | 2017/3/31 | GCA_002118135.1 |
| *uncultured Bathycoccus* | 5.18 | 44.02 | 44.02 | | 2011/10/31 | GCA_000259855.1 |
| *Volvox carteri* | 137.68 | 42.83 | 1491.50 | 84.50 | 2010/7/8 | GCF_000143455.1 |
| *Yamagishiella unicocca* | 140.84 | 543.04 | 543.04 | 85.80 | 2018/4/6 | GCA_003117035.1 |
| *Zostera marina* | 203.91 | 79.96 | 485.58 | 73.20 | 2015/7/22 | GCA_001185155.1 |
| *Zostera muelleri* | 632.07 | 4.90 | 36.73 | 78.90 | 2016/7/3 | http://appliedbioinformatics.com.au/download/Zmu_v1_scaffolds.fa.tar.gz |
| **Marine invertebrate genome** | | | | | | |
| *Acanthaster planci* | 383.86 | 49.68 | 1521.12 | 90.80 | 2018/1/3 | The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest |
| *Acartia tonsa* | 989.16 | 3.24 | 3.61 | 37.22 | 2018/1/3 | Timing of embryonic quiescence determines viability of embryos from the calanoid copepod, Acartia tonsa (Dana) |
| *Acropora digitifera* | 447.50 | 10.92 | 483.56 | 56.03 | 2016/1/15 | Using the Acropora digitifera genome to understand coral responses to environmental change |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Amphimedon queenslandica* | 166.70 | 11.82 | 120.37 | 60.43 | 2010/5/28 | The Amphimedon queenslandica genome and the evolution of animal complexity |
| *Anemonia viridis* | 400.60 | 1.32 | 2.09 | 35.79 | 2018/8/1 | GCA_900234385.1 |
| *Anopheles melas* | 224.16 | 11.31 | 18.10 | 96.42 | 2014/1/17 | GCA_000473525.2 |
| *Anopheles merus* | 288.05 | 48.12 | 1489.98 | 97.55 | 2014/1/17 | GCA_000473845.2 |
| *Aplysia californica* | 927.31 | 9.59 | 917.54 | 85.48 | 2013/5/15 | GCA_000002075.2 |
| *Apostichopus japonicus* | 804.62 | 307.42 | 487.24 | 84.66 | 2017/11/6 | Draft genome of the sea cucumber Apostichopus japonicus and genetic polymorphism among color variants |
| *Apostichopus parvimensis* | 873.09 | 9.59 | 89.13 | 84.46 | 2015/2/27 | The sea cucumber genome provides insights into morphological evolution and visceral regeneration |
| *Bankia setacea* | 3.87 | 176.24 | - | 3.68 | 2016/12/28 | GCA_001922985.1 |
| *Bathymodiolus platifrons* | 1658.19 | 12.60 | 343.34 | 80.88 | 2017/4/5 | Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes |
| *Calanus finmarchicus* | 3.59 | 0.84 | - | 0.00 | 2017/11/1 | GCA_002740975.1 |
| *Calanus glacialis* | 3.94 | 0.86 | - | 0.00 | 2017/11/1 | Mitochondrial genomes of the key zooplankton copepods Arctic |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | Calanus glacialis and North Atlantic Calanus finmarchicus with the longest crustacean non-coding regions |
| *Caligus rogercresseyi* | 398.15 | 1.65 | - | 65.34 | 2015/5/8 | GCA_001005385.1 |
| *Calvadosia cruxmelitensis* | 209.39 | 11.65 | 16.44 | 48.88 | 2018/1/27 | GCA_900245855.1 |
| *Capitella teleta* | 333.28 | 21.93 | 188.40 | 91.92 | 2013/1/25 | Insights into bilaterian evolution from three spiralian genomes |
| *Cassiopea xamachana* | 393.52 | 12.96 | 15.56 | 47.65 | 2018/2/26 | GCA_900291935.1 |
| *Clunio marinus* | 85.49 | 154.80 | 1871.16 | 88.55 | 2016/11/28 | The genomic basis of circadian and circalunar timing adaptations in a midge |
| *Colubraria reticulata* | 67.10 | 0.89 | - | 62.27 | 2016/3/9 | GCA_900004695.1 |
| *Conus tribblei* | 2160.49 | 0.85 | 2.68 | 25.56 | 2015/8/4 | Structural features of conopeptide genes inferred from partial sequences of the Conus tribblei genome |
| *Crassostrea gigas* | 557.74 | 31.24 | 401.69 | 87.22 | 2012/9/19 | The oyster genome reveals stress adaptation and complexity of shell formation |
| *Crassostrea virginica* | 684.74 | 1971.21 | 75944.02 | 86.91 | 2017/9/1 | GCA_002022765.4 |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Enteromyxum leei* | 67.98 | 1.00 | 1.00 | 1.43 | 2015/12/2 | The Multipartite Mitochondrial Genome of Enteromyxum leei |
| *Eriocheir sinensis* | 1549.19 | 45.09 | 490.42 | 31.70 | 2018/7/24 | GCA_003336515.1 |
| *Eucidaris tribuloides* | 2187.26 | 6.63 | 39.19 | 79.14 | 2015/4/14 | GCA_001188425.1 |
| *Eurytemora affinis* | 389.03 | 67.72 | 252.28 | 56.13 | 2017/12/12 | GCA_000591075.2 |
| *Exaiptasia pallida* | 256.13 | 14.40 | 442.15 | 67.89 | 2015/10/28 | The genome of Aiptasia, a sea anemone model for coral symbiosis |
| *Gyrodactylus salaris* | 67.38 | 14.67 | 18.39 | 46.63 | 2014/6/27 | Comparative genomics of flatworms (platyhelminthes) reveals shared genomic features of ecto- and endoparastic neodermata |
