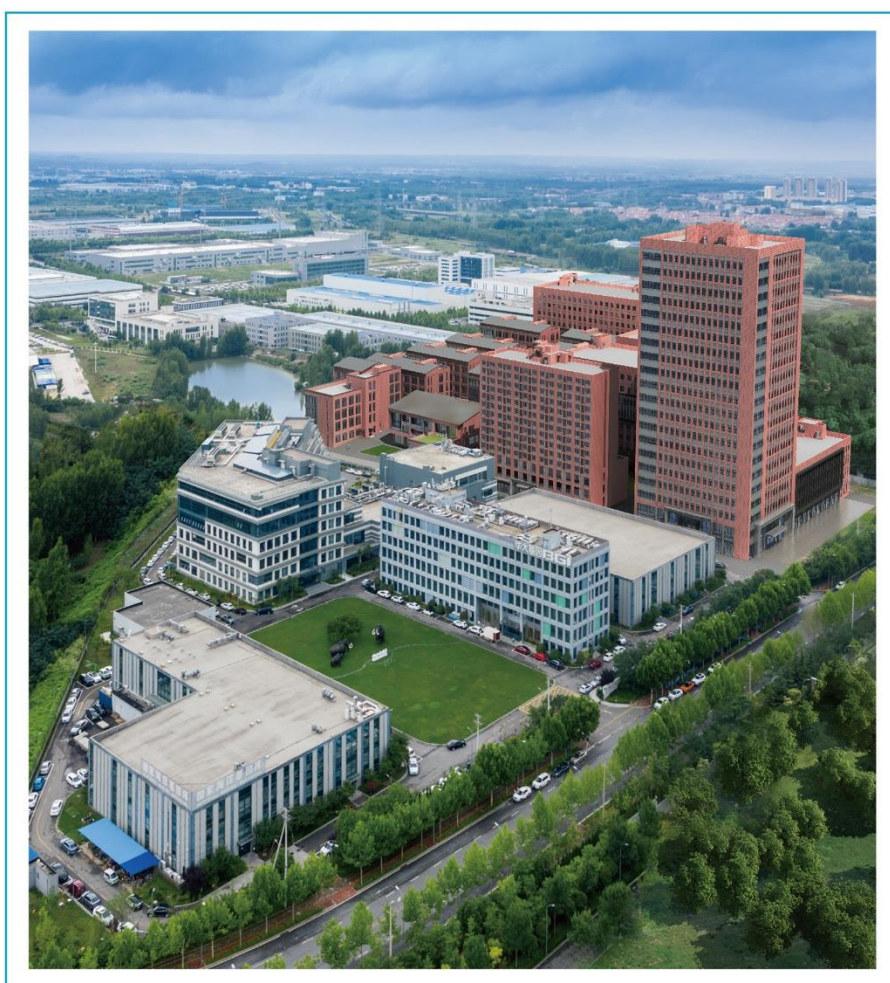


海洋生物基因组学 白皮书

REPORT ON MARINE GENOMICS



华大生命科学研究院
BGI·Research

华大智造
MGI

中德生态园
SINO-GERMAN ECOPARK

China
National
CNCB
Genobank
国家基因库

致谢

本书自《海洋生物基因组学白皮书（2018版）》发布后持续搜集资料，并于2023年6月开始撰写。直至定稿前一刻，仍在更新全球最新研究进展和数据。本书的发布集中了很多人的辛劳和心血，在此特别向他（她）们致以最衷心的感谢！

首先，感谢杨焕明老师、汪建老师、徐讯老师等对海洋生物研究的关注和大力支持，让青岛华大基因研究院团队能够聚焦于海洋生物基因组学的研究。

感谢所有参与撰写本书的作者们（以姓氏汉语拼音为序）：陈建威、杜晓、范洁、高昊阳、葛奇金、郭立东、郭扬、韩凯、贾利敏、贾洋洋、李斌、李登辉、李良伟、厉磊、刘桂林、刘美如、刘群、刘小川、刘雅琳、刘宇婧、卢怡方、吕美琪、孟亮、彭玲、秦雅婷、施成成、宋跃、孙佳惠、孙帅、孙颖、王佳昊、王睿、徐大有、于丽丽、袁增宝、张瑞华、张苏予、张要磊、赵芳、周昌浩。

感谢参与本书最终校稿的范广益、孟亮、徐梦阳、韩凯、张楠楠。

感谢为本书插图构思、设计和手工描绘的王军、齐凯如。

感谢给予本书的指导、协调并提供资源的陈力群、章文蔚、刘心、倪鸣、顾颖、刘俊年、刘姗姗、罗永伦、黎瀚博、杨献伟、郭浩冰、魏玮、战丽萍等。

还要特别感谢为本书提供部分素材与相关信息的来自全球各高校和科研院所的研究员（以姓氏汉语拼音为序）：包振民、陈松林、丁少雄、董波、郭希明、何舜平、江凌、Karsten Kristiansen、李盛英、李松海、林强、刘琨、Thomas Mock、邱强、Inge Seim、邵长伟、王大志、王堃、王禄山、王师、王文、Gregory Wray、杨连东、张琳琳、赵龙、周茜、周旭明、庄昀筠。

衷心感谢青岛市政府、青岛自贸区和中德生态园各级领导布局了基因科技未来产业，为青岛华大基因研究院的发展和海洋生物基因组学研究提供了沃土。

最后，向为全球海洋生物基因学研究作出贡献的科学家们致以最崇高的敬意，你们的学术思想、科学贡献和不懈努力，是成就本书的基石。



2023年10月25日

前言

“建设海洋强国是实现中华民族伟大复兴的重大战略任务”。习近平总书记 2018 年在山东考察时强调，“建设海洋强国，必须进一步关心海洋、认识海洋、经略海洋，加快海洋科技创新步伐。”在全球海洋开发愈演愈烈之际，我国作为海洋大国，维护好、发展好、利用好海洋资源是加快建设海洋强国的必然选择。面对数量庞大的海洋生命类群，具有大规模、高通量优势的基因测序技术成为助力海洋生命科学发展的的重要支点和强大保障。

依海而生、向海而兴。为更好的贯彻和践行新时代海洋发展观、打造全球一流的海洋发展引领和创新高地，青岛自贸片区、中德生态园加速布局海洋生物、基因技术等关键产业，取得了一系列突破性成果和进展，成为全国 67 个自贸区中唯一组织并开展大规模海洋生物基因检测任务的片区。作为基因技术发展的主要承担机构，青岛华大基因研究院在青岛市政府、青岛自贸区、中德生态园管委等政府部门的大力支持下，于 2016 年正式扎根青岛市。短短几年间，借助青岛的地理优势和资源，依靠自主研发的大规模测序技术平台和深厚的科研积累，青岛华大在向建设全球引领性的海洋生物基因组学研究中心的目标持续迈进。2017 年 9 月，青岛华大牵头成立了国际海洋基因组学联盟，旨在汇聚全球资源和技术优势、为保障海洋的可持续健康发展做出贡献。2018 年 10 月，青岛华大发布了全球首部《海洋生物基因组学白皮书》，填补了我国在海洋生物基因组学领域长期缺少综合性研究名录的空白。

六年弹指一挥间。如今，国际海洋基因组学会议（ICG-Ocean）已连续召开 6 届，《海洋生物基因组学白皮书》每年也在持续更新，青岛华大参与的多项海洋生物基因组学研究成果登上了《自然》、《科学》、《细胞》等国际顶级期刊，开展的国际间合作更是遍布全球。我们坚信，在各界领导的关怀下、在领域同行的共同努力下，蕴藏着丰富生命奥秘的广袤蓝色世界将逐渐揭开她神秘的面纱。向海图强，经略海洋，未来青岛华大将在“问海洋要答案，向海洋要资源”的路上坚定不移的走下去。

道阻且长，行则将至；行而不辍，未来可期。



2023 年 10 月 25 日

目 录

第一章 引 言	1
第二章 海洋微生物	5
2.1 全球海洋微生物组数据库摘要	6
2.2 GOMC 的构建	7
2.3 全球海洋微生物群落结构及其生物地理学分布	11
2.4 原核微生物超大基因组演化	14
2.5 海洋原核微生物防御系统	17
2.6 海洋微生物生物合成基因簇	21
2.7 体外实验验证 PET 高活性水解酶	24
2.8 讨论与展望	27
2.9 补充图表	28
第三章 藻类	51
3.1 已公布的藻类基因组特征	52
3.2 藻类基因组水平基因转移	56
3.3 藻类基因组研究热点	58
3.4 讨论与展望	62
第四章 海洋无脊椎动物	67
4.1 海洋无脊椎动物测序进展	69
4.2 海洋无脊椎动物基因组特征	71
4.3 系统发生关系研究	73
4.4 研究热点	75
4.4.1 动物发育研究	75
4.4.2 免疫功能研究	77
4.4.3 环境适应性分子机制研究	77
4.4.4 生态学研究	79
4.4.5 遗传改良研究	80
4.4.6 系统发生研究	80
4.5 讨论与展望	80
第五章 鱼类	89
5.1 全面的鱼类基因组数据集	91
5.2 基于全基因组数据的鱼类系统发生树	94
5.2.1 早期真骨鱼的进化关系	94
5.2.2 骨鳔鱼类的进化关系	96
5.3 鱼类中独特的转座元件分布	98
5.4 真骨鱼特异保守调控元件	100
5.5 鱼类进化过程中新基因的出现	103
5.6 讨论与展望	105
第六章 海洋四足动物	113

6.1 海洋四足动物简介	114
6.2 海洋四足动物基因组研究基本情况	115
6.3 海洋四足类基因组研究热点	117
6.3.1 种系发育研究	118
6.3.2 环境适应性研究.....	118
6.3.3 基于全基因组数据的海洋四足动物保护.....	120
6.4 讨论与展望	121
第七章 水生生物细胞组学.....	125
7.1 水生生物细胞组学研究概述	126
7.2 水生生物细胞图谱数据库构建	126
7.2.1 数据库数据量统计.....	127
7.2.2 细胞图谱重建	128
7.3 水生生物细胞图谱数据库应用	131
7.3.1 在胚胎发育研究中的应用.....	131
7.3.2 在神经系统演化研究中的应用.....	134
7.4 讨论与展望	137
附 录.....	142

第一章 引言

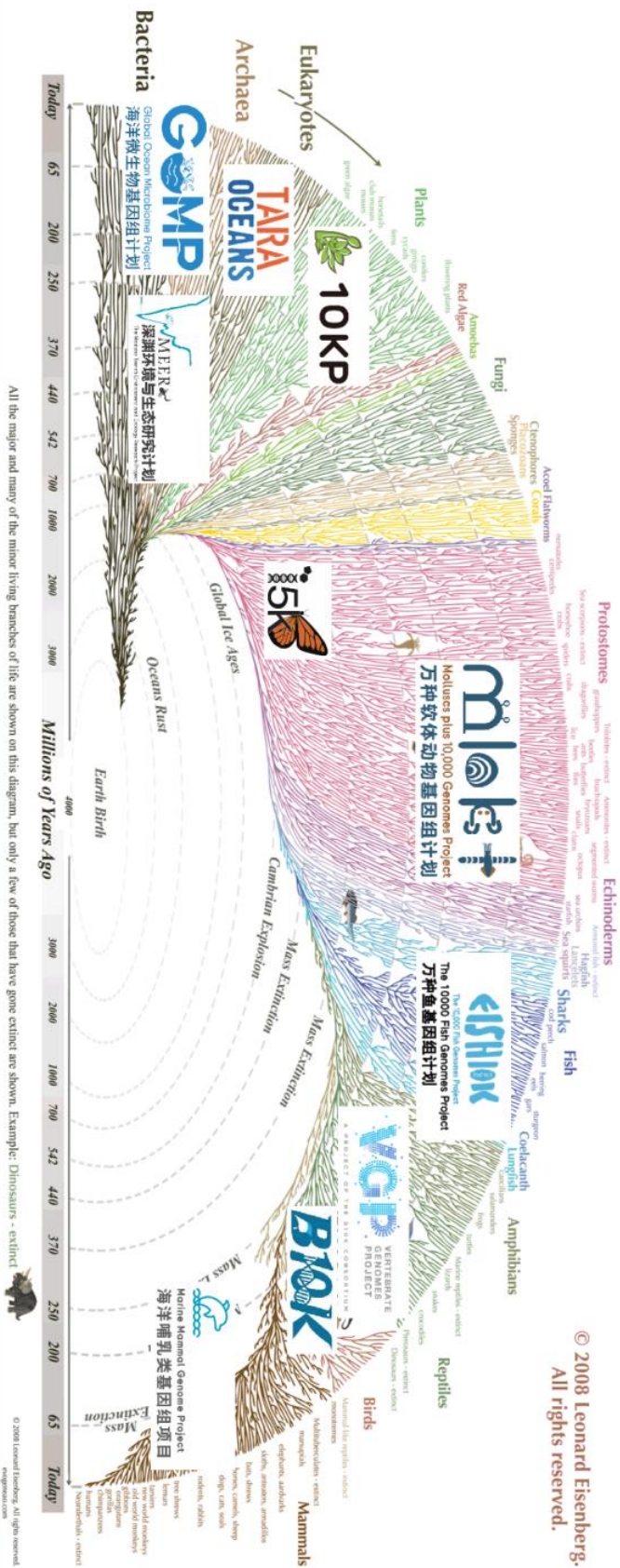
海洋占据地球表面积的 71%，是生命的起源地。经过数十亿年的演化，海洋生态环境栖息着超过 200 万个物种，其中被报道的仅有约 23 万种。海洋中极高的生物多样性为探索生命的起源和演变、生物对环境的适应性、生态学等一系列科学问题提供了空前的机遇。同时，海洋生物不仅在科学研究上具有重要意义，还是人类社会发展的食物资源之一，为人类提供充足和持续的蛋白质等营养物质。然而，与陆地动植物（如人类）的研究相比，海洋资源的勘探、海洋研究设备的开发以及海洋生物基因组学研究严重滞后。近十年来，随着测序技术和基因组学研究的日新月异，海洋生物学研究取得了较大进展，极大地促进了海洋生物探索的进程以及相关产业应用的发展。

近年来，测序技术不仅大幅提高了数据产出的通量，而且测序读长也得到显著改善。与此同时，测序成本的逐步降低也为海洋生物基因组学研究奠定了良好的基础。自 2002 年第一个海洋鱼类——红鳍东方鲀基因组发表以来，截止到目前，共发布了约 1770 种海洋动植物的全基因组序列（包含藻类、无脊椎动物、鱼类和四足动物等）和约 240 Tb 的微生物宏基因组测序数据。虽然取得了一系列重要的成果，但海洋生物基因组学未来仍然存在诸多挑战。未测序物种基因组、群体基因组、其他多组学技术联合等研究将成为未来海洋生物基因组学研究的重点，使海洋生物基因组学研究更加系统和深入。

在过去的十年中，科研人员已经启动了多个重要的大规模海洋生物基因组测序计划。由 EMBL（欧洲分子生物学实验室）领导的 Tara Oceans 计划于 2009 年 9 月启动，该计划在全球范围内的 200 多个海洋站点采集到 30,000 多个海洋环境样本，构建了目前为止最大的海洋环境微生物测序数据库[1]；2014 年启动的千种鱼类转录组测序计划（Fish-T1K）旨在对 1,000 种鱼类的转录组进行测序[2]；Genome 10K 计划（G10K）旨在对 10,000 种脊椎动物的基因组进行测序，其中包括 4,000 多种鱼类[3]。万种植物基因组计划（10KP）[4]——包含 4,000 种藻类，和地球生物基因组计划（Earth BioGenome Project, EBP）[5]——旨在完成包含大量海洋生物的所有已知真核生物基因组测序。特别地，自 2018 年以来，

针对特有的海洋生物物种类群，启动了一系列基因组研究计划，如万种鱼基因组计划（The 10,000 Fish Genomes Project, Fish10K）、万种软体动物基因组计划（Molluscs plus 10,000 Genomes Project, M10K+）、全球海洋微生物基因组计划（Global Ocean Microbiome Project, GOMP）、深渊环境与生态研究计划（The Marina Trench Environment and Ecology Research Project, MEER）等，旨在重构海洋生命之树，解析其演化历史（图 1.1）。

自 2017 年首届国际海洋基因组学会议（International Conference on Genomics of Ocean, ICG-Ocean）在中国青岛成功举办，以及 2018 年发布首部《海洋生物基因组学白皮书》以来，海洋生物基因组学研究已越来越多的引起社会各界的关注。该《海洋生物基因组学白皮书》总结了海洋生物基因组学的最新进展，包括涵盖以下六个方面：海洋微生物、藻类、无脊椎动物、鱼类、海洋四足哺乳动物和细胞组学的相关研究。这些内容包括了对过去研究成果的总结和对未来科研发展的展望，我们希望这部白皮书能够吸引来自各个领域的专家对其中内容进行深入的讨论。虽然我们编写能力有限，在内容上可能存在一些不足和错误，但我们真诚地希望本书能够帮助研究者、资助机构以及产业从业者扩展视野，为海洋生物基因组学领域的研究和应用提供一个重要参考，并促进更多的合作和创新。



微生物

细胞总数: 约 3.6×10^{29}

基因组数: 24,195

热点: 分类、功能、应用

藻类

物种: 超过3万

基因组数: 400+

热点: 演化、生态

无脊椎动物

物种: 超过43万

基因组数: 640

热点: 演化、发育、应用

鱼类

物种: 超过3万

基因组数: 533

热点: 演化、应用

四足动物

物种: 约1,128

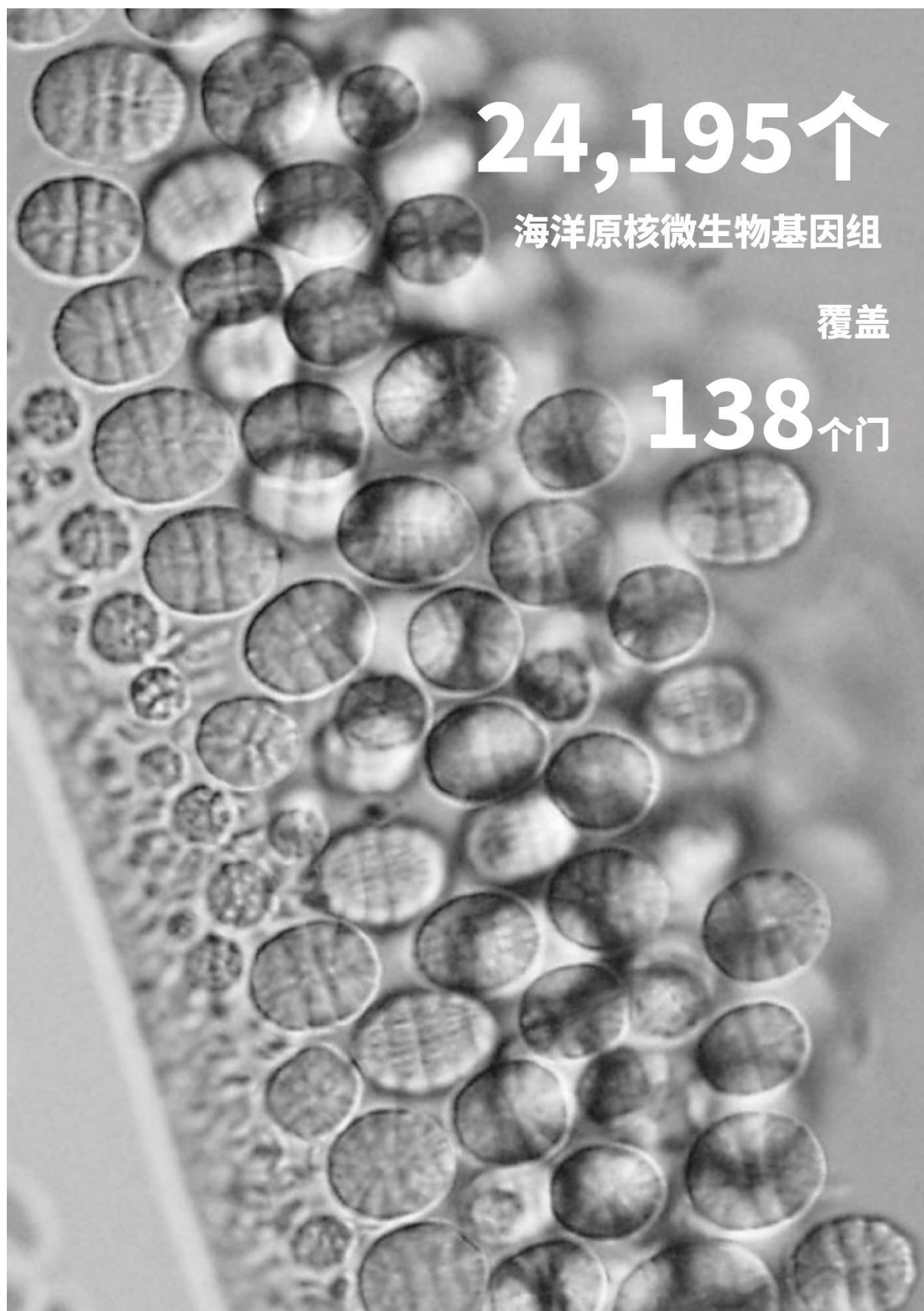
基因组数: 198

热点: 演化、遗传保护

图 1.1 生命之树及基因组大科学计划

本章参考文献

1. Sunagawa, S., et al., Ocean plankton. Structure and function of the global ocean microbiome. 2015. 348(6237): p. 1261359.
2. Sun, Y., et al., Fish-T1K (Transcriptomes of 1,000 Fishes) Project: large-scale transcriptome data for fish evolution studies. 2016. 5(1): p. 18.
3. Bernardi, G., et al., The fishes of Genome 10K. 2012. 7(3): p. 3-6.
4. Cheng, S., et al., 10KP: A phylodiverse genome sequencing plan. 2018. 7(3): p. giy013.
5. Lewin, H.A., et al., Earth BioGenome Project: Sequencing life for the future of life. 2018. 115(17): p. 4325-4333.