| *Haliotis rufescens* | 1498.70 | 283.65 | 1895.87 | 83.95 | 2018/7/27 | GCA_003343065.1 |
| *Hemicentrotus pulcherrimus* | 568.91 | 9.64 | 142.56 | 78.22 | 2018/3/21 | HpBase: A genome database of a sea urchin, Hemicentrotus pulcherrimus |
| *Hydroides elegans* | 1026.05 | 5.98 | 17.33 | 72.09 | 2016/8/11 | Dissection of the Initial Stages of Bacteria-Induced Metamorphosis in a Model Bacterium-Tubeworm Interaction |
| *Intoshia linei* | 0.03 | 41603.07 | 26.27 | 4.70 | 2016/5/10 | The Genome of Intoshia linei Affirms |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | Orthonectids as Highly Simplified Spiralians |
| *Kudoa iwatai* | 31.18 | 39.53 | 40.20 | 9.71 | 2015/10/22 | Genomic insights into the evolutionary origin of Myxozoa within Cnidaria |
| *Lepeophtheirus salmonis* | 790.05 | 9.74 | - | 67.69 | 2015/5/8 | GCA_000181255.2 |
| *Limnoperna fortunei* | 1673.22 | 32.20 | 309.12 | 60.63 | 2018/5/16 | A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, Limnoperna fortunei |
| *Limulus polyphemus* | 1828.27 | 11.44 | 254.09 | 79.35 | 2014/1/3 | The Draft Genome and Transcriptome of the Atlantic Horseshoe Crab, Limulus polyphemus. |
| *Lingula anatina* | 406.31 | 56.06 | 460.09 | 88.65 | 2018/1/26 | The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization |
| *Lottia gigantea* | 359.51 | 96.03 | 1870.06 | 88.14 | 2012/12/20 | Insights into bilaterian evolution from three spiralian genomes |
| *Lytechinus variegatus* | 1061.20 | 9.66 | 46.35 | 72.09 | 2015/3/11 | Genomes of Strongylocentrotus franciscanus and |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | Lytechinus variegatus: are there any genomic explanations for the two order of magnitude difference in the lifespan of sea urchins? |
| *Macrostomum lignano* | 764.41 | 215.28 | 246.20 | 82.52 | 2017/8/24 | Efficient transgenesis and annotated genome sequence of the regenerative flatworm model Macrostomum lignano |
| *Mizuhopecten yessoensis* | 987.59 | 65.01 | 803.63 | 84.97 | 2017/6/12 | Scallop genome provides insights into evolution of bilaterian karyotype and development |
| *Mnemiopsis leidyi* | 155.87 | 11.91 | 187.31 | 57.67 | 2011/9/19 | The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution |
| *Modiolus philippinarum* | 2629.56 | 18.39 | 100.16 | 73.52 | 2017/4/5 | Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes |
| *Mytilus galloprovincialis* | 1500.15 | 2.63 | 2.93 | 9.10 | 2017/3/16 | A First Insight into the Genome of the Filter- |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | Feeder Mussel Mytilus galloprovincialis |
| *Nematostella vectensis* | 356.61 | 19.84 | 472.59 | 78.94 | 2007/8/22 | The diversity of C-type lectins in the genome of a basal metazoan, Nematostella vectensis |
| *Notospermus geniculatus* | 858.60 | 22.60 | 239.24 | 88.24 | 2017/10/23 | Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. |
| *Octopus bimaculoides* | 2338.19 | 5.53 | 475.18 | 77.91 | 2015/8/18 | The octopus genome and the evolution of cephalopod neural and morphological novelties |
| *Oithona nana* | 1828.27 | 11.44 | 254.09 | 76.38 | 2017/2/17 | GCA_900157175.1 |
| *Ophionereis fasciata* | 1184.53 | 0.20 | 0.48 | 4.60 | 2017/2/27 | GCA_900067615.1 |
| *Ophiothrix spiculata* | 2764.32 | 6.47 | 72.78 | 81.19 | 2015/2/3 | GCA_000969725.1 |
| *Orbicella faveolata* | 485.55 | 12.47 | 1162.45 | 70.86 | 2017/3/20 | GCA_002042975.1 |
| *Parhyale hawaiensis* | 85.01 | 38.62 | 400.61 | 74.74 | 2018/6/22 | The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion |
| *Patiria miniata* | 811.03 | 9.47 | 52.61 | 90.18 | 2012/8/13 | GCA_000285935.1 |
| *Patiriella regularis* | 949.33 | 0.22 | 0.56 | 1.74 | 2017/2/10 | GCA_900067625.1 |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Penaeus japonicus* | 2752.56 | 10.44 | 20228.73 | 18.81 | 2017/9/12 | Genomic resources and comparative analyses of two economical penaeid shrimp species, Marsupenaeus japonicus and Penaeus monodon |
| *Penaeus monodon* | 1660.27 | 0.70 | 0.91 | 18.81 | 2017/9/12 | The genome and occlusion bodies of marine Penaeus monodon nudivirus (PmNV, also known as MBV and PemoNPV) suggest that it should be assigned to a new nudivirus genus that is distinct from the terrestrial nudiviruses |
| *Phoronis australis* | 498.44 | 68.15 | 655.06 | 91.00 | 2017/10/23 | GCA_002633005.1 |
| *Pinctada martensii* | 990.98 | 21.52 | 59032.46 | 75.05 | 2017/7/20 | The pearl oyster Pinctada fucata martensii genome and multi-omic analyses provide insights into biomineralization. |
| *Platynothrus peltifer* | 100.53 | 1.24 | 1.57 | 32.82 | 2015/5/6 | GCA_000988905.1 |
| *Pleurobrachia bachei* | 156.12 | 6.13 | 20.63 | 60.22 | 2014/5/21 | The ctenophore genome and the evolutionary |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | origins of neural systems |
| *Porites rus* | 470.01 | 5.32 | 137.19 | 52.04 | 2018/5/28 | GCA_900290455.1 |
| *Priapulus caudatus* | 511.74 | 10.62 | 209.73 | 80.16 | 2015/11/19 | GCA_000485595.2 |
| *Ptychodera flava* | 1228.69 | 13.43 | 196.30 | 76.38 | 2015/12/1 | Hemichordate genomes and deuterostome origins |
| *Renilla reniformis* | 131.55 | 1.74 | 1.84 | 21.68 | 2017/4/20 | GCA_900177555.1 |
| *Saccoglossus kowalevskii* | 775.84 | 10.07 | 245.82 | 75.46 | 2009/9/9 | Hemichordate genomes and deuterostome origins |
| *Sphaeromyxa zaharoni* | 173.59 | 4.47 | 4.47 | 4.81 | 2015/12/2 | Genomic insights into the evolutionary origin of Myxozoa within Cnidaria |
| *Strigamia maritima* | 176.21 | 24.75 | 139.45 | 89.47 | 2011/12/22 | GCA_000239455.1 |
| *Strongylocentrotus purpuratus* | 990.92 | 16.79 | 419.55 | 83.23 | 2015/3/10 | The Genome of the Sea Urchin Strongylocentrotus purpuratus |
| *Stylophora pistillata* | 400.12 | 20.60 | 457.45 | 73.21 | 2017/10/17 | Comparative analysis of the genomes of Stylophora pistillata and Acropora digitifera provides evidence for extensive differences between species of corals |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Trichoplax adhaerens* | 105.63 | 204.19 | 5978.66 | 69.12 | 2008/6/17 | The Trichoplax genome and the nature of placozoans |
| **Fish genomes** | | | | | | |
| *Acanthochaenus luetkenii* | - | - | - | - | - | PRJEB12469 |
| *Acanthochromis polyacanthus* | 992 | 8 | 334 | 95.7 | - | GCA_002109545.1 |
| *Acanthopagrus schlegelii* | 688 | 17.2 | 7600 | - | 2018/2 | Draft genome of the protandrous Chinese black porgy,*Acanthopagrus schlegelii* |
| *Ageneiosus marmoratus* | - | - | - | - | - | PRJNA427361 |
| *Amphilophus citrinellus* | 845 | 19 | 1216 | 97.1 | - | GCA_000751415.1 |
| *Amphiprion ocellaris* | 816 | 323 | 401 | 97.3 | 2018/1 | Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly |
| *Amphiprion percula* | - | - | - | - | - | PRJNA436093 |
| *Anabas testudineus* | - | - | - | - | - | PRJEB25768 |
| *Anguilla anguilla* | 1019 | 1.8 | 59.7 | 67.9 | 2017/8 | GCA_000695075.1 |
| *Anguilla japonica* | 1151 | 6 | 472 | 87.6 | - | GCA_002723815.1 |
| *Anguilla rostrata* | 1413 | 6.3 | 86.6 | 86.7 | - | GCA_001606085.1 |
| *Anoplopoma fimbria* | 699 | 4.9 | 5.1 | 53.5 | - | GCA_000499045.1 |
| *Antennarius striatus* | - | - | - | - | - | PRJEB12469 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Arapaima gigas* | - | - | - | - | - | PRJEB22808 |
| *Arctogadus glacialis* | - | - | - | - | - | PRJEB12469 |
| *Astatotilapia calliptera* | 883 | 12523 | 4534 | 97.3 | - | GCA_900246225.1 |
| *Astyanax mexicanus* | - | - | - | - | - | PRJNA237016 |
| *Austrofundulus limnaeus* | - | - | - | - | - | PRJNA294420 |
| *Bathygadus melanobranchus* | - | - | - | - | - | PRJEB12469 |
| *Benthosema glaciale* | - | - | - | - | - | PRJEB12469 |
| *Beryx splendens* | - | - | - | - | - | PRJEB12469 |
| *Betta splendens* | 465.24 | 9.01 | 949.03 | - | 2018/7 | Chromosome-level reference genome of the Siamese fighting fish *Betta splendens*, a model species for the study of aggression. |
| *Boleophthalmus pectinirostris* | 956 | 20 | 2376 | 94 | - | Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes |
| *Boreogadus saida* | - | - | - | - | - | PRJEB12469 |
| *Borostomias antarcticus* | - | - | - | - | - | PRJEB12469 |
| *Bregmaceros cantori* | - | - | - | - | - | PRJEB12469 |
| *Brosme brosme* | - | - | - | - | - | PRJEB12469 |
| *Brotula barbata* | - | - | - | - | - | PRJEB12469 |
| *Callorhinchus milii* | 937 | 46.6 | 4500 | - | 2014/1 | Elephant shark genome provides uniquein sights into gnathostome evolution |