24,195个

海洋原核微生物基因组

覆盖

138个门

第二章 海洋微生物

海洋原核生物细胞总数约为 10^{29} 个，是海洋的重要组成部分，也是全球生物地球化学循环和生态过程的重要基石[1, 2]。海洋微生物在分类和代谢方面具有显著的多样性，并可根据环境变化进行快速的进化适应[3, 4]。最近，测序技术的进步消除了微生物低可培养性造成的障碍，使采用宏基因组学分析揭示海洋微生物的“暗物质”成为可能。特别是环球大洋采样项目（Global Ocean Sampling, GOS）[5]和塔拉海洋科考计划（Tara Oceans Expedition）[6]使我们对全球海洋微生物的认识发生了质的变化。然而，海洋、淡水和陆地等各类环境中新产生的基因组和遗传数据表明仍有许多原核生物未被发现，尤其是在深海、极地、热液喷口和冷泉等极端生态系统中。

尽管在全球范围内开展了大量相关的测序工作，但采用综合系统性的方法来评估全球海洋微生物组多样性的研究却很少[7, 8]。因此，目前仍未建立能够涵盖从南极到北极、从表层海洋到深层海洋等各种海洋生态系统类型的基于宏基因组学的统一海洋微生物基因组目录。然而，该目录对以下几个海洋微生物研究领域至关重要：A）在物种水平上鉴定海洋微生物基因组的分类和功能；B）揭示海洋微生物的进化轨迹；C）基于比较基因组学解析海洋微生物的生态适应性功能；D）识别可用于生物勘探的基因簇。

2.1 全球海洋微生物组数据库摘要

在过去的二十年中，已发起的一些大规模海洋微生物组测序计划对海洋微生物功能多样性的认识产生了许多重要的见解。然而，我们仍然缺乏一个维度上从南极到北极、垂直层带上从表层到深层的、全生态的海洋微生物组数据集。为建立统一的数据参考集，我们收集了 NCBI、EBI 和 JGI 公开数据库中 2009 年 8 月至 2020 年 7 月期间从 NCBI、EBI 和 JGI 获得的公开海洋微生物宏基因组，通过组装和分箱分析构建了 43,191 个宏基因组组装微生物基因组（metagenome-assembled genomes, MAGs）。然后，我们对这些基因组进行去冗余，最终获得了广泛分布在 138 个门类中的 16,240 个非冗余的 MAGs。将这些非冗余 MAGs 与公开的海洋原核微生物基因组（NCBI、OMD 和 OceanDNA）相结合[7, 8]，我们构建了迄今为止最完整的全球海洋微生物组数据库（The Global Ocean Microbiome

Catalogue, GOMC)。该数据库将已知海洋微生物的多样性丰富程度提升超两倍，在不同的分类等级上鉴定到许多新的 MAGs，其中包括从深海热液喷口中发现的一个潜在的新门。我们对原核生物 MAGs 的丰度进行解析并绘制了全球范围内微生物群落的生物地理分布模式。此外，基于 GOMC 中的高质量 MAGs，我们进行了基因组扩张分析、识别了微生物 CRISPR-Cas 和抗生素基因防御系统的差异性分布及其相互间的复杂作用、鉴定了各种 BGC 的类型和分布。利用该数据集，我们从深海和热液喷口生态下的微生物类群中挖掘到具有高活性嗜盐能力的 PET 酶并从功能上实验验证了其潜在的应用前景。由此，GOMC 可为未来研究海洋微生物功能、进化、生物地理学和生态相互作用提供宝贵的数据资源。

2.2 GOMC 的构建

我们从 24,395 个公开的海洋宏基因组中收集了共计 237.02 Tb 的测序数据，样本涵盖了从南极到北极（纬度范围从 77.90°S 到 89.99°N）、从海洋表层到深达 10,000 米超深渊海沟的广泛生境（图 S2.1 和表 S2.1）。我们使用单个样本分别组装分箱的策略（图 S2.1），获得了 43,191 个中高质量的 MAGs（图 2.1A、B）。将新组装的 MAGs 按 99% 的全基因组平均核苷酸一致性（average nucleotide identity, ANI）进一步聚类，得到 16,240 个具有亚种分辨率的非冗余代表性 MAGs。在这些非冗余 MAGs 中，基因组总体平均完整度高达 82.33%，潜在污染率低至 1.79%，包括了 5,673 个（34.93%）接近完整的基因组和 7,147 个（44.01%）高质量的基因组（图 2.1B）。这些非冗余 MAGs 代表了海洋微生物的广泛多样性，涵盖了基因组分类数据库（Genome Taxonomy Database, GTDB, v207_2）中大部分已知的原核生物门类（~83.13%），包括 148 个细菌门类中的 120 个和所有 18 个古菌门类。系统发育分析显示，最常见的细菌门是变形菌门（Proteobacteria）、拟杆菌门（Bacteroidota）、放线菌门（Actinobacteriota）、浮霉菌门（Planctomycetota）、绿弯菌门（Chloroflexota）和蓝细菌门（Cyanobacteria），而最多的古菌门是 Thermoplasmata 和 Thermoproteota（图 2.1B）。在 16,240 个非冗余 MAGs 中，分别有 11、44、194、719 和 3,473 个基因组无法归入 GTDB 中已知的门、纲、目、科和属一级的类群。在物种水平上，有很大一部分细菌和古菌（分别占 64.07% 和 58.98%）无法鉴定为已知类群（共计

10,308 个 MAGs)。因此, 与 GTDB 相比, 它们可能代表候选新物种 (图 2.1B)。这一大部分未经分类的基因组表明, 海洋中还存在大量未被探索的原核生物多样性。

为了构建一个全面的海洋微生物基因组集, 我们整合了本研究中新发现的 16,240 个非冗余 MAGs、两个已公开的海洋微生物基因组数据集 OMD [7]、OceanDNA [8]以及来自 NCBI 的 8,050 个海洋原核微生物基因组, 并进一步使用代表物种水平的 95% ANI 一致性进行聚类, 最终得到了一个由 24,195 个非冗余基因组组成的数据集, 即 GOMC (图 S2.1)。据我们所知, GOMC 是迄今为止最全面、最完整的海洋微生物组数据库。其中, 9,939 个新发现的 MAGs (占总数的 41.04%) 是本研究所特有的, 在之前的任何数据库 (包括 NCBI、OMD[7]和 OceanDNA [8]) 中都没有记录 (图 2.1C、D)。这 9,939 个新的 MAGs 主要来自以前未在全球范围内进行过系统性研究的深海 (3,713 个 MAGs)、沉积物 (1,371 个 MAGs) 和共附生 (1,250 个 MAGs) 等生态系统 (图 2.1D)。值得注意的是, 大多数这些新发现的特有 MAGs (占 82.06%) 都可能代表潜在的新物种。因此, GOMC 使已知海洋古菌和细菌的系统发育多样性分别增加了 255.98%和 133.73% (图 2.1E、F)。新发现的 Thermoproteota 和 Halobacteriota 的基因组是古菌系统发育多样性扩大的主要贡献者 (图 2.1D、E)。此外, 我们还首次从海洋生态系统中发现了古菌门 Hadarchaeota 和 Micrarchaeota。相较于古菌, 脱硫杆菌 (Desulfobacterota)、弯曲杆菌 (Campylobacterota) 和 Patescibacteria 对细菌系统发育多样性的扩展贡献最大 (图 2.1D、F)。具体来说, 本研究中弯曲杆菌和脱硫杆菌中新发现的 MAGs 占该类群基因组的 80% 以上 (图 2.1D)。此外, 我们的数据库涵盖了蓝细菌系统发育的大部分范围, 包括分类等级相对较低的物种 (图 S2.2A、B), 从而扩大了全球蓝藻泛基因组的功能多样性 (图 S2.2C)。

同时, GOMC 为我们提供了一种新的视角, 以便更好地理解整个海洋生态系统, 尤其是极地海洋和深海热液喷口等极端环境中的微生物多样性。我们重新分析了来自全球热液喷口的共 777 个样本, 并成功重构了 3,064 个非冗余 MAGs。其中, 三个基因组 (GOMC.bin.39、GOMC.bin.1651 和 GOMC.bin.3975, 平均完整度为 80.77%, 污染度为 3.09%) 形成了一个独特的单系群, 迄今为止在现有数据库中没有发现与之相似的基因组 (图 S2.3A)。它们与最接近的相邻门类 Aquificota 和 Campylobacterota 的相对进化差异

（relative evolutionary divergence, RED）低于 0.25[9]、ANI 小于 60%（图 S2.3A）。因此，这些数据表明它们有可能代表一个新的候选细菌门。

这三个 MAGs 是从三个不同的热液喷口获得的：大西洋中脊 Lucky Strike（大西洋）、Brothers Volcano（太平洋）和 Axial Volcano（北太平洋）（图 S2.3B），可能代表了同一个目下的不同科（RED 值介于 0.61 和 0.79 之间）。它们的基因组相对较小（2.0~2.5 Mb），GC 含量较低（31%~41%）。根据其基因组特征预测的最适生长温度（optimal growth temperature, OGT）[10]表明它们属于嗜热微生物（图 S2.3A）。根据丰度分析，这些微生物在广泛分布于全球海洋中的 48 个样本中被检测到（图 S2.3B）。尽管它们的总体丰度较低，但与开放海洋（约 0.0171%）相比，它们在热液喷口生态系统中的丰度较高（约 0.0456%）。在每个 MAGs 中预测到的约 2,000 个基因中，约有 70% 可以被 NCBI NR 和 KEGG 等数据库进行注释。这些 MAGs 所代表的功能类型中包含了主要的碳水化合物代谢途径，包括糖酵解、葡萄糖生成、卡尔文-本森-巴沙姆途径（Calvin-Benson-Bassham, CBB）、TCA 循环和脂肪酸代谢等。然而，需要注意的是，这些经典途径中的一些基因在这些 MAGs 中并不存在（图 S2.3C）。此外，我们还发现了一种非典型的伍德-隆达尔途径（Wood-Ljungdahl pathway, WLP），这可能赋予它们使用一碳底物的能力。我们还发现了一条还原甘氨酸途径（reductive glycine pathway, RGP），它与 WLP 相结合，支持甘氨酸、丝氨酸、半胱氨酸和甲硫氨酸的生物合成。这种非典型的 WLP 途径被认为是最古老的厌氧固碳途径之一，很可能反映了对深海热液喷口等地质古老生态位的适应。同时，也鉴定出了参与支链氨基酸合成的基因和跨膜氨转运体，这表明它们获取必需氨基酸的策略具有多样性。膜结合的 Nuo 复合物和 F-type H⁺-transporting ATPase 可能负责 NADH 还原和能量转换。此外，在所有三个 MAGs 中都发现了多组分 Na⁺:H⁺反转运体可调节细胞质内钠离子水平并产生质子动力[11]。

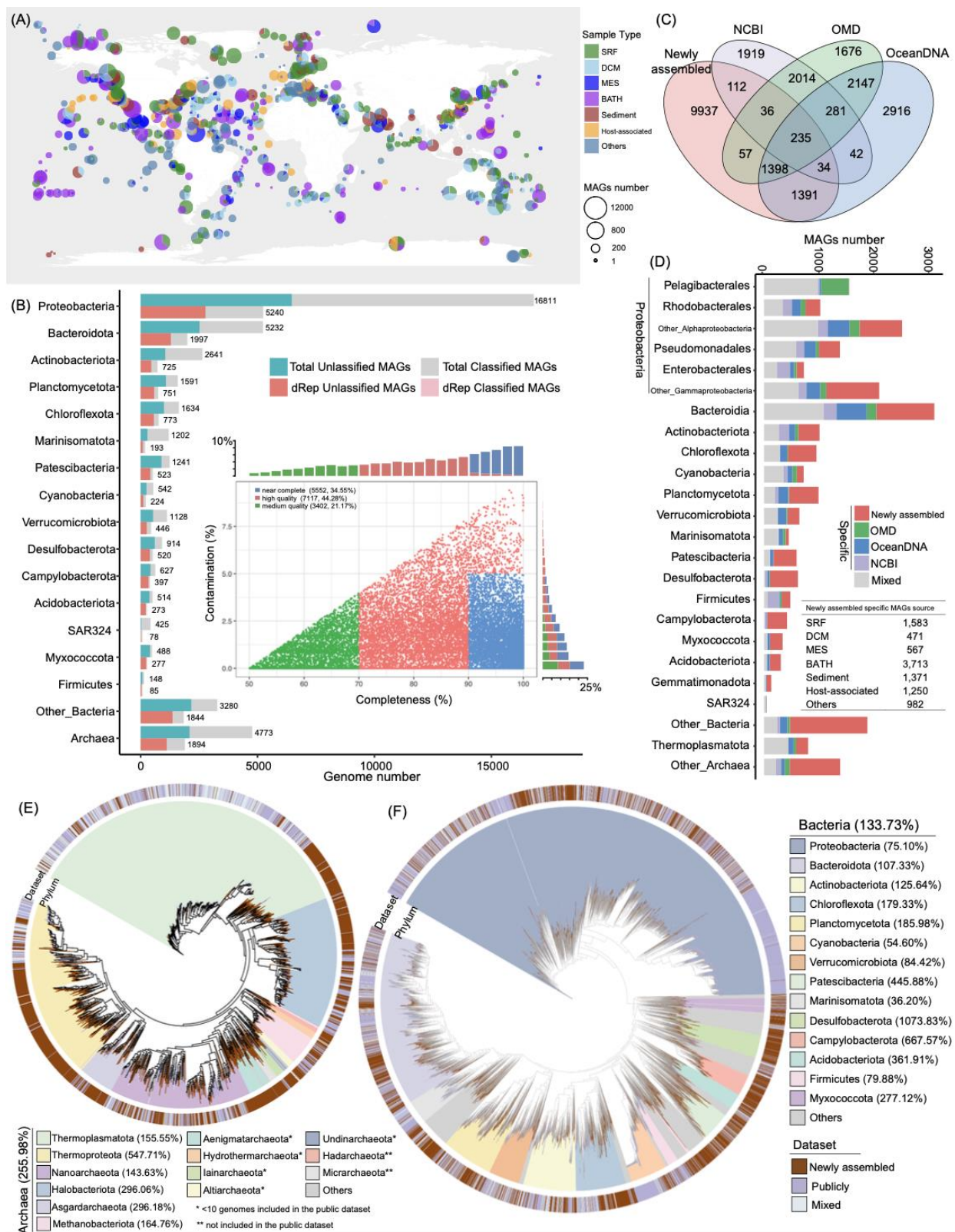


图 2.1. 宏基因组组装基因组 (MAGs) 的地理分布和生态系统分布特征及其对微生物生命树的贡献。

(A) 43,191 个 MAGs 的地理分布。“SRF”代表“表层海水”，“DCM”代表“叶绿素浓度最大层”，“MES”代表“中层带”，“BATH”代表“深层带”。(B) 重构的 43,191 个 MAGs 和 16,240 个非冗余 MAGs，很大一部分为未分类基因组。所有 MAGs 的质量在中等至较高水平，其中约 35% 接近完整。(C) 维恩图显示了新组装 MAGs 与 NCBI、OMD 和 OceanDNA 中特有或共有的基因组。(D) 本研究 and 现有已发表数据库对每个原核生物门的贡献。右侧表格显示了本研究中 9,937 个特有 MAGs 的生态系统来源。(E 和 F) 基于 GOMC 中 122 或 120 个单拷贝基因的古菌 (E) 和细菌 (F) 基因组的系统发生树。图 1 中的 Myxococcota 门代表 GTDB 分类的 Myxococcota 和 Myxococcota_A，Methanobacteriota 门代表 Methanobacteriota、Methanobacteriota_A 和 Methanobacteriota_B。

2.3 全球海洋微生物群落结构及其生物地理学分布

不同的海洋生态系统可能拥有不同的原核生物群落，这可能与环境过滤有关 (图 2.2A)。例如，变形菌门遍布表层至深层水体，而蓝细菌则富集在叶绿素浓度最大层 (deep chlorophyll maximum, DCM)。相对而言，中层带 (mesopelagic, MES) 中浮霉菌门的含量比浅海中多。在中层和深层 (bathypelagic, BATH) 海水中，绿弯菌门和 Marinisomatota 最为丰富，这表明它们更适应深海环境。在整个水柱和所有海域中分布最广的是 SAR324，在深渊也有发现，这与其先前确定的生态型深度分区模式一致[12]。脱硫杆菌是底栖环境中最普遍的硫酸盐还原菌之一，主要富集在沉积物样本中。本研究发现，不同深度的原核生物群落之间的地理距离与 Jensen-Shannon divergence (JSD) 之间总体呈正相关，但相关性较弱 (Mantel test: $0.09 \sim 0.39$, $p < 0.01$)，其中表层海水 (SRF) 群落的相关性最弱。这一结果可能与温盐环流和表层洋流有关，因为这两个过程通常在深层水域不太明显。

尽管我们观察到了在整个水柱不同深度层中的微生物分布总体模式，但在每个特定深度层中，原核生物群落仍然表现出显著的多样性 (图 2.2A)。我们根据每个样本中微生物的相对丰度，基于 UMAP (uniform manifold approximation and projection) 降维方法，成功划分出 56 个宏基因组分区 (metagenomic provinces, MPs) [13] (图 2.2B 和图 S2.4)，并基于 JSD 矩阵进行 ANOSIM 检验 ($R = 0.61$, $p < 0.01$) 来评估不同 MPs 之间的分离程度。结果显示，不同气候带构成了不同的 MPs 组合，但大多数 MPs 都局限于特定的海洋

深度，只有少数例外跨越了相邻的深度边界（图 2.2C），这与图 2A 的观察一致。一般而言，我们发现拟杆菌、 α -变形菌和 γ -变形菌在许多 MPs 中都很普遍[6, 14]，而其他类群只在特定生态系统中大量存在，如弯曲菌门（Campylobacterota）主要出现在深海热液喷口中[15]（图 S2.4）。

在全球范围内，同一 MP 并不限制于地理位置上相邻的采样点，相反，MP 中的采样点可以呈现广泛的地理分布（图 S2.4）。因此，同属一个气候带和深度水层的 MPs 可以占据不同的生态位，其中一些跨越多个海区，而另一些则只局限于较小的区域。例如，MP U0 的样本数量最多，在 SRF 和 DCM 层中覆盖了广阔的地域，从北温带一直延伸到南温带。高丰度的蓝细菌、拟杆菌和变形菌以及相当一部分放线菌、疣微菌和 SAR324 使得 U0 与其他 MPs 明显区分开来（图 S2.4）。在所有沉积物样本中，MP U18 表现出最高的香农（Shannon index）多样性（图 2.2D），并具有独特的分类组成，包括脱硫杆菌、浮霉菌、绿弯菌和泉古菌（图 S2.4）。环境变量（如温度、盐度等）并不能很好的解释不同微生物群落香农多样性上的差异（ $p < 0.01$, adjusted $R^2 < 0.5$ ）（图 S2.5A、B）。而营养元素对主要由浮游植物占据的 DCM 水层（图 S2.5C）的微生物群落多样性影响更大[16]。要深入探究全球海洋微生物群落组成的驱动因素，需要更准确、更全面的环境条件数据集，以确定潜在的的决定性生物因素（如捕食者与被捕食者之间的相互作用），同时也需要考虑尚未被考虑的非生物因素（如原核生物使用的新资源）等[17]。

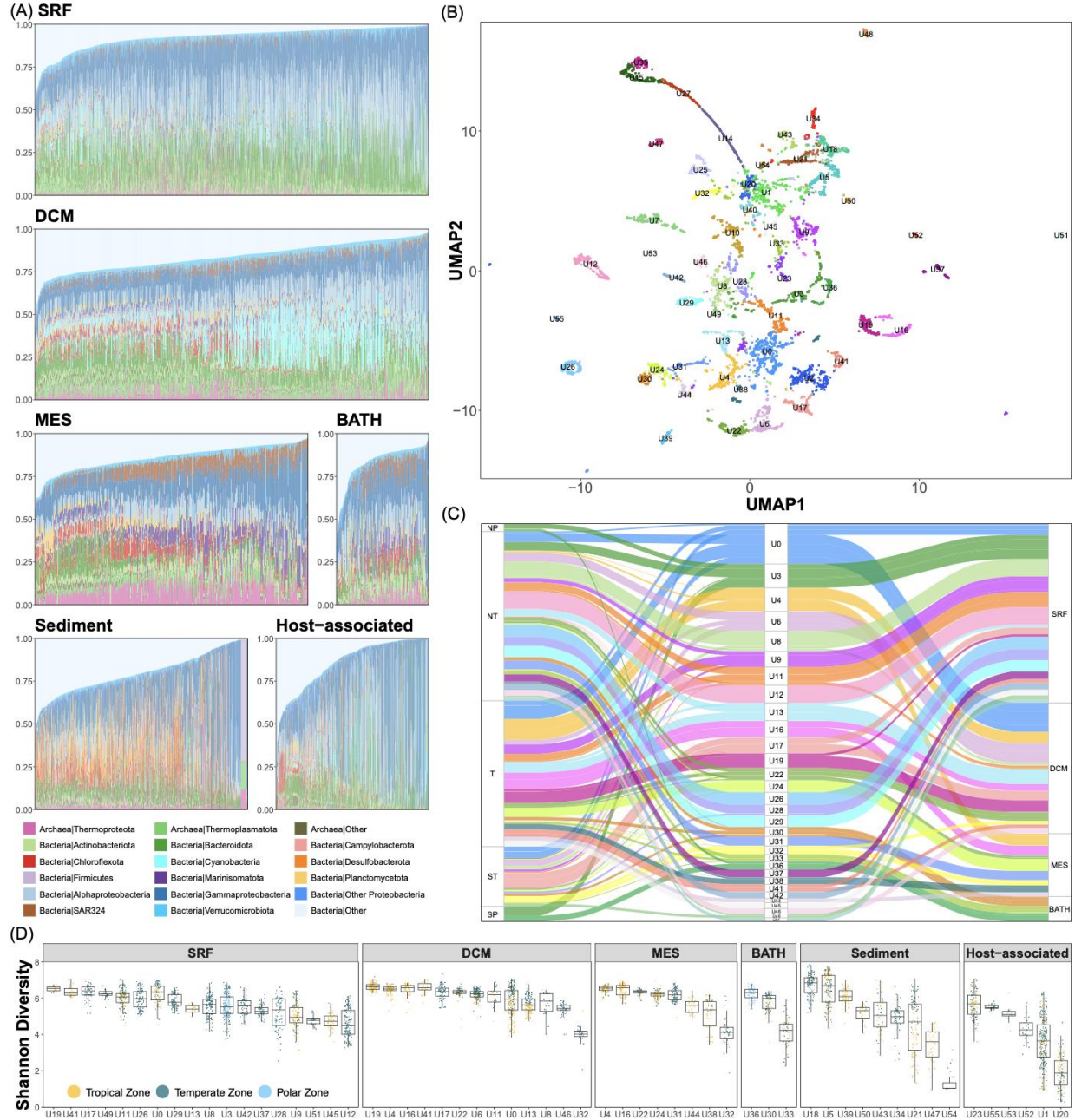


图 2.2. 原核微生物生物地理分布和宏基因组分区（metagenomic provinces, MPs）的划分。（A）不同深度样本的微生物群落组成。（B）MPs 在 UMAP 降维空间中的分布。不同颜色表示不同的 MPs。每个 MP 的标识符标注在每个簇的中间。图 S4 显示了 MPs 的详细信息。（C）冲积图显示仅由海水样本组成的 MPs 的地理分组。左侧显示了主要气候区类别，并用以下缩写表示：“NP”表示“北极”，“NT”表示“北温带”，“T”表示“热带”，“ST”表示“南温带”，“SP”表示“南极”。MP 的颜色含义与图 B 相同。（D）MPs 所包含的样本的 Alpha 多样性分布。属于同一 MP 但来自不同深

度区域的样本会被分开统计（例如，U0 中来自于 SRF 和 DCM 样本被分开显示）。来自不同气候带的样本由不同的颜色显示，其对应的色谱展示在图的底部。

2.4 原核微生物超大基因组演化

根据进化理论，环境的高变异性会选择代谢潜力更大的基因组[18, 19]。该观点在陆地和淡水生态系统中得到了很好的证实，但在海洋生境中却鲜有研究[20, 21]。GOMC 包括 303 个超大原核基因组（基因组大小 ≥ 8 Mb），其中，三个浮霉菌门新 MAGs 进一步提高了海洋原核生物基因组大小的上限。这三个高质量的 MAGs 属于 *Pirellulacea* 科，其基因组大小介于 16.7 和 18.4 Mb 之间（GOMC.bin.14112、GOMC.bin.14342 和 GOMC.bin.14350，平均完整度 = 98.76%，污染度 = 6.67%），超过了目前 NCBI GenBank 中已有的海洋基因组的大小。并且，它们均通过了基于熵方法的嵌合检验（chimerism examination），表明它们的大基因组并不是分箱错误所导致的嵌合体[22]（图 S2.6A）。这三个 MAGs 均来自加勒比海委内瑞拉北部大陆架的缺氧海洋盆地卡里亚科盆地的两个样本[23]，有趣的是，它们的近源物种——基因组大小为 11.7 Mb 的 *Pirellulaceae* bacterium 是在黑海缺氧浮游系统的上层被发现的[24]（图 S2.6B、C）。虽然这两种环境在一些理化特性上有所不同，但它们都具有营养物质供应波动和氧化还原反应显著的特点。因此，这两个生态系统中普遍存在的显著环境变化可能对那些具有大型复杂基因组的细菌产生有利的选择压力[25]。

与进化树中亲缘关系最近的支系相比，这四个 MAGs 的基因组大小显著增大（图 S2.6C）。为了研究这些庞大的基因组会带来哪些生态优势，我们基于更大的基因组会优先积累涉及基因组稳定性、细胞周期进程、信号转导和调控等相关基因的假设，研究了各类功能基因的数目与基因组大小之间的相关性趋势[19]。我们采用了系统发育回归分析、祖先基因重建、基因组大小与基因拷贝数的相关性等分析（图 S6DEF）以阐明基因组大小的扩张可能导致的遗传功能改变。最终，我们通过这三项分析筛选出了 77 个 Pfam 结构域（图 S2.6G）。这些结构域功能十分广泛，涉及各种生物过程，包括催化活性、分子识别和细胞调控等。总的来说，大多数筛选得到的 Pfam 结构域在各个门类中均与基因组大小呈现出显著正相关，与我们的预期相符（图 2.3A、B、C、D 和图 S2.7）。此外，我们还观察到不同物种在基因组扩张时对特定功能域有不同的偏好（图 2.3E、F、G、H 和图

S8)。在粘细菌 (Myxococcota) 中, 蛋白激酶结构域 (PF00069)、von Willebrand factor type A 型结构域 (PF13519、PF13768) 和 ATPase GHKL 结构域 (PF02518) 是导致基因组大小变化的重要因素 (图 2.3E)。而在浮霉菌门 (Planctomycetota) 中, 3-酮二糖水解酶 (PF06439) 对基因组大小有显著影响 (图 2.3F)。有趣的是, 众所周知的膜转运蛋白超家族 MFS (major facilitator superfamily, PF07690) 是酸性杆菌门 (Acidobacteriota) 中基因组大小扩张的一个强有力的影响因素 (图 2.3G)。在放线菌门 (Actinobacteriota) 中, 甲基转移酶结构域 (PF13649) 和 FAD binding 结构域 (PF00890) 随着基因组大小的增加而显著增加 (图 2.3H)。这些蛋白质涉及参与了多方面的功能, 它们的扩展表明了细胞过程潜在复杂性的增加。这种复杂性往往与较大的基因组相对应, 暗示较大的基因组从本质上促进了复杂细胞机制的衍生, 从而提高了原核微生物在非生物和生物条件显著波动的环境中的生存和适应能力。

除了特征明确的蛋白质结构域外, 我们还观察到基因组大小与含有 WD40-motif 结构的蛋白质 (WD40-motif proteins, PF00400) 之间存在显著的正相关性 (图 2.3A、B、C、D 和图 S8)。WD40 是一个古老的蛋白质家族, 最初在真核生物中被发现, 但随后也在表型复杂的原核生物中被发现[26, 27], 其通常是蛋白质-蛋白质相互作用的支架[28, 29]。例如, WD40 蛋白可能有助于浮霉菌门形成特有的胞质内膜, 从而促进类真核细胞的胞内分隔[30]。WD40 蛋白的相邻蛋白质在功能上多种多样, 没有富集特定功能的蛋白质家族。综上所述, 大型细菌基因组似乎是蛋白家族扩张的结果, WD40 蛋白可能在维持细胞内复杂性方面发挥了核心作用。

此外, 我们还评估了整体基因组特征的变化。虽然小基因组往往具有更高的编码密度, 但在主要原核生物门类中, 编码密度与基因组大小之间没有发现一致的趋势 (图 S2.9)。我们发现基因长度和基因间区长度随着基因组大小的增加而明显增加 (图 S2.9) [31, 32]。同样, 较大的基因组具有较高的 GC 含量, 最高可达约 75% (图 S2.9), 这可能是由内在突变偏差和环境因素共同限制所致[33-36]。

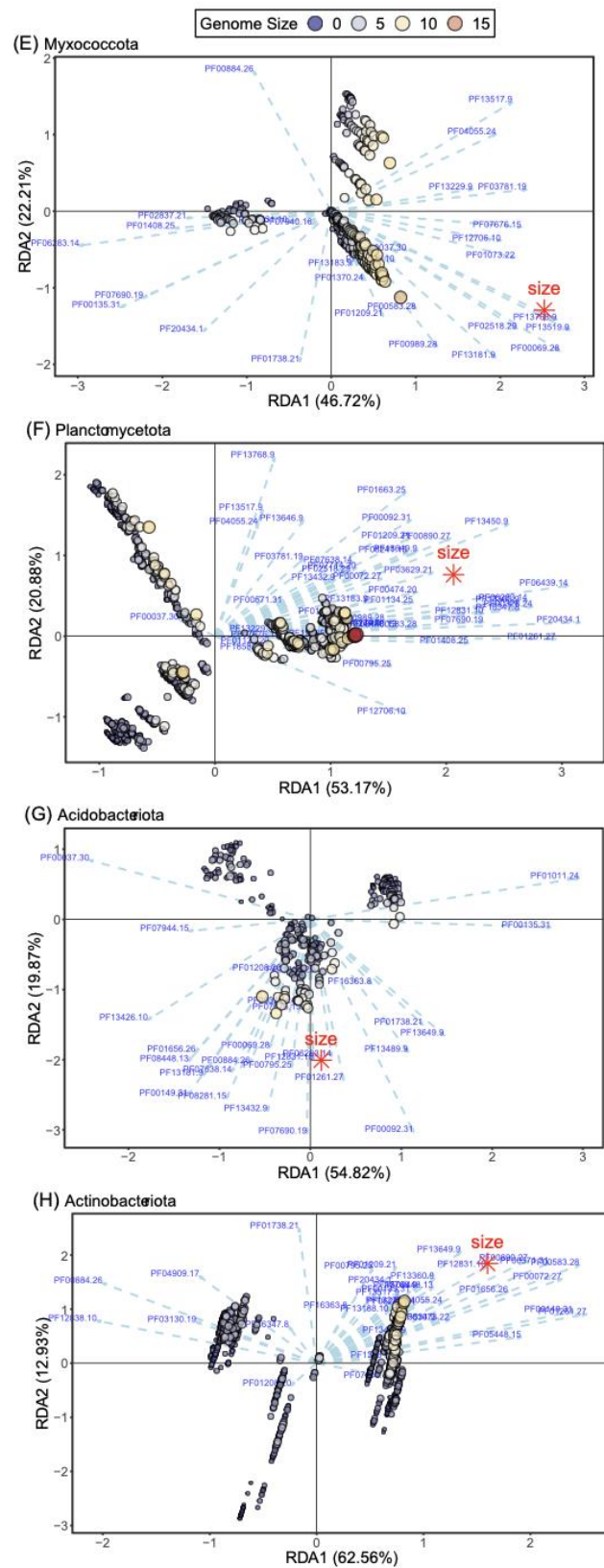
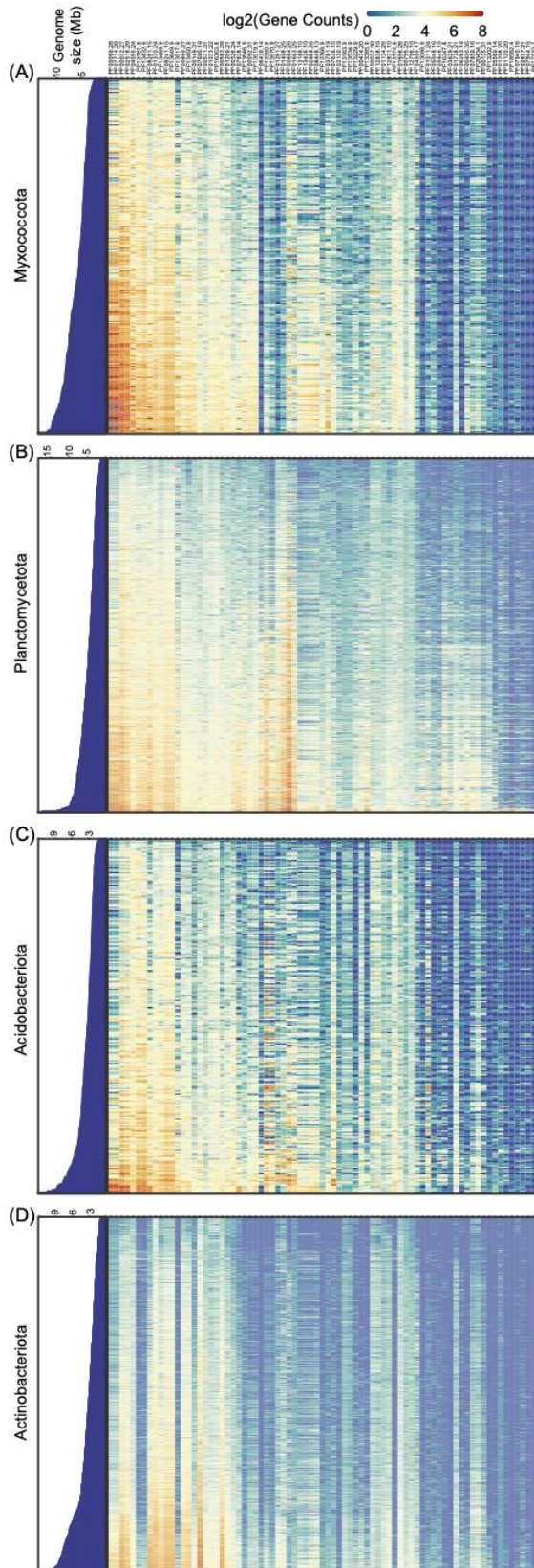


图 2.3. 原核微生物基因组扩张的总体特征和特点。我们研究了粘细菌门（A）、浮霉菌门（B）、酸杆菌门（C）和放线菌门（D）中 77 个 Pfam 结构域对应的基因数量的变化随基因组大小递增的变化趋势。此外，我们还进行了冗余分析（Redundancy Analysis, RDA），研究了粘细菌门（E）、浮霉菌门（F）、酸杆菌门（G）和放线菌门（H）中基因组大小与包含特定 Pfam 结构域的基因拷贝数之间的关系。值得注意的是，通过前向选择发现的显著相关性用蓝色箭头表示。门类的排列是根据其在高质量基因组（完整性 > 70%，污染小于 < 10%）中的平均估计基因组大小确定的。这些分析包括 9 个门类，其余 5 个门类见图 S8。图 3 中的 Myxococcota 包含 GTDB 分类的 Myxococcota 和 Myxococcota_A。

2.5 海洋原核微生物防御系统

基于 CRISPR-Cas 的微生物“免疫”系统是微生物抵御病毒感染和防止异源基因入侵的主要机制之一[37, 38]。自发现以来，CRISPR-Cas 系统在不同微生物种系和不同生态系统中的分布引起了人们的广泛关注[39]。在 GOMC 数据集中，我们一共从 3212 个 MAGs（约占总 MAGs 的~15%）中鉴定出 40 种不同 Cas 类型的 5127 个 Cas 操纵子蛋白，其中 1708 个基因组（~7.1%）携带了完整的 CRISPR-Cas 系统（图 2.4A）。Firmicutes_B 基因组中编码 Cas 操纵子的 MAGs 比例最高（约 81%），而其他类群，如 Margulisbacteria、Rhizobiales 和一些变形菌（TMED109、TMED127、UBA8366、GCA-002705445）则很少编码 Cas 蛋白（图 2.4A）。本研究中观察到的 Cas 操纵子总体检出率和不均匀分布模式与之前的研究基本一致[40]。

温度是影响原核生物中 CRISPR-Cas 系统丰度的一个已知因素。为了研究温度的潜在影响，我们首先根据基因组信息预测了 GOMC 中所有 MAGs 的最适生长温度 OGT [10]。结果显示，在大多数原核生物门类中，编码 Cas 蛋白的微生物的平均 OGT 明显高于没有编码 Cas 蛋白的微生物（图 S2.10）。在不同的门中，预测 OGT 较高的 MAGs 的 Cas 蛋白检出频率显著较高（图 2.4B）。同样，与海水样本相比，处在热液喷口高温环境的 MPs 中编码 CRISPR-Cas 系统的 MAGs 比例明显更高（图 2.4D）。

为了研究 CRISPR-Cas 系统在各种不同生境中的分布情况，在 GOMC 数据集的基础上，我们还对从冰川到热泉、从水生到陆生、从营自由生到宿主相关样本的数据进行了扩展分析[41, 42]。结果表明，热泉生境中的微生物编码 Cas 蛋白的频率最高（图 2.4C 和图 S2.11）。除温度外，还观察到与海水相比，与宿主相关的生态系统 MAGs 的 Cas 蛋白检出比例更高（图 2.4D）。此外，一些厌氧生境系统中微生物的 Cas 蛋白检出率也相对较