| *Carapus acus* | - | - | - | - | - | PRJEB12469 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Carassius auratus* | - | - | - | - | - | PRJNA487739 |
| *Chaenocephalus aceratus* | - | - | - | - | - | PRJEB12469 |
| *Channa argus* | 615.3 | 81.4 | 4500 | - | 2017/3 | Draft genome of the Northern snakehead, *Channa argus* |
| *Chatrabus melanurus* | - | - | - | - | - | PRJEB12469 |
| *Chromis chromis* | - | - | - | - | - | PRJEB12469 |
| *Clupea harengus* | 808 | 22 | 1861 | 96.2 | 2016/3 | The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing |
| *Coryphaenoides rupestris* | 829 | 12.8 | 159 | 82.8 | 2018/3 | Genomics of habitat choice and adaptive evolution in a deep-sea fish |
| *Cottus rhenanus* | 563 | - | - | 55.3 | - | GCA_001455555.1 |
| *Ctenopharyngodon idellus* | 1070 | 40.8 | 6457 | - | 2015/5 | The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution adaption |
| *Cynoglossus semilaevis* | 470 | 27 | 20010.6 | 97.1 | 2014/2 | Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | adaptation to a benthic lifestyle |
| *Cyprinodon nevadensis* | - | - | - | - | - | PRJNA254053 |
| *Cyprinodon variegatus* | - | - | - | - | - | PRJNA308224 |
| *Cyprinus carpio* | 1713 | 59 | 7828 | 83.4 | 2014/9 | Genome sequence and genetic diversity of the common carp, *Cyprinus carpio* |
| *Cyttopsis rosea* | - | - | - | - | - | PRJEB12469 |
| *Danio rerio* | 1679 | 854 | 52186 | 95.9 | 2013/4 | The zebrafish reference genome sequence and its relationship to the human genome |
| *Danionella dracula* | - | - | - | - | - | PRJEB27320 |
| *Dicentrarchus labrax* | 676 | 53 | 26439 | 95.4 | - | GCA_000689215.1 |
| *Eptatretus burgeri* | - | - | - | - | - | PRJEB21290 |
| *Esox lucius* | 904 | 68 | 32939 | 97 | - | GCA_000721915.3 |
| *Fundulus heteroclitus* | - | - | - | - | - | PRJNA286680 |
| *Gadiculus argenteus* | - | - | - | - | - | PRJEB12469 |
| *Gadus chalcogrammus* | - | - | - | - | - | PRJEB12469 |
| *Gadus morhua* | 650 | 2.8 | 688 | - | 2011/8 | The genome sequence of Atlantic cod reveals a unique immune system |
| *Gambusia affinis* | - | - | - | - | - | PRJNA386810 |
| *Gasterosteus aculeatus* | 463 | 83.2 | 1080 | 93 | 2012/4 | The genomic basis of adaptive evolution in threespine sticklebacks |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Glyptosternon maculatum* | 662.34 | 993.67 | 20900 | - | 2018/8 | Draft genome of *Glyptosternon maculatum*, an endemic fish from Tibet Plateau |
| *Guentherus altivela* | - | - | - | - | - | PRJEB12469 |
| *Haplochromis burtoni* | 831 | 21.9 | 1194 | 97.1 | - | GCA_000239415.1 |
| *Helostoma temminkii* | - | - | - | - | - | PRJEB12469 |
| *Hippocampus comes* | 494 | 39.5 | 2034 | 95.1 | 2017/4 | The seahorse genome and the evolution of its specialized morphology |
| *Hippocampus erectus* | 494 | 138 | 4145 | - | 2017/6 | Draft genome of the lined seahorse, *Hippocampus erectus* |
| *Holocentrus rufus* | - | - | - | - | - | PRJEB12469 |
| *Hucho hucho* | - | - | - | - | - | PRJNA475010 |
| *Ictalurus punctatus* | 784 | 77.2 | 27425 | 96.3 | - | GCA_001660625.1 |
| *Ictalurus punctatus* | - | - | - | - | 2011/12 | The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts |
| *Kryptolebias marmoratus* | 680 | 43 | 2229 | 96.6 | - | GCA_001663955.1 |
| *Labeotropheus fuelleborni* | - | - | - | - | - | PRJNA29479 |
| *Labrus bergylta* | 805 | 704 | 795 | 95.9 | 2018/3 | Loss of stomach, loss of appetite? Sequencing of the ballan wrasse (*Labrus bergylta*) |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | genome and intestinal transcriptomic profiling illuminate the evolution of loss of stomach function in fish |
| *Laemonema laureysi* | - | - | - | - | - | PRJEB12469 |
| *Lampris guttatus* | - | - | - | - | - | PRJEB12469 |
| *Lamprogrammus exutus* | - | - | - | - | - | PRJEB12469 |
| *Larimichthys crocea* | 678 | 68 | 1034 | 96.4 | 2015/4 | Genome Sequencing of the Perciform Fish *Larimichthys crocea* Provides Insights into Molecular and Genetic Mechanisms of Stress Adaptation |
| *Larimichthys crocea* | 728 | 25.717 | 495.7 | 96.9 | 2014/11 | The draft genome of the large yellow croaker reveals well-developed innate immunity |
| *Lateolabrax maculatus* | 668 | 31 | 1040 | - | 2018/9 | Chromosome-level genome assembly of the spotted sea bass, *Lateolabrax maculatus* |