高，包括肠道生境、污水厌氧发酵系统以及陆地深层地下生境等（图 2.4C），这可能是由于先前报道的宿主相关条件和/或低氧浓度导致的[39]。

虽然微生物 CRISPR-Cas 系统的存在限制了外来 DNA 的侵入，但在某些条件下，获得新功能和特殊性状对微生物具有极大益处，例如抗生素抗性能力[43]。因此，我们研究了不同菌株中在编码 Cas 操纵子蛋白的 MAGs 中抗生素抗性基因（antibiotic resistance genes, ARGs）的检出频率。在几个微生物门类中，编码 Cas 蛋白的 MAGs 与不编码的 MAGs 相比，ARG 的频率显著降低，表明这些微生物门类的典型生境条件有利于 CRISPR-Cas 防御系统的选择（图 2.4E）。例如，经常出现在热液喷口环境中的嗜热菌 Thermoplasmatota 和 Halobacteriota（图 S2.10）[15]，以及在厌氧或与宿主相关生境中常见的细菌门，如 Patescibacteria、WOR-3、Gemmatimonadota、Marinisomatota 和 Firmicutes 等[44]。然而，在其他门类中并没有发现这种现象（图 2.4E）。我们进一步分析基因组中 Cas 操纵子的数量对 ARG 数量的影响。结果发现基因组中编码的 Cas 操纵子数量限制了基因组可能编码的 ARGs 数量的上限（Nested ANOVA test, $p < 0.001$ ）。因此，随着 Cas 操纵子的增多，基因组可能编码的 ARGs 数量的上限呈下降趋势（Kruskal Wallis test, $p < 0.001$ ，图 2.4F）。此外，我们对基因组中的其他移动遗传元件（mobile genetic elements, MGEs），例如质粒、整合共轭元件、插入序列、转座子和整合子也进行了分析，结果显示，与 ARGs 情况相似，MGEs 总数的上限也受到基因组中 Cas 操纵子数量的限制（图 S2.12）。

先前的研究已经报道了这两种不同的微生物防御系统之间复杂的相互作用。例如，在几种病原体微生物菌株中观察到了 CRISPR-Cas 和 ARG 之间的负相关趋势[45, 46]，但在大肠杆菌基因组中观察到 CRISPR-Cas 与磷霉素和利福平相关的 ARG 之间的正相关趋势[46]。对微生物的长期进化分析研究表明，CRISPR-Cas 系统的存在既不会对其他防御系统的存在产生积极影响，也不会产生消极影响[43]。因此，主动免疫系统与获取新遗传物质之间存在某种平衡 [43, 45-47]。并且，许多微生物编码抗 CRISPR-Cas 基因（anti-CRISPR genes, *acr*），其对应的产物能抑制 CRISPR-Cas 系统的功能，以避免自我靶向免疫，并在相应的环境选择压力下促进新功能的获得[45, 48]（图 S2.11）。在 GOMC 数据集中，与宿主相关 MPs 和热液喷口 MPs 相比，在远洋海水的 MPs 中观察到的编码 ARGs 的 MAGs 比例明显更高（图 2.4D）。然而，在对外源 DNA 的免疫更为重要的共附生相关 MPs 中观察到更高比例的 MAGs 编码 CRISPR-Cas 免疫系统（图 2.4D）。在与各种海洋生

态系统相关的大多数 MPs 中都观察到了两种免疫策略兼而有之的 MAGs，尽管它们的频率相对较低（图 2.4D）。未来，还需要从计算模拟（*in silico*）和体外研究（*in vitro*）两方面开展进一步的研究，以破译来自不同栖息地的微生物所依赖的看似相互矛盾的免疫系统策略之间复杂的相互作用。

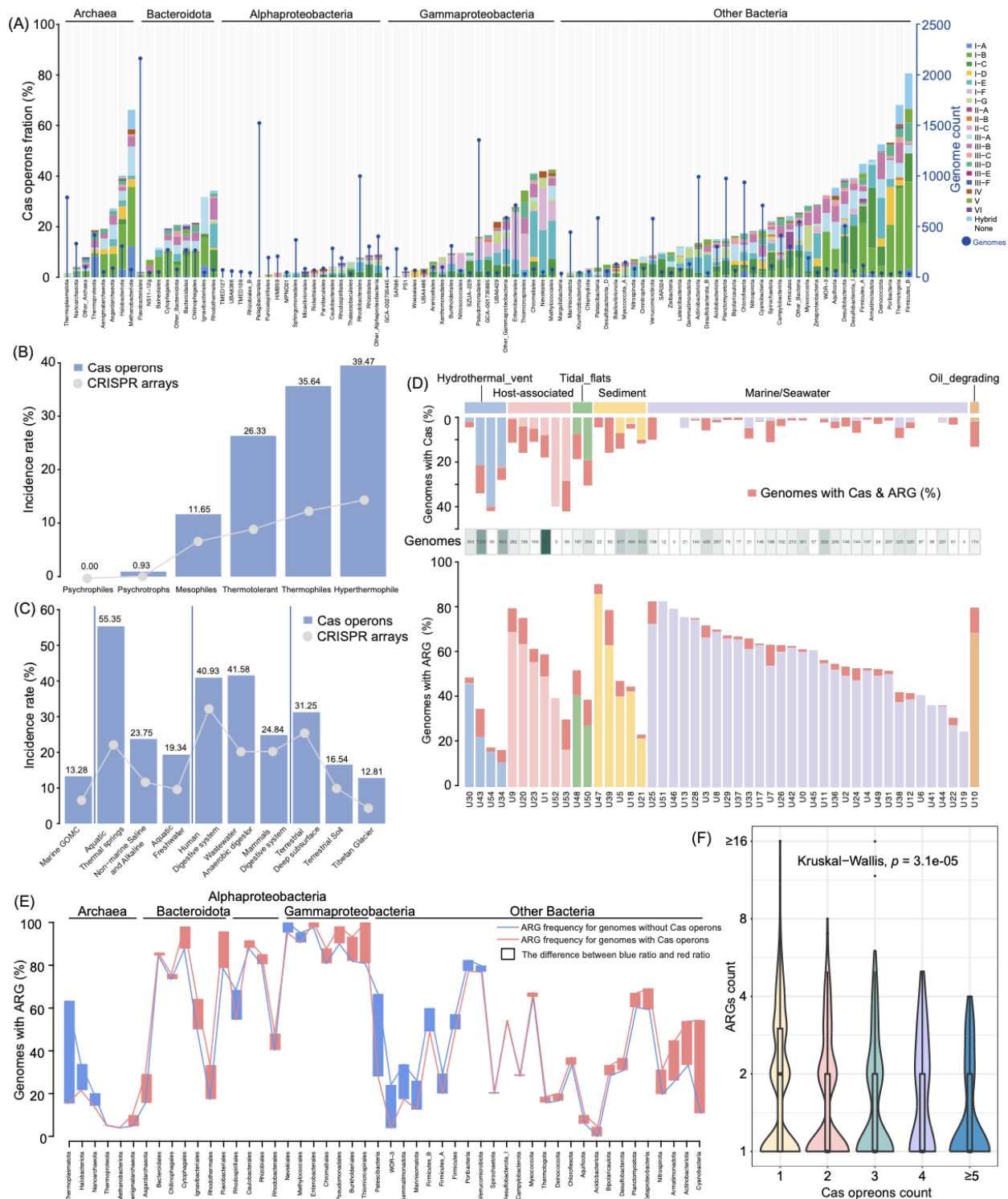


图 2.4. 防御系统的分布。(A) 柱状图显示了所有 GOMC 基因组中不同门类的 Cas 操纵子检出频率。图中只显示了基因组数超过 50 个的门类，蓝线代表每个类群的基因组数。(B 和 C) 直方图显示了不同最适生长温度 (optimal growth temperature, OGT) 基因组 (B) 和不同生态系统基因组 (C) 中 Cas 操纵子的检出率。灰线显示的是 CRISPR-Cas 系统的检出率。(D) 代表不同海洋生态系统的 MPs 中编码 Cas 操纵子 (上)

或 ARG（下）的 MAGs 数量（中间热图）和比例。（E）折线图显示了编码 Cas 操纵子（红线）或不编码 Cas 操纵子（蓝线）基因组中 ARG 的检出比例；箱线图表示这两个比例之间的差异，蓝色方框表示不编码 Cas 操纵子的基因组提高了 ARG 频率的比例，红色方框表示相反的情况。（F）随着 Cas 操纵子数量的增加，ARG 的上限数量呈下降趋势。

2.6 海洋微生物生物合成基因簇

海洋微生物可以合成具有重要的生态功能、生物技术和治疗用途的次级代谢物[7]。这些分子由许多相邻的基因编码，它们被称为生物合成基因簇（Biosynthetic Gene Clusters, BGCs）。我们从 24,195 个 GOMC 基因组中预测出了 66 种不同类型的 64,217 个 BGCs（图 S2.13A）。这些 BGCs 长度从 1,001 到 576,743 bp 不等，并且其长度与宿主基因组大小无显著相关性（图 S2.13A）。Terpene、RiPP-like、NRPS-like、NRPS 和 RRE-containing 基因簇是海洋微生物中最主要的 BGCs，这与先前发表的一项关于海洋微生物 BGCs 的研究结果一致[7]。此外，考虑到 BGCs 固有的冗余性和不完整性，我们将所有 BGCs 聚类成 13,063 个基因簇家族（gene cluster families, GCFs）（图 2.5A）。其中，来自 5,793 个 GCF 的 16,369 个 BGCs（约 25.49%）被鉴定为新型 BGC，其与参考数据库（BiG-FAM）中已注释的 GCFs 相关性较低（余弦距离 > 0.2）[48]，这些新型 BGC 中约 60.83%是由我们研究中的新 MAGs 编码的。大多数新型 BGCs 来自变形菌门（38.36%）和拟杆菌门（10.65%）（图 S2.13B），其中 RiPPs（43.12%）和 Terpene（23.12%）是最主要的类型（图 S2.13C）。此外，共鉴定出 419 个古菌 GCFs，其中 233 个为古菌特异性 GCFs，它们主要来自 Halobacteriota 和 Thermoplasmatota；细菌的 GCFs 种类更为丰富，占鉴定出的所有 GCFs 的 96.79%（图 2.5B）。变形菌、放线菌和厚壁菌的 GCFs 多样性最高，并且其中超过 80%的 GCFs 具有门特异性（图 2.5B），这意味着某些次级生物大分子的生物合成可能仅会出现在特定的类群中。

由于 GCF 在不同微生物门类中的分布并不均衡，我们评估了 GCFs 在整个微生物系统不同种系中的多样性，来寻找编码最丰富次生代谢物的微生物。我们根据观察到的数据，通过外推算法预测了所有门类的 GCF 编码潜力，发现变形菌门、放线菌门、拟杆菌门和浮霉菌门具有最高的次生代谢物生物合成潜力（图 2.5C 和图 S2.13D）[48]。我们进一步研究了每个较低分类水平的 BGC 合成能力，并对每个分类水平编码的 GCFs 数量进行了

方差分析（图 S2.13E）。与之前观察结果一致，我们发现在属水平的方差显著减小[48]。因此，属可能是展现整个微生物系统演化过程中 GCF 编码潜力的最适分类层级。在属水平上，我们发现链霉菌属（*Streptomyces*）、小单孢菌属（*Micromonospora*）和假单胞菌属（*Pseudomonas*_E）是 GCF 多样性最高的三个细菌属。在当前的数据集中，放线菌属链霉菌和小单孢菌富含与 NRPS 和 PKS 相关的 GCFs，而假单胞菌属的 GCF 则包含 NRPS 和 RiPPs。与之前在链霉菌、小孢子菌和假单胞菌中的研究结果一致[49-51]，这三个菌属的多种 NRPS 能够合成具有广泛生物活性的天然产物，包括抗菌和抗肿瘤化合物，因此可以作为发现新型化学物质的靶标。

此外，我们从通常具有各种抗菌和抗肿瘤活性的 RiPPs 中挖掘潜在的新型抗菌肽（antimicrobial peptides, AMPs）[52]。在本研究中，我们通过使用深度学习模型，共从 629 个 RiPPs BGCs 中鉴定出 1,079 个核心肽，其中从 115 个 BGCs 中发现了 133 个候选 AMPs（图 2.5D）。它们主要来自放线菌门（36 个 AMPs）、厚壁菌门（31 个 AMPs）和变形菌门（21 个 AMPs）的 anthipeptide-class-II 和 anthipeptide-class-I BGCs（图 2.5E）。在 133 种候选 AMPs 中，有 4 种属于已知的 AMP（Planosporicin、Micrococцин、Subtilin 和 Roseocin），而其他 129 种候选 AMPs 则极有可能是新型 AMPs。进一步研究这些潜在抗菌肽 AMPs 的来源基因簇 BGCs 和基因组表明，在本研究中鉴定的新型 BGCs 中，总共鉴定出 16 个 AMPs（占 11.38%），其中 60%来自新基因组，这表明新型海洋细菌基因组具有挖掘新型 AMPs 的巨大潜力，也提示了海洋原核生物可能存在尚未发现的新型抗生素[53]。

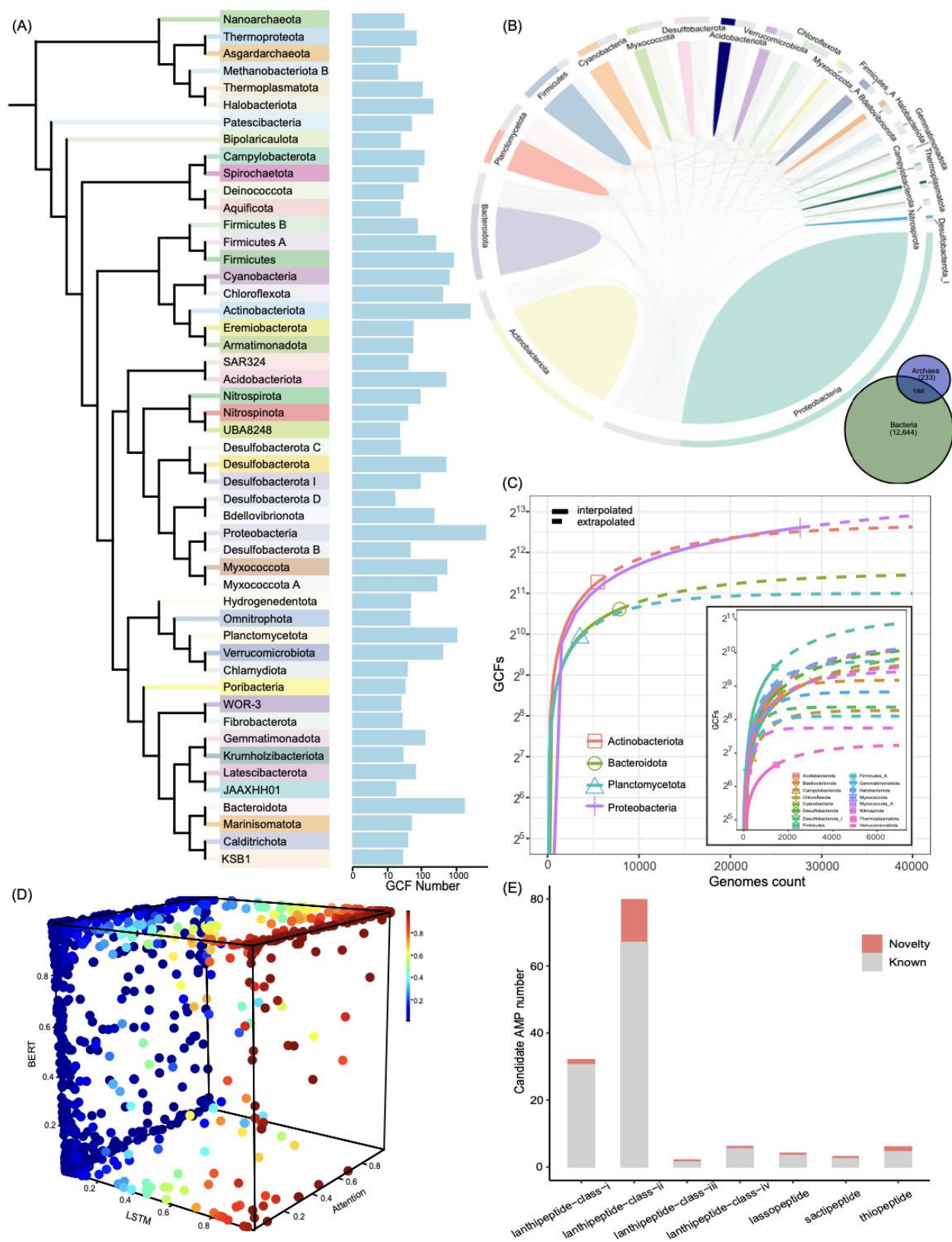


图 2.5. 生物合成基因簇。（A）各个门类生物合成基因簇的比较。条形图显示了各门中检测到的特有 GCFs 的数量。（B）显示各门特有 GCFs 的环状图（实心图形）和各门之间成对重叠的 GCF 图（带状图形）。维恩图显示细菌和古菌领域的 GCF 重叠。（C）预测生物合成潜力最大的前 4 个门类（嵌入图显示前 20 个门类中的其他 16 个门类）的稀释曲线。（D）深度学习模型 attention、LSTM 和 BERT 对 1079 种核心肽的预测得分。（E）柱状图显示了 133 种候选 AMPs 中 RiPPs 亚型的新颖性或已知数量。

2.7 体外实验验证 PET 高活性水解酶

除海洋微生物基因组数据库（GOMC）之外，我们还预测了所有宏基因组组装序列上的开放阅读框及其对应的蛋白质序列，并构建了全球海洋蛋白质序列数据库（Global Ocean Protein Catalogue, GOPC）（图 S2.1）。GOPC 包含超过 24.58 亿个非冗余蛋白，约为之前海洋微生物参考基因集（OM-RGC_v2）的 52 倍[54]。全球海洋蛋白质序列数据库不仅可以作为功能基因注释的参考数据库，更重要的是，它为生物技术应用领域（如持久性废物的生物降解）的各种天然酶/产品的生物勘探提供了宝贵的资源。

自从在 PET 消化细菌菌株中发现一种新型 PET 水解酶（*IsPETase*）以来，聚对苯二甲酸乙二醇酯（polyethylene terephthalate, PET）的酶解日益受到关注[55-57]，同时，定向进化等技术大大提高了 PET 材料降解和再循环的催化效率[58-60]。考虑到 PET 水解酶在不同生态系统中普遍存在，以及塑料废弃物对海洋环境日益严重的威胁[61, 62]，我们关注海洋（尤其是深海）微生物是否具有降解 PET 的能力。我们以 *IsPETase* 序列为参考，对 GOPC 进行了有针对性的搜索，来寻找潜在的新型 PET 水解酶[55]。我们从不同的海洋生态系统中发现了 1,598 个包含 PET 水解催化三联体 Ser-Asp-His 的候选 *IsPETase* 序列（图 S2.14A）[63]。这些序列显示出明显的系统发育多样性，但其形成的不同支系不受地理起源的限制（图 S2.14B、C）。为了寻找对极端环境条件具有高度耐受性和稳定性的 PET 水解酶，我们聚焦于与极端海洋环境相关的候选 *IsPETase* 序列[64]，选择了三个来自海斗深渊的序列和另外三个来自热液喷口的序列（*dsPETase*）在大肠杆菌中进行异源表达，并进行体外生化鉴定。

我们以常用 GfPET 薄膜（ES301445，Goodfellow）为底物，测试异源表达的六种 *dsPETase* 的水解活性。如之前的研究中所报道，水解主要产物（包括单（2-羟乙基）对苯

二甲酸（MHET）和对苯二甲酸（TPA）的总浓度（图 2.6A）来代表其催化活性[59, 65]。在六个候选酶序列中，我们发现三个嗜盐 PET 水解酶（从 North Su 热液喷口发现的 dsPETases05 以及从马里亚纳海沟发现的 dsPETases01 和 dsPETases06）对 GfPET 薄膜表现出卓越的催化活性，尤其是在高浓度 NaCl 的条件下（图 S2.15）。它们的催化活性随着 NaCl 浓度的升高而增加，37 °C 反应中在 4.5 M 或 5.3 M NaCl 的条件下催化活性最高，分别比 *IsPETase* 高 12.0 倍、16.0 倍和 5.6 倍（图 2.6B）。同时，在不同的盐度条件下，dsPETase02、dsPETase03 和 dsPETase04 没有观察到明显的催化活性（图 S2.15）。有趣的是，热液喷口来源的 dsPETase05 的最适温度（55 °C）高于海沟水来源的 dsPETase01 和 dsPETase06（40 和 45 °C）。如图 S15 所示，与不耐盐的 *IsPETase* 相比，这三种嗜盐 dsPETase 在最佳盐度和温度条件下的活性高出 5.6 至 44.3 倍（图 2.6C）。为了进一步观察可见的 PET 解聚过程，我们使用活性最高的 dsPETase05 和已验证催化活性的 *IsPETase* 来降解实验室制备的平均厚度为 28 μm 的薄膜（in-lab prepared solvent cast PET, scPET）。与 *IsPETase* 相比，dsPETase05 在 3 天的孵育过程中对 scPET 薄膜的降解更为明显（图 2.6D）。值得注意的是，在 500 nM dsPETase05 浓度条件下培养 3 天后，所有 scPET 都被分解，解聚率达到 83%，高于 *IsPETase* 的解聚率（41%）。当使用 300 nM 的酶时，dsPETase05 的解聚率为 41%，也高于 *IsPETase* 的 27%。通过对具有不同 PET 降解性能的嗜盐 PET 水解酶进行序列分析和实验验证，证明了 GOPC 作为潜在数据库资源的价值，GOPC 数据库可进一步应用于生物技术应用中各种功能的生物勘探等研究。

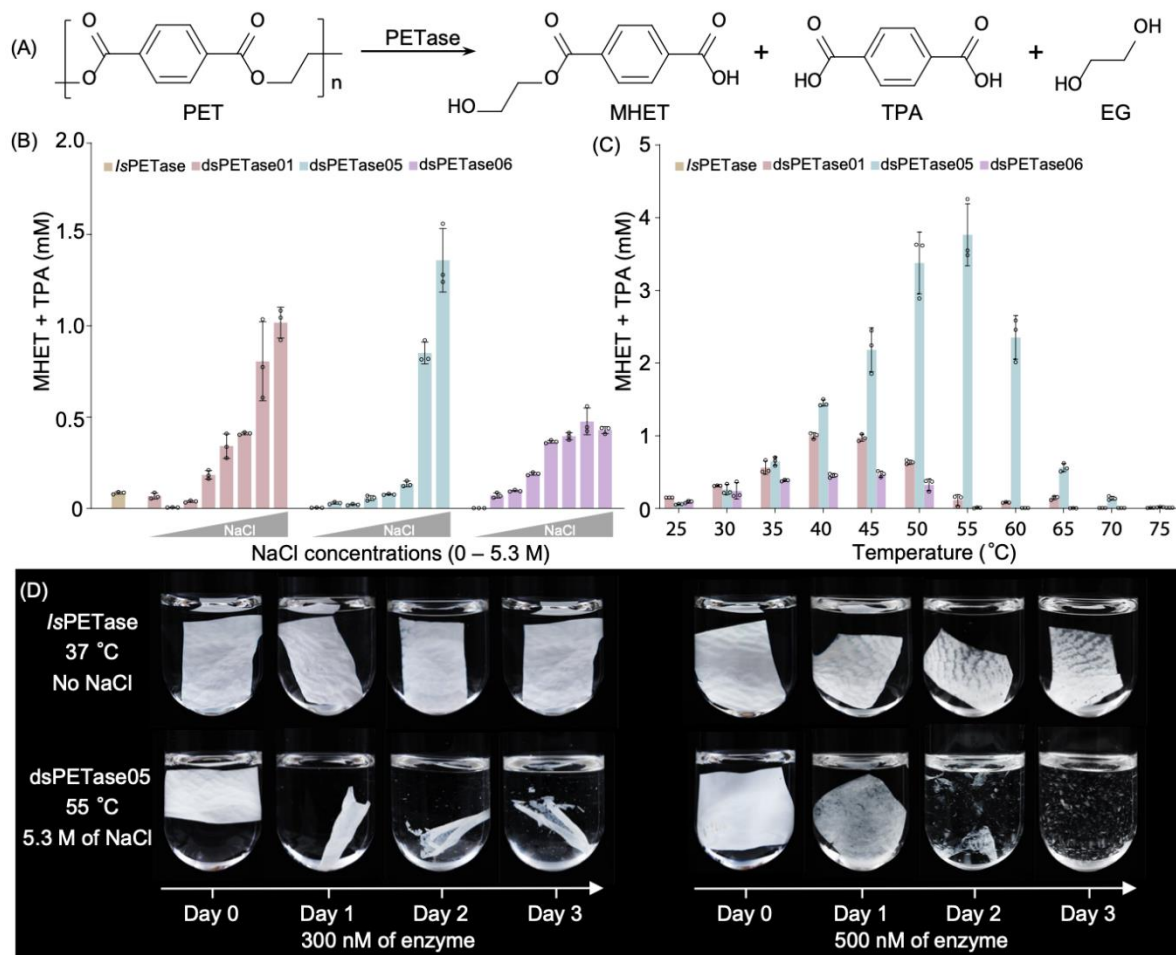


图 2.6. 嗜盐 dsPETase 的水解活性。 (A) PET 酶催化 PET 解聚，主要产生 MHET、TPA 和 EG 等可溶性产物。(B) 深海来源 dsPETase 的嗜盐特性。对 GfPET 薄膜 ($\Phi = 6 \text{ mm}$) 的水解活性由总释放产物的浓度 (通过 HPLC 分析的 MHET 和 TPA 的总和) 来表征。50 nM dsPETase 在 pH 9.0 的 Tris-HCl 缓冲液中进行 120 小时的催化反应。每个 dsPETase 反应都加入了一系列不同浓度的 NaCl，分别为 0、0.6、1.2、1.9、2.8、3.7、4.5 和 5.3 M (接近饱和)。所有反应均在 37°C 下进行，一式三份。横杠和圆圈分别代表平均值和单个值。同时测定了无 NaCl 时 IsPETase 的活性作为参考。(C) 三种嗜盐 dsPETase 在不同反应温度下对 GfPET 薄膜的水解活性。反应开始时，将酶加入根据 (B) 推断出的最佳盐度缓冲液中，dsPETase01 和 dsPETase05 为 5.3 M NaCl，dsPETase06 为 4.5 M NaCl。所有反应均一式三份。(D) 嗜盐 dsPETase05 对 scPET 薄膜的可见降解。根据 (B) 和 (C) 的盐度和温度条件设定为最佳反应条件。以 IsPETase 在 37°C 不含 NaCl 的 Tris-HCl 缓冲液 (pH9.0) 中催化的反应为参考。每个样品中，将 3mg scPET 与 300 或 500 nM 的酶 (如所示) 在 3 毫升的总体积中进行孵育。

2.8 讨论与展望

我们收集并分析了分布在世界各地的海洋宏基因组和海洋微生物基因组，从而建立了迄今为止最全面的基于宏基因组学解析的海洋微生物基因组数据库。我们的数据库包括 24,195 个物种水平基因组，代表了海洋原核生物基因组的百科全书式数据库，因此有助于全球研究人员进行从生态层面到生物技术挖掘层面的多种分析。例如针对潜在高活性生物酶的筛选，深海黑暗、高压、低温、高盐等极端复杂的环境蕴含具有结构新颖、功能多样的生物活性物质，是一个天然的极端酶宝库和酶筛选平台，结合人工智能的方法可以实现海洋源微生物序列—功能、或序列—结构及结构—功能的关联，从而使以序列作为输入，通过计算模拟预测酶的结构，从而推断其活性及功能、实现各类分子工具酶及工业酶等的高通量挖掘成为可能。因此，未来的分析研究不仅将从我们最全面的海洋微生物基因组数据中受益，还将受益于以前数据库中匮乏的 9,939 个原核生物基因组所代表的生物新特性。

2.9 补充图表

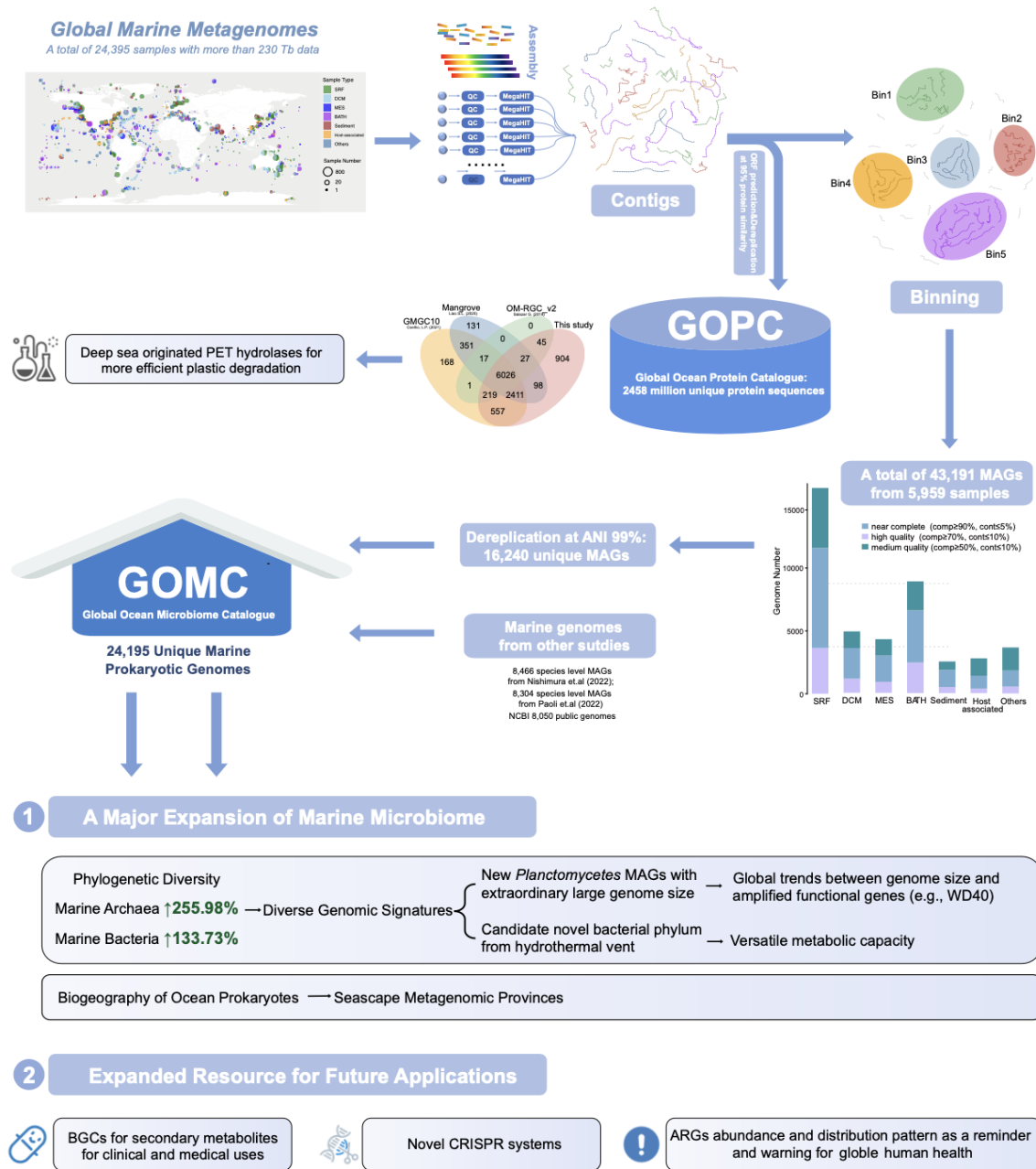


图 S2.1. 本研究概述和流程示意图。收集全球分布的海洋宏基因组并重新进行组装及分箱分析，以重构海洋微生物基因组（metagenome-assembled-genomes，MAGs）。同时下载了之前公共数据库（NCBI、OMD 和 OceanDNA）中的微生物基因组，并将其与新重构获得的 MAGs 汇总，构建了一个全面、完整的全球海洋微生物组数据库（Global Ocean Microbiome Catalogue，GOMC），其为下游分析和未来研究的参考数据库。与之前公开的海洋微生物基因组数据库相

比，我们新重构的 MAGs 显著扩大了已知的海洋微生物多样性。我们基于新组装的序列进行基因预测及去冗余，从而构建了一个非冗余全面的全球海洋微生物蛋白质数据库（Global Ocean Protein Catalogue , GOPC）。维恩图显示了 GOPC KEGG 功能与之前发表的其他基因集的异同。

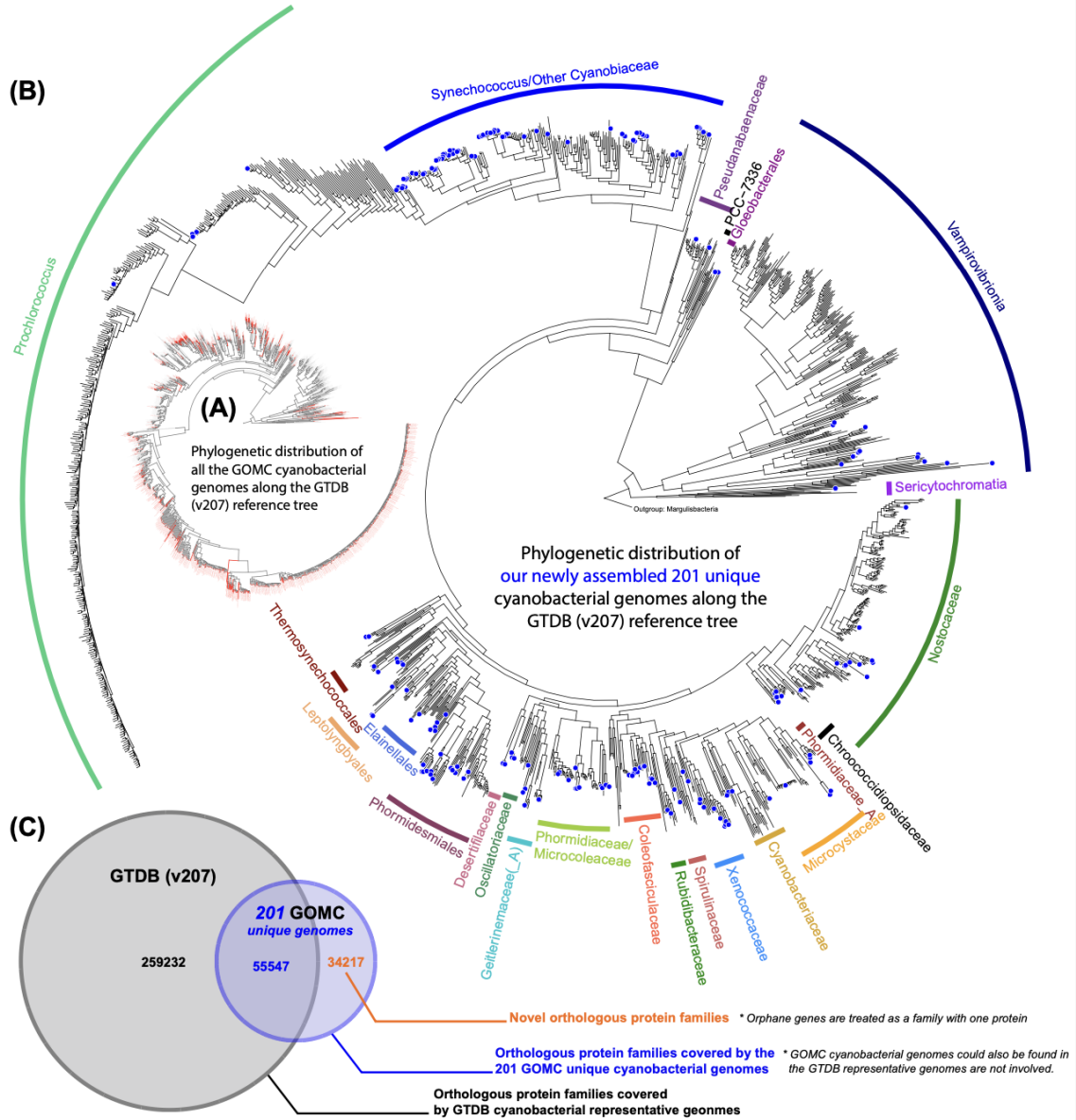


图 S2.2. 蓝藻系统发育树及其基因组间共有的同源蛋白家族数量。(A) GOMC 全部 722 个蓝藻基因组与 GTDB (v207) 参考基因组的系统发育分布。红色分枝代表 GOMC 基因组。(B) 将 GOMC 独有的 MAGs 与 GTDB 参考系统树整合而得到的蓝细菌系统发育树。代表 GOMC 新 MAGs 的分枝用放大的蓝点标出，而其他已被纳入最新 GTDB 数据库的 GOMC 蓝细菌 MAGs

单系支系，树旁显示了基因组的预测最适生长温度（optimal growth temperature , OGT）。(B) 候选新门类中三个 MAGs 的地理分布和相对丰度。(C) 基于基因组信息的候选新门类的主要碳代谢潜能和跨膜转运体及能量守恒系统。发现了参与糖酵解和葡萄糖生成的基因，以及利用甘露糖和果糖的能力。这些基因组编码了部分 Calvin-Benson-Bassham (CBB) 途径，但缺乏几种必需的酶，如磷酸布洛激酶（phosphoribulokinase）和用于碳固定的核酮糖-二磷酸羧化酶（ribulose-bisphosphate carboxylase）。我们还发现了一种非典型的 Wood-Ljungdahl pathway (WLP)，其中缺少典型 WLP 中的三种关键酶，包括甲酸脱氢酶（fdh）、CODH/ACS 酶复合物的一个亚基（acesB）和甲基转移酶（acesE）。这类新细菌可能不会将 CO₂ 还原成甲酸盐，而是能够结合外源甲酸盐和/或通过四氢叶酸（THF）直接介导一碳（C₁）代谢。另一条天然 CO₂ 同化途径——还原甘氨酸途径（RGP）能完全鉴定到，并与 WLP 相结合，支持下游的甘氨酸、丝氨酸、半胱氨酸和甲硫氨酸生物合成。与促进后续氨基酸合成和平衡的甲基分支不同，羰基分支因 CODH/ACS 复合物的缺失而中断。此外，还发现了一种[FeFe]电子分叉氢化酶和一种潜在的一氧化碳脱氢酶铁硫亚基（coof）。WLP 被认为是最古老的厌氧碳固定途径之一，已随着宿主基因组进化了数百万年。因此，这里观察到的非典型不完全途径可能反映了对特殊生态位的适应，并有助于更好地了解 WLP 在海洋环境中的多样性和分布。此外，还检测到参与脂肪酸代谢的酶，包括长链酰基-CoA 合成酶、酰基-CoA 脱氢酶、丁酰-CoA 脱氢酶、3-羟基酰基-CoA 脱氢酶和乙酰-CoA 酰基转移酶。此外，MAGs 还编码了丁酸和长链脂肪酸代谢的完整 β -氧化途径，以及丙酸代谢的乙酰丙二酰-CoA（ethylmalonyl-CoA , EMC）途径。这三个基因组都编码 TCA 循环所有步骤的酶和可逆的 2-氧代戊二酸铁氧还蛋白氧化还原酶。