| *Lates calcarifer* | 605 | 1921.8 | 1921.8 | 97.3 | - | GCA_900066035.1 |
| *Latimeria chalumnae* | 2860 | 12.7 | 924 | 87.4 | 2013/4 | The African coelacanth genome provides insights into tetrapod evolution |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Lepisosteus oculatus* | 945 | 68.3 | 50348 | 94.3 | 2016/3 | The spotted gar genome illuminates vertebrate evolution and facilitates humanteleost comparisons |
| *Lesueurigobius sanzi* | - | - | - | - | - | PRJEB12469 |
| *Lethenteron camtschaticum* | - | - | - | - | - | PRJNA237018 |
| *Leuciscus waleckii* | 725 | 37.4 | 21959 | 70.1 | - | GCA_900092035.1 |
| *Leucoraja erinacea* | - | - | - | - | - | PRJNA60893 |
| *Lota lota* | - | - | - | - | - | PRJEB12469 |
| *Maccullochella peelii* | 633 | 52.687 | 110 | 93.6 | 2017/8 | De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read |
| *Macrourus berglax* | - | - | - | - | - | PRJEB12469 |
| *Malacocephalus occidentalis* | - | - | - | - | - | PRJEB12469 |
| *Mastacembelus armatus* | - | - | - | - | - | PRJEB25769 |
| *Maylandia zebra* | - | - | - | - | - | PRJNA60369 |
| *Mchenga conophoros* | - | - | - | - | - | PRJNA29477 |
| *Megalobrama amblycephala* | 111 | 49 | 839 | - | 2017/7 | The draft genome of blunt snout bream |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | (*Megalobrama amblycephala*) reveals the development of intermuscular bone and adaptation to herbivorous diet |
| *Melanochromis auratus* | - | - | - | - | - | PRJNA29481 |
| *Melanogrammus aeglefinus* | 652 | 77 | 209 | 92.3 | - | GCA_900291075.1 |
| *Melanonus zugmayeri* | - | - | - | - | - | PRJEB12469 |
| *Merlangius merlangus* | - | - | - | - | - | PRJEB12469 |
| *Merluccius capensis* | - | - | - | - | - | PRJEB12469 |
| *Merluccius merluccius* | - | - | - | - | - | PRJEB12469 |
| *Merluccius polli* | - | - | - | - | - | PRJEB12469 |
| *Micropterus floridanus* | 1001 | 11 | 11 | 71.9 | - | GCA_002592385.1 |
| *Miichthys miiuy* | 619 | 73 | 1145 | 91.2 | - | GCA_001593715.1 |
| *Mola mola* | 639 | 23 | 8767 | 96.7 | 2016/9 | The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast grow rate |
| *Molva molva* | - | - | - | - | - | PRJEB12469 |
| *Monocentris japonicus* | - | - | - | - | - | PRJEB12469 |
| *Monopterus albus* | - | - | - | - | - | PRJNA380265 |
| *Monopterus albus* | 684 | 22.239 | 2106 | 96.5 | 2018/4 | Chromosome-scale assembly of the *Monopterus* genome |
| *Mora moro* | - | - | - | - | - | PRJEB12469 |
| *Morone saxatilis* | 585 | 17 | 29 | 79.3 | - | GCA_001663605.1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Muraenolepis marmoratus* | - | - | - | - | - | PRJEB12469 |
| *Myoxocephalus scorpius* | - | - | - | - | - | PRJEB12469 |
| *Myripristis jacobus* | - | - | - | - | - | PRJEB12469 |
| *Neolamprologus brichardi* | 847 | 13 | 4430 | 92.7 | - | GCA_000239395.1 |
| *Neoniphon sammara* | - | - | - | - | - | PRJEB12469 |
| *Nothobranchius furzeri* | 1242 | 15 | 57367 | 93.5 | 2015/12 | Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish |
| *Nothobranchius kuhntae* | - | - | - | - | - | PRJNA33401 |
| *Notothenia coriiceps* | 637 | 11.5 | 218 | 70.7 | 2014/11 | The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment |
| *Oncorhynchus kisutch* | 2369 | 43.7 | 50431 | 91.2 | - | GCA_002021735.1 |
| *Oncorhynchus mykiss* | 2179 | 13.8 | 71056 | 90.2 | 2014/4 | The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates |
| *Oncorhynchus tshawytscha* | 2425 | 85 | 47556 | 85.5 | - | GCA_002872995.1 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Oreochromis niloticus* | 1010 | 3090 | 37007 | 98.1 | 2017/5 | A high-quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions |
| *Oryzias latipes* | 744 | 3516 | 32853 | 95.5 | 1905/7/9 | Centromere evolution and CpG methylation during vertebrate speciation |
| *Oryzias melastigma* | 779 | 28 | 23737 | 97.3 | - | GCA_002922805.1 |
| *Osmerus eperlanus* | - | - | - | - | - | PRJEB12469 |
| *Pagrus major* | - | - | - | - | - | PRJDB5593 |
| *Pampus argenteus* | 550 | 0.499 | 1.58 | - | 1905/7/7 | Draft genome sequence of the silver pomfret fish, *Pampus argenteus* |
| *Parablennius parvicornis* | - | - | - | - | - | PRJEB12469 |