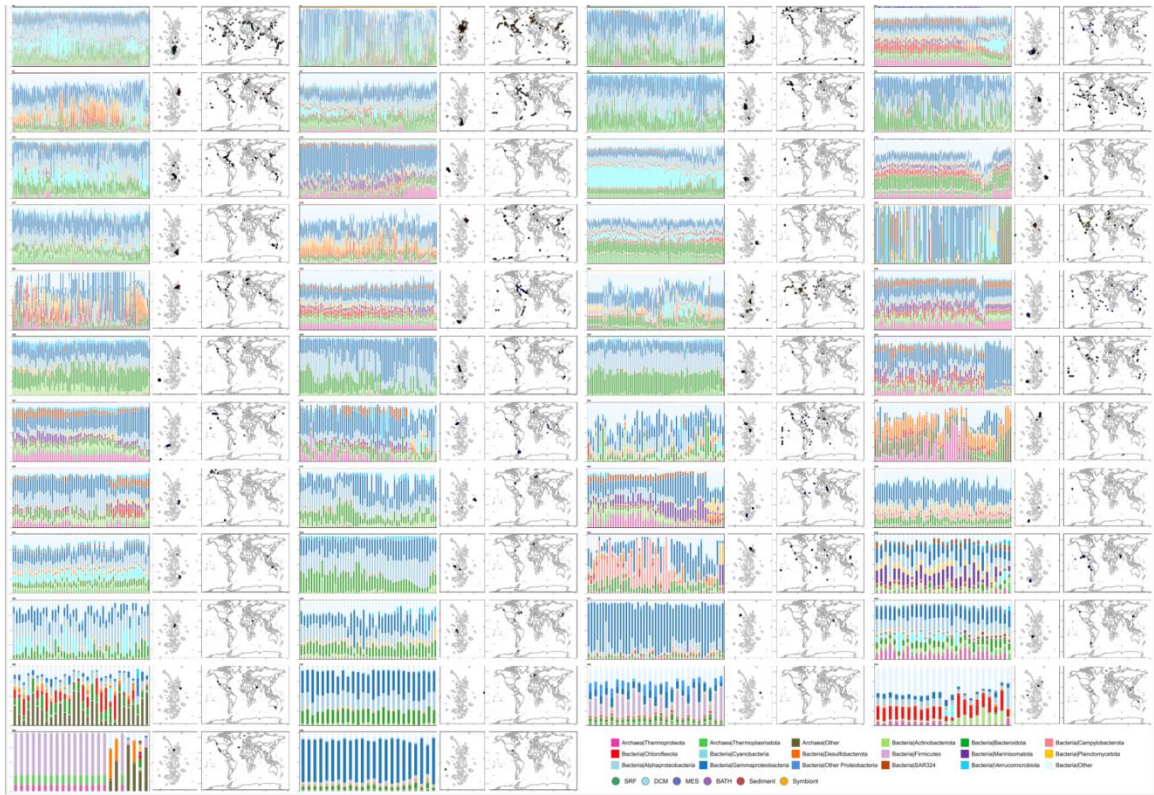


图 S2.4. 宏基因组分区（metagenomic provinces, MPs）的物种分类和生物地理分布特征。每个 MP 由三个图组成，从左到右分别显示了每个样本的微生物群落组成、样本在 UMAP 降维空间中的位置以及该 MP 在全球海洋中覆盖的地理区域。分类群的配色方案与图 2A 相同，并显示在本图的底部。在第一幅图中根据样本与 UMAP 空间中 MP 中心的距离进行排序。在每个样本的上方叠加条形图，用不同颜色的点表示样本所属的水层深度。深度的配色方案也显示在图的底部。每个点的不透明度与特定样本与 UMAP 空间中 MP 簇中心的距离成正比。样本离中心越近，点越不透明。在第二和第三幅图中，每个样本的颜色和透明度与第一幅图中的点相同。

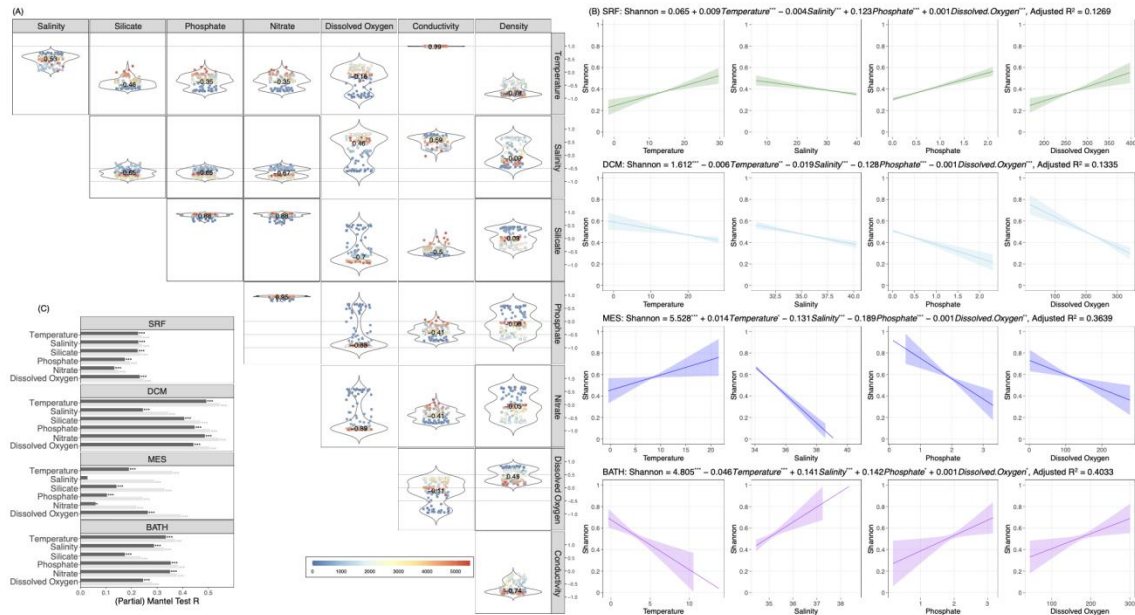


图 S2.5. 评估环境参数和随机过程对样本间群落多样性和分布模式的影响。(A) 不同环境参数之间的斯皮尔曼 (Spearman) 相关性, 以为多元回归模型的最佳拟合选择适当的预测环境参数。小提琴图上方标出了各比较值的中位数。不同水层深度的相关性是分别计算得出的, 各个点根据其深度来源进行着色的。深度的配色方案显示在底部。(B) 多元回归分析各选定环境参数 (包括温度、盐度、营养元素 (磷酸盐) 和溶解氧) 对样本香农多样性的影响。每个预测变量的统计显著性水平用星号表示: * $0.01 < P \leq 0.05$, ** $0.001 < P \leq 0.01$, *** $P \leq 0.001$ 。在保持其他自变量均值不变的情况下, 通过为每个自变量绘制折线图, 可以直观地看出香农多样性与每个环境参数之间的关系。每幅图都显示了拟合线和 95% 的置信预测区间。(C) 基于 JSD 矩阵, 我们评估了每个环境参数对不同深度样本间 Beta 多样性的影响。条形图显示了 (partial) Mantel tests 的相关系数。浅灰色柱形图显示的是 Mantel tests 的结果, 深灰色柱形图显示的是考虑了地理距离的 partial Mantel tests 结果。

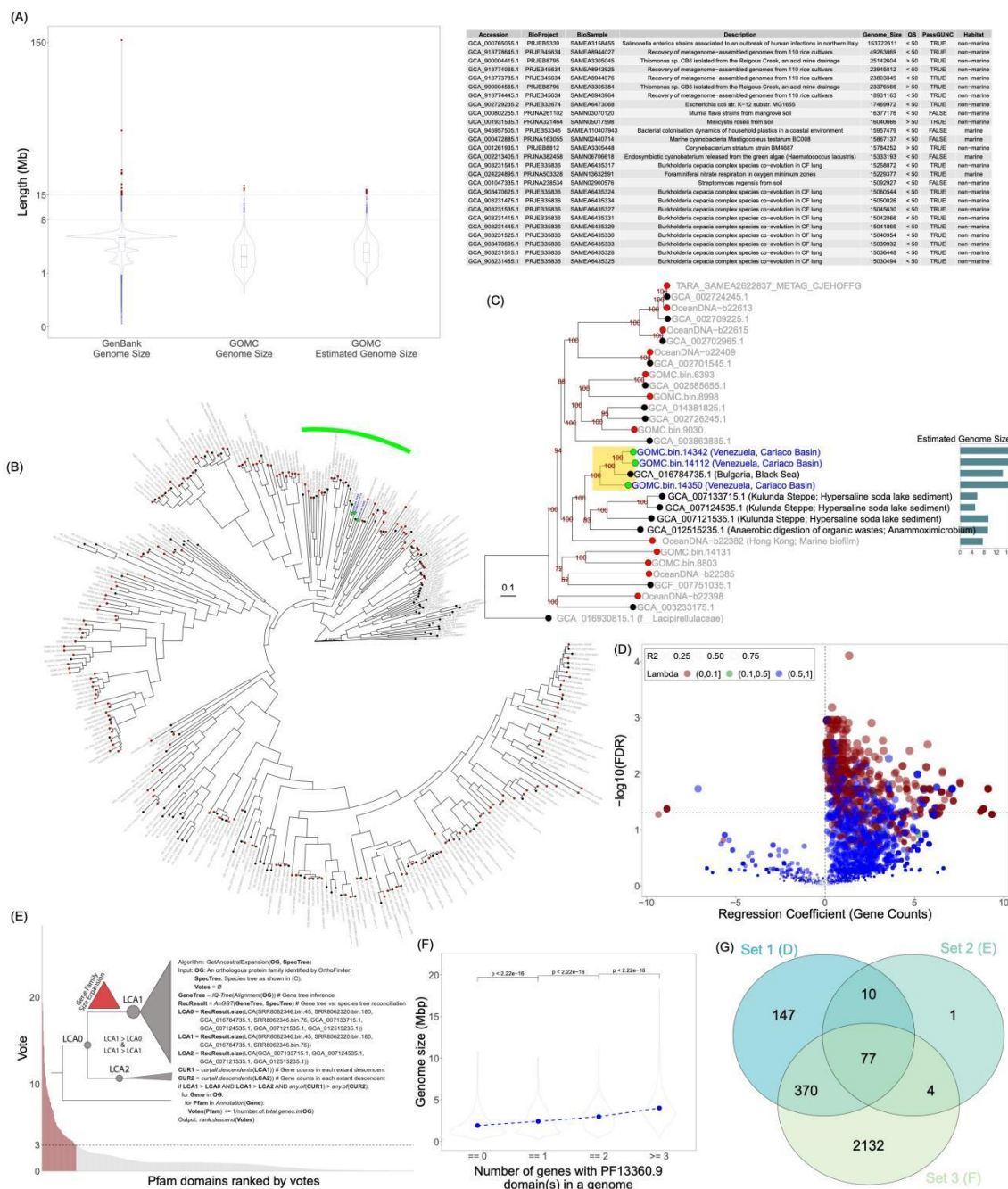


图 S2.6. (A) 小提琴图显示了 NCBI 和我们的 GOMC 数据库中已组装和预估基因组大小的分布情况。两条水平线分别代表 8 Mb 和 15 Mb 的基因组大小。基因组大于 15 Mb 的异常值用深红点表示。右侧表格显示了 26 个大小超过 15 Mb 的 NCBI GenBank 基因组的信息，包括基因组来源的项目和样本特性、分离来源以及栖息环境（海洋或非海洋）的描述、组装基因组大小、根据 CheckM 完整度和污染度计算的质量分数（quality score, QS），以及基因组是否通过了 GUNC 的嵌合检验。(B) 来自 GOMC 数据库的 192 个 Pirellulaceae MAGs 在 Pirellulaceae 科

基因组进化树上的分布。该树是使用 GTDB-Tk 工具包 (v2.1.1) 的 `de_novo_wf` 函数生成的。以 GTDB 数据库 (207 版) 中的基因组为参考并用黑点标出, GOMC MAGs 用红点标出。大于 16 Mb 的 3 个 *Pirellulaceae* MAGs 用大绿点突出显示。与它们相邻的基因组用绿色条带标出, 用于 (C) 的分析。(C) 最大似然系统发生树显示了 3 个 GOMC 超大 *Pirellulaceae* MAGs 与其近亲的关系。各分支的支持值以深红色数字表示。条形图显示了构成高置信度单系支系的 9 个基因组的大小, 其中包括我们的目标超大基因组。(D) 火山图显示使用系统发育回归分析 (phylogenetic regression, PR) 得出的基因拷贝数量与基因组大小的回归系数与其相应的 FDR 值。火山图中的每个点代表一个功能家族 (Pfam 结构域)。各点根据其相应的 λ 值着色, 表示系统发育回归信号的强度 (方法)。每个点的大小代表相应的 R^2 值。x 轴显示系统发育回归对每个 Pfam 的回归系数。y 轴显示 $-\log_{10}$ 转换后相应的 FDR 值。水平虚线表示 FDR 值为 0.05。PR 方法精确定位了 604 个与基因组大小呈显著正相关的 Pfam 结构域, 即使考虑到系统发育相关性 ($R^2 \geq 0.5$, $FDR < 0.05$, $\text{regression coefficient} > 0$), 在统计学上仍然是稳定的正相关。(E) 如何利用祖先蛋白重建结果选择扩展 Pfam 的示例概述。基于简约法则, 我们提出了一个广义的解析标准, 即特定蛋白质家族的扩张倾向于发生在四个超大基因组的最后共同祖先 (last common ancestor, LCA), 而不是在它们分化后各自独立扩张。按等级排列的柱状图说明了单个 Pfam 的投票得分分布, 其中 x 轴代表 Pfam 结构域, y 轴代表其对应的得票数。得票数高于 3 的 Pfam 结构域颜色更深。根据这一方法, 我们确认了 92 个随基因组增大而扩张的 Pfam 结构域。(F) 小提琴图展示了精确定位 Pfam 家族所采用的方法, 其中基因组越大, 注释到指定 Pfam 家族中的基因数量就越多。随后, 我们将分析范围扩大到整个 GOMC 数据集, 寻找基因组大小扩增随特定 Pfam 蛋白家族拷贝数一致增加的趋势 (Wilcoxon 秩和检验, $FDR < 0.05$), 结果检测到了 2,583 个 Pfam 结构域候选者。这里以 Pfam 域 “PF13360.9” 为例。在 x 轴上, 基因组中注释有 “PF13360.9” 结构域的基因数量分为四类: 0、1、2 和 3 或更多。y 轴代表基因组大小。整个小提琴图通过密度曲线描绘了每个等级 (0、1、2 和 ≥ 3) 对应的基因组大小分布情况。保留了使用 Wilcoxon 秩和检验在所有三个配对比较中均表现出统计学意义的 Pfam 家族。这些被选中的蛋白质族标志着在四个不同类别中呈现出一致的上升趋势。(G) 维恩图直观地表示了通过三种不同方法选出的 Pfam 之间的交叉点。

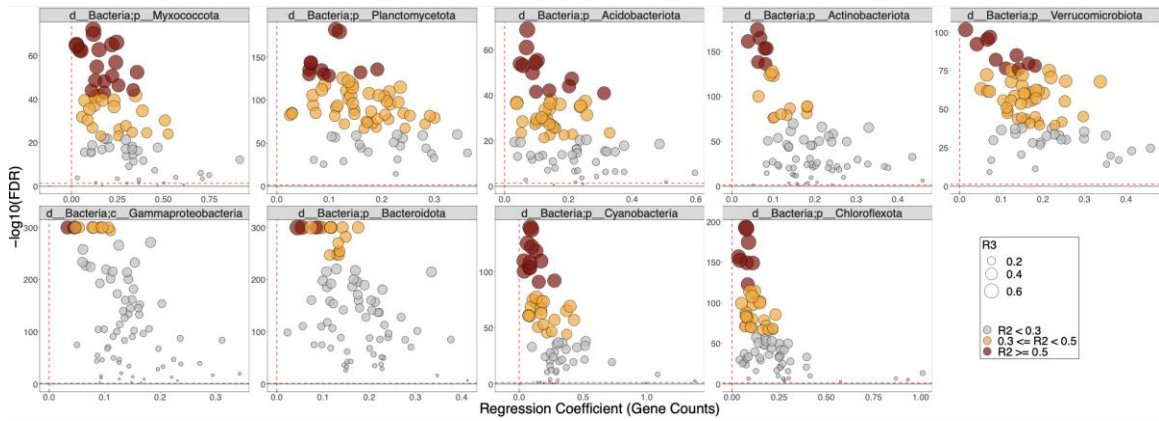


图 S2.7. 火山图显示了系统发育回归分析得出的基因数量回归系数与相应的错误发现率（false discovery rate, FDR）值的关系。这些分析使用了来自 9 个门类的高质量基因组，每个门类至少包括 5 个大基因组（ ≥ 8 Mb）。每个点的大小代表相应的 R^2 值。x 轴显示每个 Pfam 的系统发育回归系数。y 轴显示 $-\log_{10}$ 转换后相应的 FDR 值。水平虚线表示 FDR 值为 0.05。垂直虚线表示回归系数为 0。因此，淡蓝色阴影划定的区域内的数据点表示系数为正且 FDR 值小于 0.05。值得注意的是，与特定 Pfam 结构域相关的基因拷贝数差异对基因组大小差异的解释程度显示出相当大的可变性。

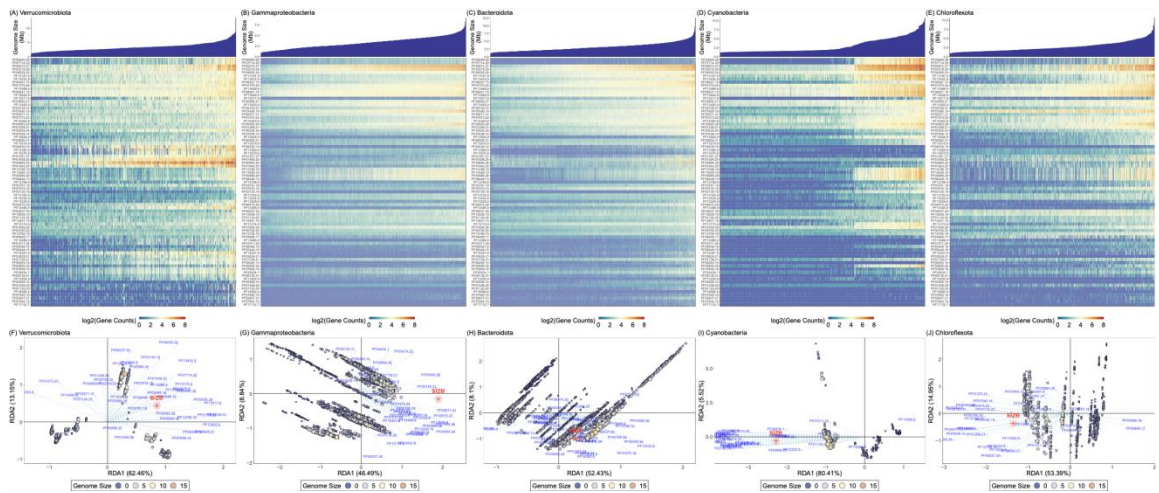


图 S2.8. 研究了 77 个指定的 Pfam 结构域中注释基因拷贝数量随基因组增大的变化趋势，基因组大小从大到小依次为疣微菌（A）、Gamma-变形菌（B）、拟杆菌群（C）、蓝细菌（D）和绿弯菌（E）。每张图的底部都标有颜色代码。此外，还进行了冗余分析（Redundancy Analysis, RDA），研究了这 5 个类群的基因组大小与 Pfam 之间的关系。值得注意的是，通过前向选择确定的显著相关性用蓝色箭头表示。门类的排列是根据其在高质量基因组（完整性 >