| *Parachaenichthys charcoti* | 795 | 6.145 | 178.362 | - | 1905/7/9 | Draft genome of the Antarctic dragonfish, *Parachaenichthys charcoti* |
| *Paralichthys olivaceus* | 546 | 30.5 | 23206 | 97.5 | 1905/7/8 | The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Paramormyrops kingsleyae* | 799 | 32 | 1731 | 96.6 | 2017/12 | The Genome and Adult Somatic Transcriptome of the Mormyrid Electric Fish *Paramormyrops kingsleyae* |
| *Parasudis fraserbrunneri* | - | - | - | - | - | PRJEB12469 |
| *Perca fluviatilis* | - | - | - | - | - | PRJNA450919 |
| *Percopsis transmontana* | - | - | - | - | - | PRJEB12469 |
| *Periophthalmodon schlosseri* | 679 | 16 | 39 | 83.3 | - | Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes |
| *Periophthalmus magnuspinnatus* | 701 | 28 | 296 | 93.9 | - | Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes |
| *Petromyzon marinus* | 1130 | 170 | 12000 | - | 2018/2 | The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution |
| *Phycis blennoides* | - | - | - | - | - | PRJEB12469 |
| *Phycis phycis* | - | - | - | - | - | PRJEB12469 |
| *Pimephales promelas* | 1219 | 3.8 | 60 | 77.1 | - | GCA_000700965.1 |
| *Poecilia formosa* | 714 | 57 | 1570 | - | 2018/2 | Clonal polymorphism and high heterozygosity |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | in the celibate genome of the Amazon molly |
| *Poecilia latipinna* | - | - | - | - | - | PRJNA305623 |
| *Poecilia mexicana* | - | - | - | - | - | PRJNA305619 |
| *Poecilia reticulata* | 731 | 35 | 31413 | 95.8 | - | GCA_000633615.2 |
| *Pollachius virens* | - | - | - | - | - | PRJEB12469 |
| *Polymixia japonica* | - | - | - | - | - | PRJEB12469 |
| *Protosalanx hyalocranius* | 536 | 17.2 | 1163 | - | 2017/2 | Whole genome sequencing of Chinese clearhead icefish, *Protosalanx hyalocranius* |
| *Pseudochromis fuscus* | - | - | - | - | - | PRJEB12469 |
| *Pseudopleuronectes yokohamae* | - | - | - | - | - | PRJDB3259 |
| *Pundamilia nyererei* | 830 | 22 | 2525 | 96.9 | - | GCA_000239375.1 |
| *Pygocentrus nattereri* | - | - | - | - | - | PRJNA331139 |
| *Regalecus glesne* | - | - | - | - | - | PRJEB12469 |
| *Rhamphochromis esox* | - | - | - | - | - | PRJNA29485 |
| *Rhincodon typus* | 2931 | 144 | 144 | 78.5 | 2017/7 | Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828 |
| *Rondeletia loricata* | - | - | - | - | - | PRJEB12469 |
| *Salmo salar* | 3412 | 35.8 | 63420 | 92.4 | 2016/4 | The Atlantic salmon genome provides |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | insights into rediploidization |
| *Salvelinus alpinus* | 2169 | 44 | 36001 | 88.1 | - | GCA_002910315.2 |
| *Scartelaos histophorus* | 695 | 8 | 15 | 76.3 | - | Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes |
| *Scleropages formosus* | 777 | 62.8 | 5970 | 95.9 | 2016/4 | The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts |
| *Scophthalmus maximus* | - | - | - | - | - | PRJEB11743 |
| *Sebastes aleutianus* | - | - | - | - | - | PRJNA229179 |
| *Sebastes minor* | - | - | - | - | - | PRJNA236304 |
| *Sebastes nigrocinctus* | 746 | 12 | 116 | 80.6 | - | GCA_001910765.2 |
| *Sebastes norvegicus* | - | - | - | - | - | PRJEB12469 |
| *Sebastes rubrivinctus* | 756 | 13 | 30 | 81.1 | - | GCA_000475215.1 |
| *Sebastes steindachneri* | - | - | - | - | - | PRJNA236323 |
| *Selene dorsalis* | - | - | - | - | - | PRJEB12469 |
| *Seriola dumerili* | 677 | 183 | 5812 | 98.5 | - | GCA_002260705.1 |
| *Seriola lalandi* | - | - | - | - | - | PRJNA314866 |
| *Seriola lalandi dorsalis* | 732 | 51 | 1269 | 97.6 | - | GCA_002814215.1 |
| *Seriola quinqueradiata* | 639 | 872 | 5610 | 96.9 | 2018/7 | GCA_002217815.1 |
| *Seriola rivoliana* | 40.76 | 740 | 9509 | 97.1 | - | GCA_002994505.1 |
| *Simochromis diagramma* | - | - | - | - | - | PRJEB26682 |
| *Sinocyclocheilus anshuiensis* | 1632 | 17 | 1284 | 97.2 | 2016/1 | The Sinocyclocheilus cavefish genome |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | provides insights into cave adaptation |
| *Sinocyclocheilus grahami* | 1750 | 29 | 1156 | 94.9 | - | GCA_001515645.1 |
| *Sinocyclocheilus rhinocerous* | 1655 | 18 | 945 | 96.8 | - | GCA_001515625.1 |
| *Sparus aurata* | - | - | - | - | - | PRJNA416845 |
| *Spondyliosoma cantharus* | - | - | - | - | - | PRJEB12469 |
| *Squalius pyrenaicus* | - | - | - | - | - | PRJEB9465 |
| *Stegastes partitus* | - | - | - | - | - | PRJNA251741 |
| *Stylephorus chordatus* | - | - | - | - | - | PRJEB12469 |
| *Symphodus melops* | 614 | 461 | 461 | 93.8 | - | GCA_002819105.1 |
| *Syngnathus scovelli* | 307 | 32.24 | 640.41 | - | 2016/12 | The genome of the Gulf pipefish enables understanding of evolutionary innovations |
| *Takifugu flavidus* | 378 | 1 | 315 | 78.7 | - | GCA_000400755.1 |
| *Takifugu rubripes* | 391 | 49 | 11516 | 95.8 | 2002/12 | Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes |
| *Tetraodon nigroviridis* | 342 | 28 | 734 | 87.2 | 2004/1 | Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Thunnus albacares* | - | - | - | - | - | PRJEB12469 |
| *Thunnus orientalis* | 800 | 7.5 | 136 | - | 2013/6 | Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna |
| *Thunnus thynnus* | - | - | - | - | - | PRJNA432036 |
| *Trachinotus ovatus* | - | - | - | - | - | PRJEB22654 |
| *Trachyrincus murrayi* | - | - | - | - | - | PRJEB12469 |
| *Trachyrincus scabrus* | - | - | - | - | - | PRJEB12469 |
| *Trisopterus minutus* | - | - | - | - | - | PRJEB12469 |
| *Typhlichthys subterraneus* | - | - | - | - | - | PRJEB12469 |
| *Xiphophorus couchianus* | - | - | - | - | - | PRJNA290781 |
| *Xiphophorus hellerii* | - | - | - | - | - | PRJNA290782 |
| *Xiphophorus maculatus* | 704 | 22 | 1110 | - | 2013/3 | The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits |
| *Zeus faber* | - | - | - | - | - | PRJEB12469 |
| **Marine tetrapods** | | | | | | |
| *Anas platyrhynchos* | 1136.42 | 88.03 | 74988.51 | 93.8 | 2017/11 | GCA_002743455.1 |
| *Anser brachyrhynchus* | 1116.99 | 97.46 | 4974.39 | 95.0 | 2018/03 | First de novo whole genome sequencing and assembly of the pink-footed goose |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Aptenodytes forsteri* | 1254.34 | 31.73 | 5071.6 | 96.7 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Arctocephalus gazella* | 2313 | 198.98 | 6207.32 | 88.4 | 2016/01 | A draft fur seal genome provides insights into factors affecting SNP validation and how to mitigate them |
| *Balaena mysticetus* | 2310 | 34.8 | 877 | 74.2 | 2015/01 | Insights into the evolution of longevity from the bowhead whale genome |
| *Balaenoptera acutorostrata* | 2431.69 | 12.84 | 22.57 | 90.7 | 2013/11 | Minke whale genome and marine adaptation in cetaceans |
| *Balaenoptera bonaerensis* | 2234.64 | 1.74 | 20.08 | 48.8 | 2015/02 | Marine adaptation and the evolution of smell and taste in whales |
| *Callorhinus ursinus* | 2706.87 | 133.02 | 31506.8 | 84.4 | 2018/06 | GCF_003265705.1 |
| *Charadrius vociferus* | 1219.85 | 39.27 | 3657.05 | 96.3 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Chelonia mydas* | 2208.4104 | 29.24 | 3864.108 | 97.1 | 2013/04 | The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | evolution of the turtle-specific body plan |
| *Crocodylus porosus* | 2049.5363 | 34.073 | 84437.66 | 96.8 | 2014/12 | Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs |
| *Delphinapterus leucas* | 2358.51 | 159.14 | 1959 | 89.5 | 2017/12 | The Genome of the Beluga Whale (Delphinapterus leucas) |
| *Egretta garzetta* | 1206.5 | 29.02 | 3067.16 | 95.5 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Enhydra lutris kenyoni* | 2455.28 | 244.53 | 38751.46 | 92.9 | 2017/12 | The Genome of the Northern Sea Otter (Enhydra lutris kenyoni) |
| *Eschrichtius robustus* | 2886.22 | 2.66 | 10.67 | 82.8 | 2017/12 | De novo assembling and primary analysis of genome and transcriptome of gray whale Eschrichtius robustus |
| *Fulmarus glacialis* | 1141.4 | 25.93 | 47.21 | 85.5 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Gavia stellata* | 1129.69 | 24.32 | 45.52 | 82.6 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Haliaeetus albicilla* | 1133.55 | 25.14 | 57.31 | 85.7 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Haliaeetus leucocephalus* | 1178.41 | 105.49 | 9145.5 | 98.4 | 2014/08 | GCF_000737465.1 |
| *Leptonychotes weddellii* | 3156.89 | 23.66 | 904.03 | 67.6 | 2013/03 | GCF_000349705.1 |
| *Limosa lapponica* | 1034.77 | 21.26 | 283.01 | 60.6 | 2017/12 | GCA_002844005.1 |
| *Lipotes vexillifer* | 2429.21 | 30.01 | 2270 | 82.1 | 2013/10 | Baiji genomes reveal low genetic variability and new insights into secondary marine adaptations |