70%，污染小于 < 10%）中的平均估计基因组大小确定的。图中分析选取了 9 个门类，前 4 个门类如图 3 所示。

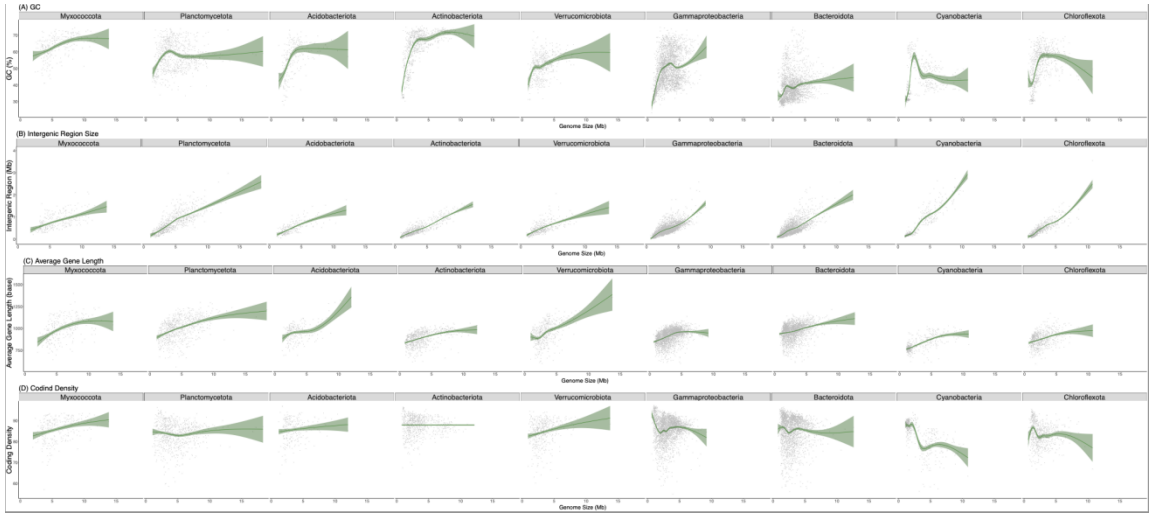


图 S2.9. 大基因组门类的基因组特征与基因组大小的相关性概览。

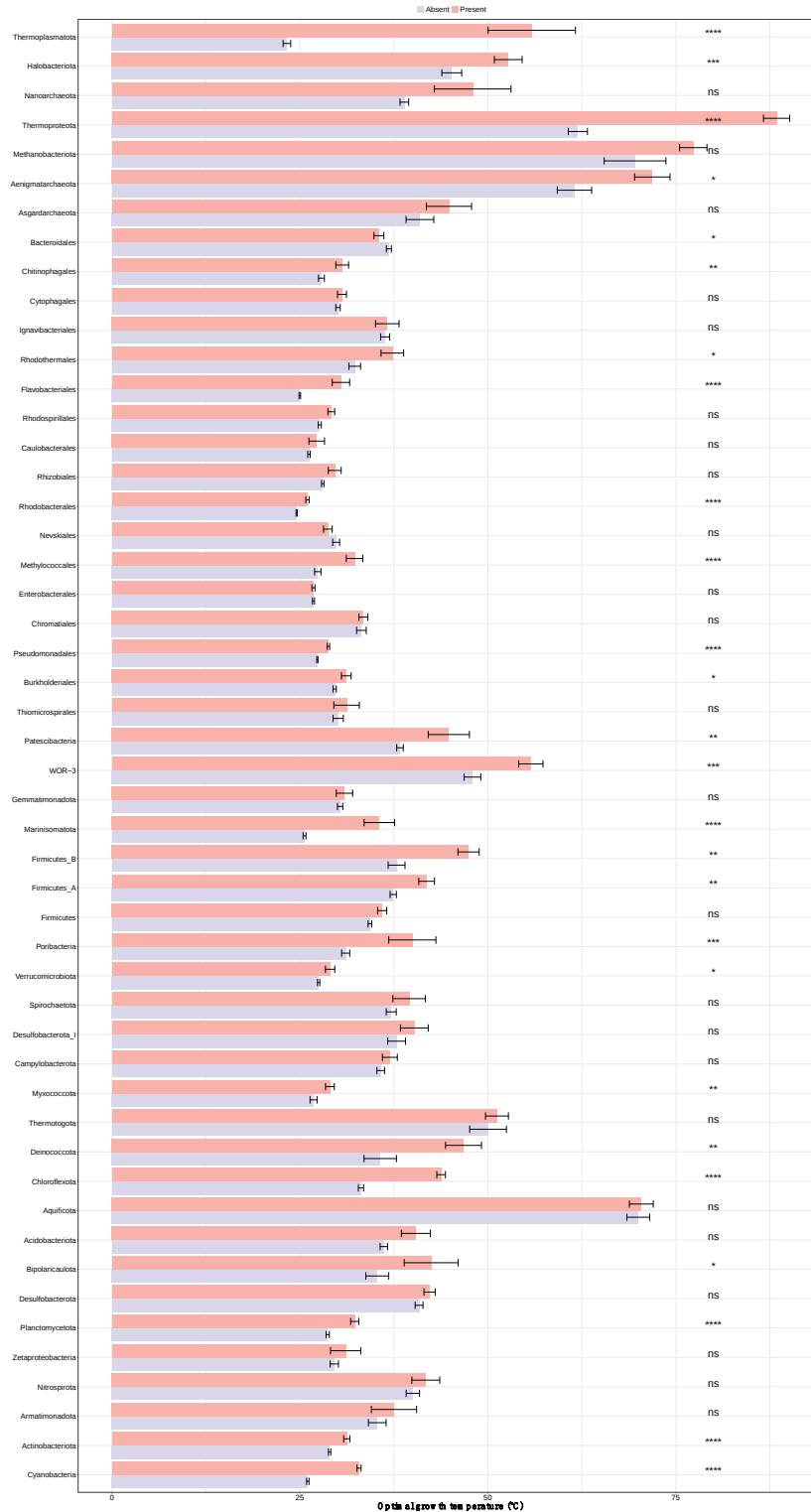


图 S2.10. 不同类群 GOMC 基因组中带有（红色条）或不带有（紫色条）Cas 操作子的预测最适生长温度（OGT）。图中仅列出了基因组数量超过 50 个的类群。

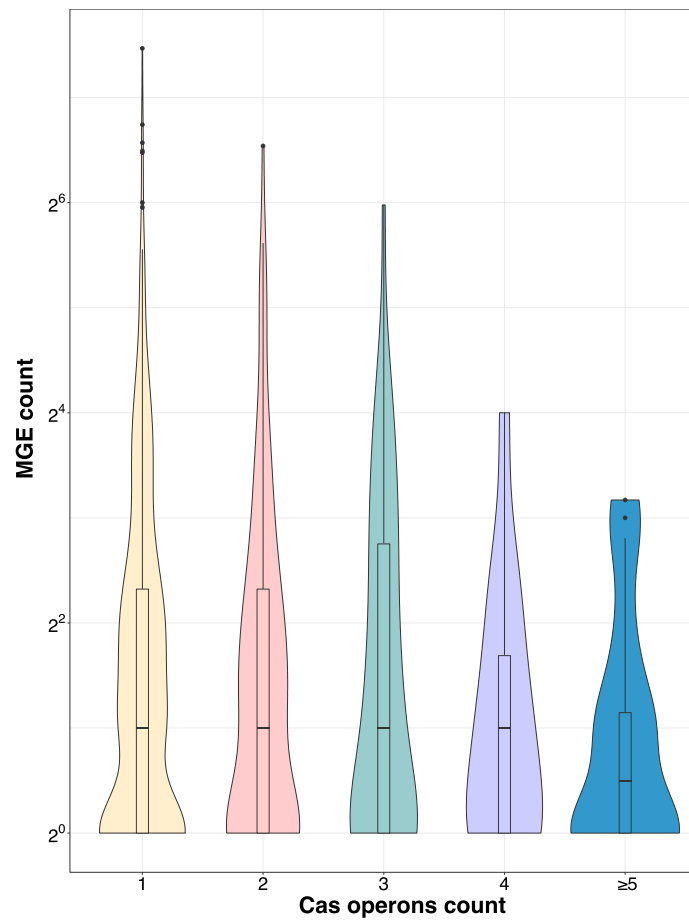


图 S2.12. 随着 Cas 操纵子数量的增加，移动遗传元件（MGEs）的上限数量呈下降趋势。

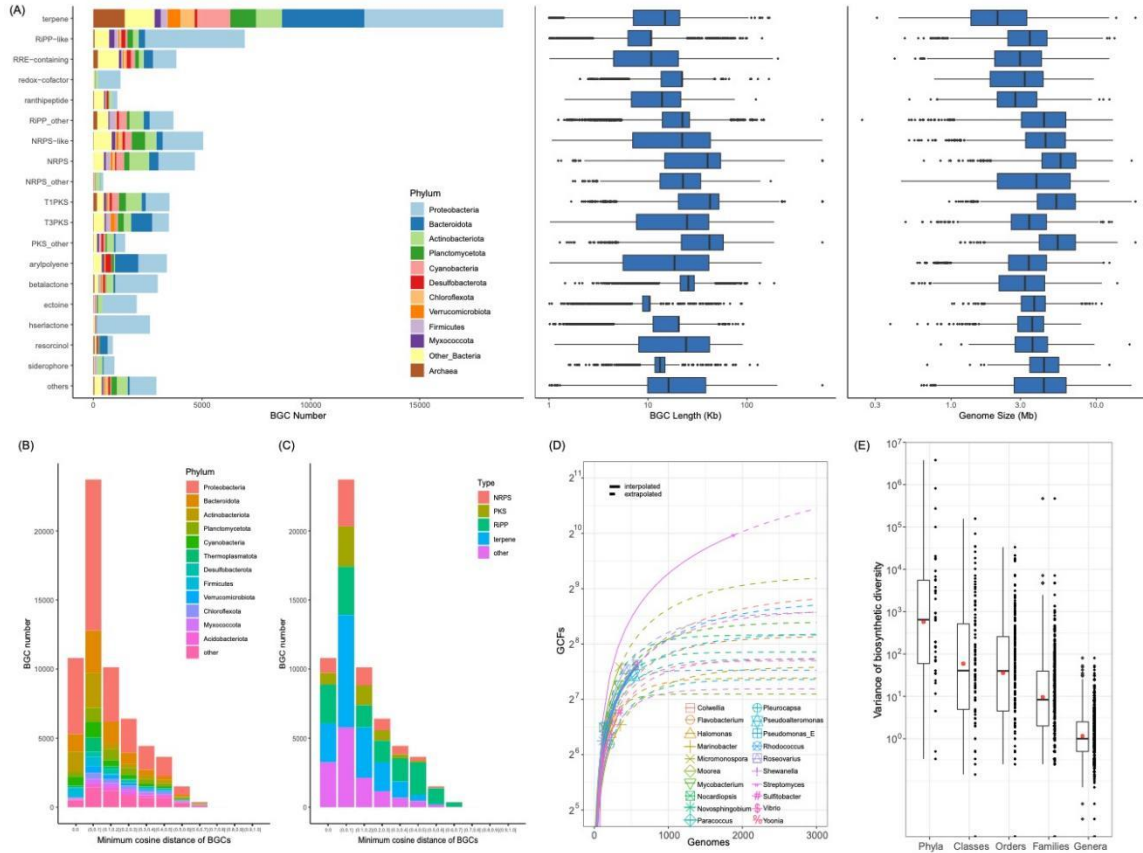


图 S2.13. GOMC 中微生物生物合成基因簇 (Biosynthetic gene clusters, BGCs) 的系统发育分布和多样性。(A) 从 24,195 个 GOMC 基因组中预测出了来自 66 个类型的 64,217 个 BGCs (左图), 长度从 1,001 到 576,743 bp 不等 (中图), 但与来源基因组大小无显著相关性 (右图)。(B-C) 将我们的数据集中发现的 BGCs 与 RefSeq BGCs 数据库 (由 BiG-FAM 构建) 进行比较, 发现了 16,369 个新的 BGCs (余弦距离大于 0.2)。大多数 BGCs 由 (B) 变形菌 (38.36%)、拟杆菌 (10.65%)、浮霉菌 (6.73%) 和放线菌 (5.24%) 预测得到, 主要编码 (C) RiPPs (43.12%)、萜类 (23.12%) 和其他假定天然产物 (15.18%)。(D) 稀释曲线显示含量最高的前 20 个细菌属的实际和潜在生物合成多样性, 表明放线菌门的链霉菌 (*Streptomyces*) 可能编码最高的生物合成多样性。(E) 从门到属不同分类等级基因组生物合成多样性的方差。观察到属的方差显著减小, 因此属是表征微生物生物合成潜力的合适分类等级。

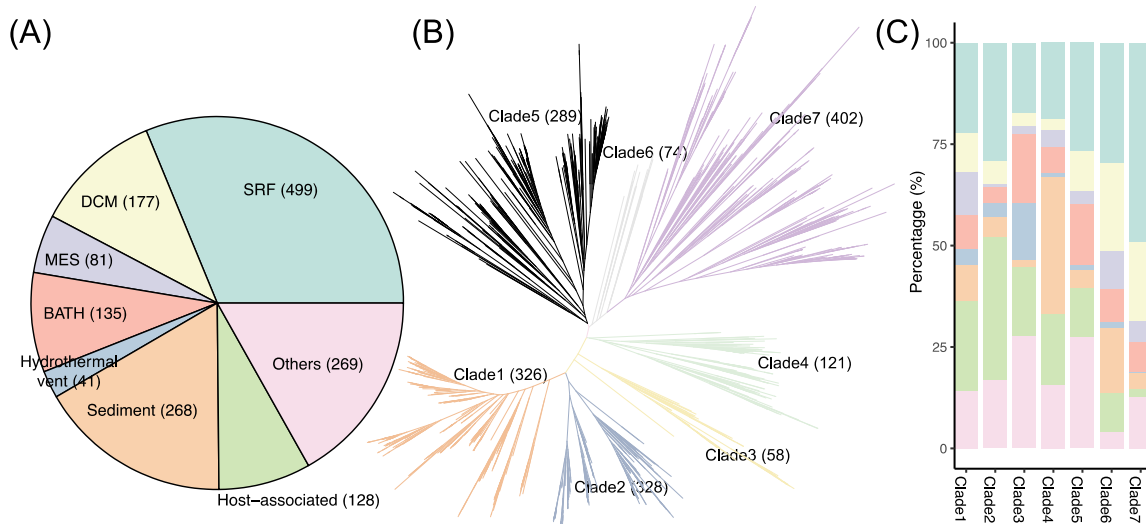


图 S2.14. 候选 *IsPETase* 的生物勘探。(A) 1,598 个具有 Ser-Asp-His 催化三连体的候选 *IsPETase* 在不同海洋生态系统中的分布。(B-C) 候选 *IsPETase* 的系统发育分析和不同支系的生态系统起源。所选的六个异源表达 *IsPETase* 序列分布在不同支系中，其中一个来自支系 3 (DMPETase06)，另一个来自支系 4 (DMPETase05)，其余 4 个 dsPETases 来自支系 5。

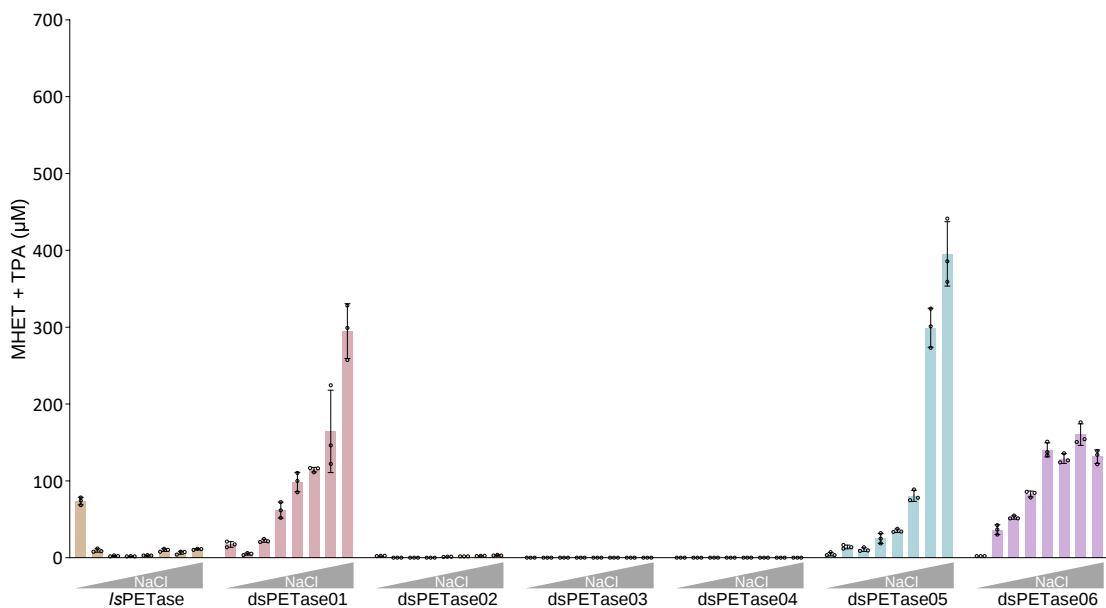


图 S2.15. 各种条件下 PET 酶活性的体外验证。所有反应都是在 pH 值为 9.0 的 Tris-HCl 缓冲液中进行的，其中含有一系列 NaCl 浓度，包括 0、0.6、1.2、1.9、2.8、3.7、4.5 和 5.3 M。GfPET 薄膜 ($\Phi = 6$ mm) 与 50 nM 酶在 37°C 孵育 48 小时后，用 HPLC 分析水解的总产物 (TPA 和

MHET) 的浓度, 作为 PET 水解活性的代表指标。所有反应均做三次重复。横杠和圆圈分别代表平均值和个体值。

表 S2.1. 从公共数据库中收集海洋原核生物基因组的分类和关键词。

NCBI_taxonomy	NCBI_taxid	EBI/JGI Keywords
algae_metagenome	1300146	ocean
ballast_water_metagenome	1954210	seawater
beach_sand_metagenome	412757	sea-water
brine_metagenome	1981201	sea
cetacean_metagenome	1822005	marine
ciliate_metagenome	1969832	bay
cold_seep_metagenome	1583376	coast
coral_metagenome	496922	deepsea
coral_reef_metagenome	471232	deep-sea
crab_metagenome	1660082	coral
crustacean_metagenome	1681198	shore
ctenophore_metagenome	1508044	seashore
desalination_cell_metagenome	1983455	microalgal
dinoflagellate_metagenome	1579005	algal
echinoderm_metagenome	1411990	hydrothermal
estuary_metagenome	1649191	estuary
eukaryotic_plankton_metagenome	2315767	seep
flotsam_metagenome	1602165	brine
gill_metagenome	1455666	trench
glacier_metagenome	1651087	mangrove
hydrothermal_vent_metagenome	652676	pelagic
hydrozoan_metagenome	1941281	Atlantic
invertebrate_gut_metagenome	1775950	Antarctic
invertebrate_metagenome	1711999	Arctic
jellyfish_metagenome	1549733	Pacific
lagoon_metagenome	1763544	phytoplankton
macroalgae_metagenome	2015907	Mariana

mangrove_metagenome	1284368
marine_metagenome	408172
marine_plankton_metagenome	1874687
marine_sediment_metagenome	412755
microbial_mat_metagenome	527640
mollusc_metagenome	1417798
oyster_metagenome	1541066
periphyton_metagenome	1825055
sand_metagenome	1671699
sea_anemone_metagenome	1825923
sea_squirt_metagenome	1041057
sea_urchin_metagenome	1873886
seagrass_metagenome	1904484
seawater_metagenome	1561972
sediment_metagenome	749907
shrimp_gut_metagenome	1588881
sponge_metagenome	1163772
starfish_metagenome	2053188
surface_metagenome	1774230
tidal_flat_metagenome	1269027
whale_fall_metagenome	412756
zebrafish_metagenome	1331678

本章作者：

陈建威、贾洋洋、孙颖、周昌浩、郭扬、刘桂林、李登辉、徐大有、刘宇婧、范洁、赵芳、王军、孙佳惠（青岛华大基因研究院）

本章参考文献

1. Munn, C.B., *Marine Microbiology: Ecology and Applications*. Boca Raton: CRC Press, 2019.
2. Overmann, J. and C. Lepleux, *Marine Bacteria and Archaea: Diversity, Adaptations, and Culturability*. 2016: p. 21-55.
3. Bobay, L.M. and H. Ochman, *The evolution of bacterial genome architecture*. Front Genet, 2017. **8**: p. 72.
4. Allison, S.D. and J.B.H. Martiny, *Resistance, resilience, and redundancy in microbial communities*. Proceedings of the National Academy of Sciences, 2008. **105**(supplement_1): p. 11512-11519.
5. Rusch, D.B., et al., *The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific*. PLOS Biology, 2007. **5**(3): p. e77.
6. Sunagawa, S., et al., *Structure and function of the global ocean microbiome*. Science, 2015. **348**(6237): p. 1261359.
7. Paoli, L., et al., *Biosynthetic potential of the global ocean microbiome*. Nature, 2022. **607**(7917): p. 111-118.
8. Nishimura, Y. and S. Yoshizawa, *The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments*. Sci Data, 2022. **9**(1): p. 305.
9. Parks, D.H., et al., *A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life*. Nat Biotechnol, 2018. **36**(10): p. 996-1004.
10. Sauer, D.B. and D.N. Wang, *Predicting the optimal growth temperatures of prokaryotes using only genome derived features*. Bioinformatics, 2019. **35**(18): p. 3224-3231.
11. Ito, M., M. Morino, and T.A. Krulwich, *Mrp Antiporters Have Important Roles in Diverse Bacteria and Archaea*. Front Microbiol, 2017. **8**: p. 2325.
12. Boeuf, D., et al., *Metapangenomics reveals depth-dependent shifts in metabolic potential for the ubiquitous marine bacterial SAR324 lineage*. Microbiome, 2021. **9**(1): p. 172.
13. McInnes, L., J. Healy, and J. Melville, *Umap: Uniform manifold approximation and projection for dimension reduction*. arXiv preprint arXiv:1802.03426, 2018.
14. Delmont, T.O., et al., *Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes*. Nature Microbiology, 2018. **3**(7): p. 804-813.
15. Zeng, X., K. Alain, and Z. Shao, *Microorganisms from deep-sea hydrothermal vents*. Marine Life Science & Technology, 2021. **3**(2): p. 204-230.
16. Cullen, J.J., *Subsurface chlorophyll maximum layers: enduring enigma or mystery solved?* Ann Rev Mar Sci, 2015. **7**: p. 207-39.