| *Neomonachus schauinslandi* | 2400.93 | 112.7 | 2951.86 | 83.9 | 2017/04 | Improved de novo Genome Assembly: Linked-Read Sequencing Combined with Optical Mapping Produce a High Quality Mammalian Genome at Relatively Low Cost |
| *Neophocaena asiaeorientalis* | 2295 | 26.73 | 6334.54 | 93.7 | 2018/04 | Population genomics of finless porpoises reveal an incipient cetacean |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | species adapted to freshwater |
| *Odobenus rosmarus* | 2400.15 | 89.95 | 2616.78 | 83.9 | 2015/01 | Convergent evolution of the genomes of marine mammals |
| *Orcinus orca* | 2373 | 70.3 | 12735.09 | 94.8 | 2015/01 | Convergent evolution of the genomes of marine mammals |
| *Pelecanus crispus* | 1160.92 | 21.68 | 43.36 | 84.1 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Phaethon lepturus* | 1152.96 | 22.94 | 47.9 | 87.6 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Phalacrocorax auritus* | 1246.05 | 97.55 | 1990.98 | 98.0 | 2017/06 | A genetic signature of the evolution of loss of flight in the Galapagos cormorant |
| *Phalacrocorax brasilianus* | 1346.19 | 35.2 | 86.91 | 88.1 | 2017/06 | A genetic signature of the evolution of loss of flight in the Galapagos cormorant |
| *Phalacrocorax carbo* | 1138.97 | 17.34 | 48.43 | 82.8 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |

 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Phalacrocorax harrisi* | 1202.99 | 100.24 | 4648.44 | 98.5 | 2017/06 | A genetic signature of the evolution of loss of flight in the Galapagos cormorant |
| *Phalacrocorax pelagicus* | 1210.66 | 67.73 | 1719.9 | 97.5 | 2017/06 | A genetic signature of the evolution of loss of flight in the Galapagos cormorant |
| *Philomachus pugnax* | 1229.09 | 109.24 | 10060.04 | 98.6 | 2015/11 | GCA_001431845.1 |
| *Phocoena phocoena* | 2441 | 2.77 | 28296.39 | 91.3 | 2018/07 | High-quality whole-genome sequence of an abundant Holarctic odontocete, the harbour porpoise (Phocoena phocoena) |
| *Physeter macrocephalus* | 2469.59 | 43.83 | 122.18 | 94.1 | 2018/07 | GCA_900411695.1 |
| *Podiceps cristatus* | 1134.92 | 17.41 | 30.08 | 69.6 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Pygoscelis adeliae* | 1216.61 | 22.19 | 5118.89 | 94.7 | 2014/12 | Comparative genomics reveals insights into avian genome evolution and adaptation |
| *Pygoscelis antarcticus* | 1209.83 | 22.34 | 5226.39 | 94.3 | 2018/06 | GCA_003264595.1 |
| *Pygoscelis papua* | 1209.81 | 22.34 | 5226.25 | 94.6 | 2018/06 | GCA_003264615.1 |
| *Spheniscus humboldti* | 1209.85 | 22.34 | 5226.26 | 94.1 | 2018/06 | GCA_003264545.1 |
| *Spheniscus magellanicus* | 1209.85 | 22.34 | 5226.29 | 94.5 | 2018/06 | GCA_003264715.1 |
| *Spheniscus mendiculus* | 1209.85 | 22.34 | 5226.23 | 94.2 | 2018/06 | GCA_003264655.1 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0156.v1



| | | | | | | |
|---|---|---|---|---|---|---|
| *Trichechus manatus latirostris* | 3104 | 37.75 | 14442.68 | 93.6 | 2015/01 | Convergent evolution of the genomes of marine mammals |
| *Tursiops aduncus* | 2504 | 214.09 | 1254.76 | 87.3 | 2018/08 | Population genomic analysis reveals contrasting demographic changes of two closely related dolphin species in the last glacial |
| *Tursiops truncatus* | 2386 | 30.68 | 26997.44 | 93.5 | 2018/07 | New de novo assembly of the Atlantic bottlenose dolphin (Tursiops truncatus) improves genome completeness and provides haplotype phasing |
| *Uria lomvia* | 1179.35 | 24.89 | 15847.59 | 93.0 | 2018/01 | Assembly and RNA-free annotation of highly heterozygous genomes: The case of the thick-billed murre (Uria lomvia) |
| *Ursus maritimus* | 2301.38 | 46.51 | 15940.66 | 82.9 | 2012/06 | Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change |

**Note:** 1, the list of microorganisms genome and microorganisms meta SRA can be download from the following

 doi:10.20944/preprints201812.0156.v1



size: https://pan.genomics.cn/ucdisk/s/i2Y7Zv, if you are interested in this list, please send an email to ICG-Ocean Organizing Committee: ICG-Ocean@genomics.cn to get the code. 2, Some of the species may not live in marine, but they are important aquatic species. 3, If a species has been published by different teams, we cited the better genome assembly in this table.