17. Richter, D.J., et al., *Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems*. Elife, 2022. **11**.
18. Bentkowski, P., C. Van Oosterhout, and T. Mock, *A model of genome size evolution for prokaryotes in stable and fluctuating environments*. Genome Biol Evol, 2015. **7**(8): p. 2344-51.
19. Konstantinidis, K.T. and J.M. Tiedje, *Trends between gene content and genome size in prokaryotic species with larger genomes*. Proceedings of the National Academy of Sciences, 2004. **101**(9): p. 3160-3165.
20. Cobo-Simon, M. and J. Tamames, *Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa*. BMC Genomics, 2017. **18**(1): p. 499.
21. Rodríguez-Gijón, A., et al., *A genomic perspective across Earth's microbiomes reveals that genome size in archaea and bacteria is linked to ecosystem type and trophic strategy*. Frontiers in Microbiology, 2022. **12**.
22. Orakov, A., et al., *GUNC: detection of chimerism and contamination in prokaryotic genomes*. Genome Biol, 2021. **22**(1): p. 178.
23. Mara, P., et al., *Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline*. ISME J, 2020. **14**(12): p. 3079-3092.
24. Cabello-Yeves, P.J., et al., *The microbiome of the Black Sea water column analyzed by shotgun and genome centric metagenomics*. Environ Microbiome, 2021. **16**(1): p. 5.
25. Mende, D.R., et al., *Environmental drivers of a microbial genomic transition zone in the ocean's interior*. Nat Microbiol, 2017. **2**(10): p. 1367-1373.
26. Guo, M., et al., *Increased WD40 motifs in Planctomycete bacteria and their evolutionary relevance*. Mol Phylogenet Evol, 2021. **155**: p. 107018.
27. Hu, X.J., et al., *Prokaryotic and Highly-Repetitive WD40 Proteins: A Systematic Study*. Sci Rep, 2017. **7**(1): p. 10585.
28. Neer, E.J., et al., *The ancient regulatory-protein family of WD-repeat proteins*. Nature, 1994. **371**(6495): p. 297-300.
29. Mahajan, M., et al., *Paralogization and new protein architectures in Planctomycetes bacteria with complex cell structures*. Mol Biol Evol, 2020. **37**(4): p. 1020-1040.
30. Fuerst, J.A. and E. Sagulenko, *Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function*. Nature Reviews Microbiology, 2011. **9**(6): p. 403-413.
31. Giovannoni, S.J., J. Cameron Thrash, and B. Temperton, *Implications of streamlining theory for microbial ecology*. ISME J, 2014. **8**(8): p. 1553-65.
32. Swan, B.K., et al., *Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean*. Proc Natl Acad Sci U S A, 2013. **110**(28): p. 11463-8.
33. Musto, H., et al., *Genomic GC level, optimal growth temperature, and genome size in prokaryotes*. Biochemical and Biophysical Research Communications, 2006. **347**(1): p. 1-3.

34. Almpanis, A., et al., *Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages*. Microb Genom, 2018. **4**(4).
35. Vinogradov, A.E., *Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship*. Cytometry, 1998. **31**(2): p. 100-109.
36. Šmarda, P., et al., *Genome size and GC content evolution of festuca: ancestral expansion and subsequent reduction*. Annals of Botany, 2007. **101**(3): p. 421-433.
37. Watson, B.N.J., R.H.J. Staals, and P.C. Fineran, *CRISPR-Cas-Mediated Phage Resistance Enhances Horizontal Gene Transfer by Transduction*. mBio, 2018. **9**(1).
38. Munita, J.M., et al., *Mechanisms of Antibiotic Resistance*. Microbiology Spectrum, 2016. **4**(2).
39. Meaden, S., et al., *High viral abundance and low diversity are associated with increased CRISPR-Cas prevalence across microbial ecosystems*. Curr Biol, 2022. **32**(1): p. 220-227 e5.
40. Burstein, D., et al., *Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems*. Nat Commun, 2016. **7**: p. 10613.
41. Nayfach, S., et al., *A genomic catalog of Earth's microbiomes*. Nat Biotechnol, 2020.
42. Liu, Y., et al., *A genome and gene catalog of glacier microbiomes*. Nat Biotechnol, 2022. **40**(9): p. 1341-1348.
43. Gophna, U., et al., *No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales*. ISME J, 2015. **9**(9): p. 2021-7.
44. Weissman, J.L., et al., *Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy*. ISME J, 2019. **13**(10): p. 2589-2602.
45. Wheatley, R.M. and R.C. MacLean, *CRISPR-Cas systems restrict horizontal gene transfer in Pseudomonas aeruginosa*. ISME J, 2021. **15**(5): p. 1420-1433.
46. Shehreen, S., et al., *Genome-wide correlation analysis suggests different roles of CRISPR-Cas systems in the acquisition of antibiotic resistance genes in diverse species*. Philos Trans R Soc Lond B Biol Sci, 2019. **374**(1772): p. 20180384.
47. Pursey, E., et al., *CRISPR-Cas is associated with fewer antibiotic resistance genes in bacterial pathogens*. Philos Trans R Soc Lond B Biol Sci, 2022. **377**(1842): p. 20200464.
48. Gavriilidou, A., et al., *Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes*. Nat Microbiol, 2022. **7**(5): p. 726-735.
49. Saati-Santamaria, Z., et al., *Unveiling the genomic potential of Pseudomonas type strains for discovering new natural products*. Microb Genom, 2022. **8**(2).
50. Belknap, K.C., et al., *Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria*. Sci Rep, 2020. **10**(1): p. 2003.
51. Yan, S., et al., *Micromonospora: A Prolific Source of Bioactive Secondary Metabolites with Therapeutic Potential*. J Med Chem, 2022. **65**(13): p. 8735-8771.

52. Ayikpoe, R.S., et al., *A scalable platform to discover antimicrobials of ribosomal origin*. Nat Commun, 2022. **13**(1): p. 6135.
53. Wei, B., et al., *Global analysis of the biosynthetic chemical space of marine prokaryotes*. Microbiome, 2023. **11**(1): p. 144.
54. Salazar, G., et al., *Gene expression changes and community turnover differentially shape the global ocean metatranscriptome*. Cell, 2019. **179**(5): p. 1068-1083 e21.
55. Yoshida, S., et al., *A bacterium that degrades and assimilates poly(ethylene terephthalate)*. Science, 2016. **351**(6278): p. 1196-1199.
56. Kawai, F., T. Kawabata, and M. Oda, *Current State and Perspectives Related to the Polyethylene Terephthalate Hydrolases Available for Biorecycling*. ACS Sustainable Chemistry & Engineering, 2020. **8**(24): p. 8894-8908.
57. DeFrancesco, L., *Closing the recycling circle*. Nat Biotechnol, 2020. **38**(6): p. 665-668.
58. Zhu, B., D. Wang, and N. Wei, *Enzyme discovery and engineering for sustainable plastic recycling*. Trends Biotechnol, 2022. **40**(1): p. 22-37.
59. Lu, H., et al., *Machine learning-aided engineering of hydrolases for PET depolymerization*. Nature, 2022. **604**(7907): p. 662-667.
60. Tournier, V., et al., *An engineered PET depolymerase to break down and recycle plastic bottles*. Nature, 2020. **580**(7802): p. 216-219.
61. Buchholz, P.C.F., et al., *Plastics degradation by hydrolytic enzymes: The plastics-active enzymes database-PAZy*. Proteins, 2022. **90**(7): p. 1443-1456.
62. Danso, D., et al., *New Insights into the Function and Global Distribution of Polyethylene Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes*. Applied and Environmental Microbiology, 2018. **84**(8).
63. Joo, S., et al., *Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation*. Nat Commun, 2018. **9**(1): p. 382.
64. Jin, M., et al., *Properties and Applications of Extremozymes from Deep-Sea Extremophilic Microorganisms: A Mini Review*. Mar Drugs, 2019. **17**(12).
65. Shi, L., et al., *Complete Depolymerization of PET Wastes by an Evolved PET Hydrolase from Directed Evolution*. Angew Chem Int Ed Engl, 2023. **62**(14): p. e202218390.

An aerial photograph showing a small boat on a vast, bright green algal bloom in the ocean. The water is a deep teal color, and the algal bloom is a vibrant, almost neon green. The boat is positioned in the lower right quadrant of the image, moving towards the upper left. The algal bloom is concentrated in a large, irregular shape, with smaller patches scattered around it. The boat has a white cabin and a green net or canopy over the deck. The overall scene suggests a marine research or monitoring activity.

400+

藻类基因组

覆盖预估藻类物种的

0.6%

第三章 藻类

3.1 已公布的藻类基因组特征

藻类是有别于陆地植物的、能进行光合作用的真核生物，多为水生，在分类、形态和遗传特征方面具有巨大的多样性，能在冷、热、酸等多种环境条件下生存，因此在世界各地分布广泛。藻类为多系起源，初级内共生事件后演化出灰藻、绿藻和红藻，绿藻和红藻在多级内共生之后演化出更多不同藻类。其中，绿藻分支主要包含绿藻门和轮藻门（链形植物门），属于绿色植物类群，是陆地高等植物的祖先；红藻分支经多次内共生后演化出硅藻、褐藻、定鞭藻等类群。目前在 AlgaeBase 数据库（<https://www.algaebase.org>）中收录藻类已超过三万种，其中绿藻分支类群的物种一万多种，红藻分支类群两万多种。藻类作为地球最古老、最重要的生物之一，一方面它们是海洋生态系统的主要生产力，是其他海洋动物的食物，作为初级生产者具有不可替代的作用，是生态学、生物演化等学科的重点研究对象；另一方面，藻类可作为人类食物，也是重要的工业、医学原料，具有很高的经济和应用价值。

鉴于藻类重要的生态地位和价值，藻类基因组和遗传信息的解读对科学研究和产业发展越来越重要。藻类组学研究也随着日新月异的测序技术迅速推进。截至 2023 年，NCBI 收录的已发表藻类基因组超过 400 个，涵盖大部分藻类门类（图 3.1a、d）。具体来看，绿藻（Chlorophyta）179 种，褐藻（Ochrophyta）90 种，硅藻（Bacillariophyta）74 种，红藻（Rhodophyta）28 种，囊泡藻（Alveolata，甲藻）26 种，链形植物门藻类（Streptophyta，轮藻）13 种，灰藻（Glaucocystophyceae）基因组 1 种。在已公布的藻类基因组中，绿藻数量最多，占有藻类的 41%；其次是褐藻、硅藻、红藻和甲藻，上述藻类为主要测序物种，占 91% 以上。从测序技术发展和应用上看，2017 年以三代（Third-generation sequencing, TGS）和二代（Next-generation sequencing, NGS）测序技术相结合的基因组组装策略开始盛行（图 3.1b），2019 年藻类基因组公布的数量达到一个顶峰，是 2018 年的三倍以上（图 3.1c），可以说三代测序技术的发展极大加速了藻类基因组研究进程。

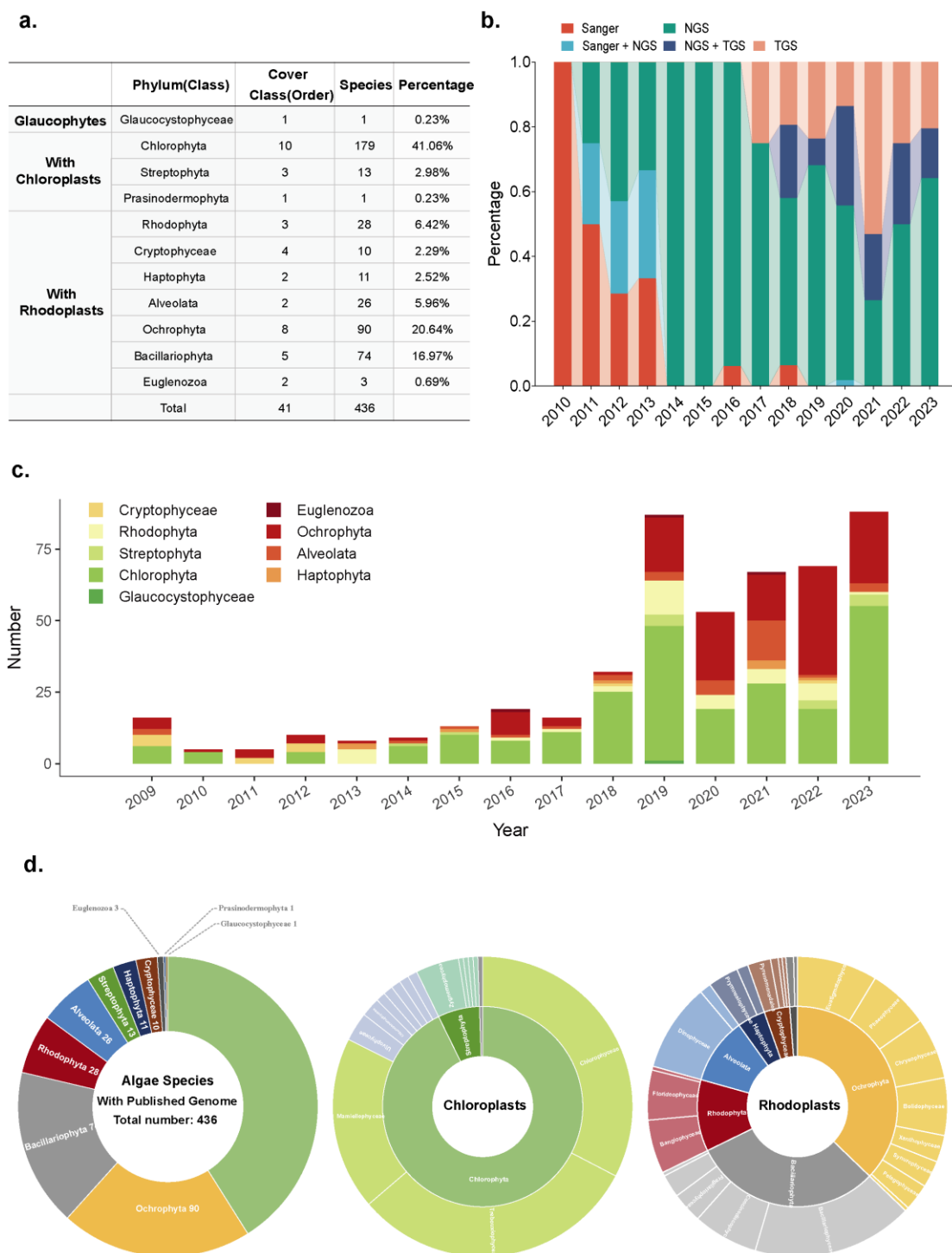


图 3.1. NCBI 已发表藻类基因组数量统计

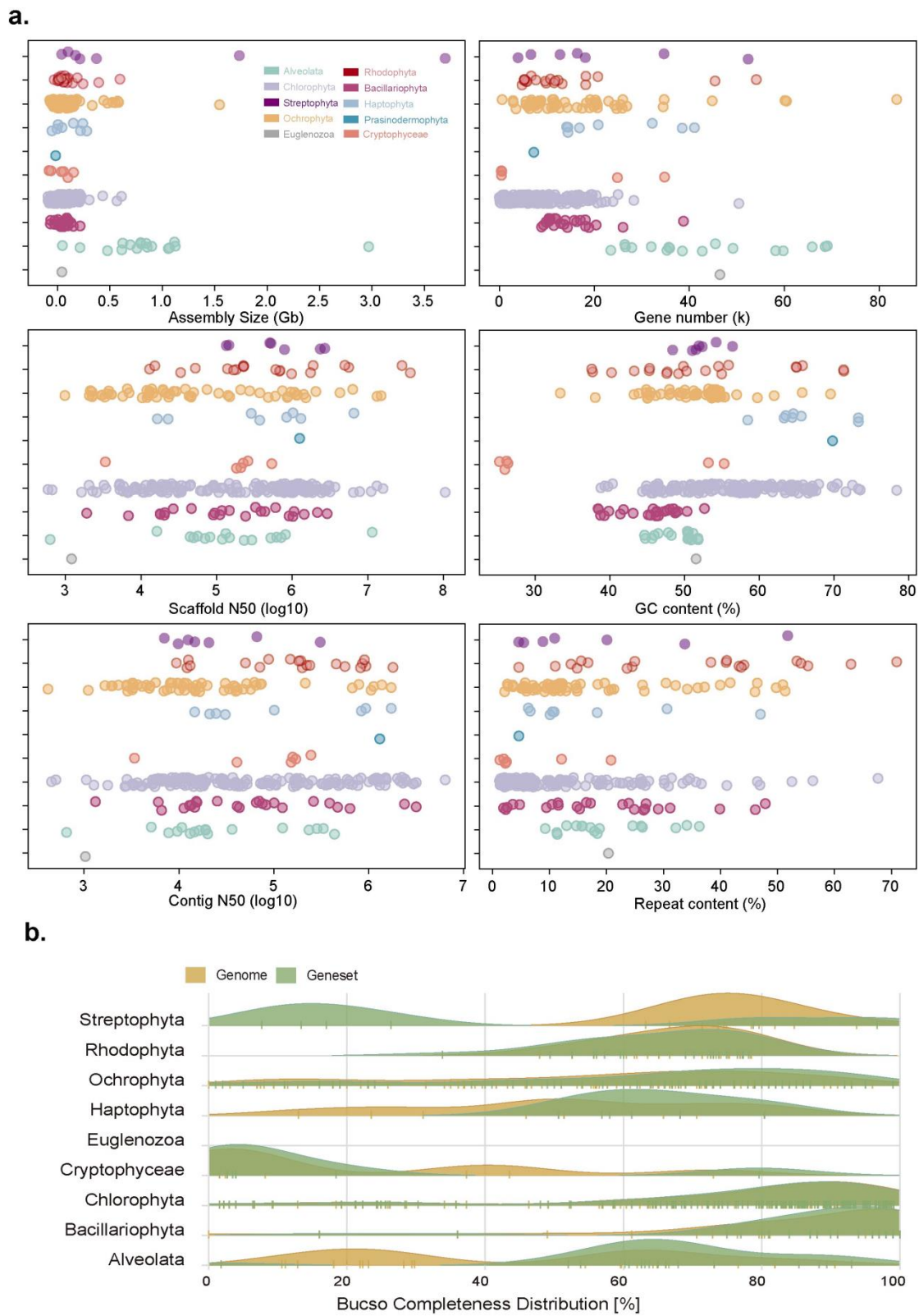


图 3.2. 藻类基因组组装质量和特征分布

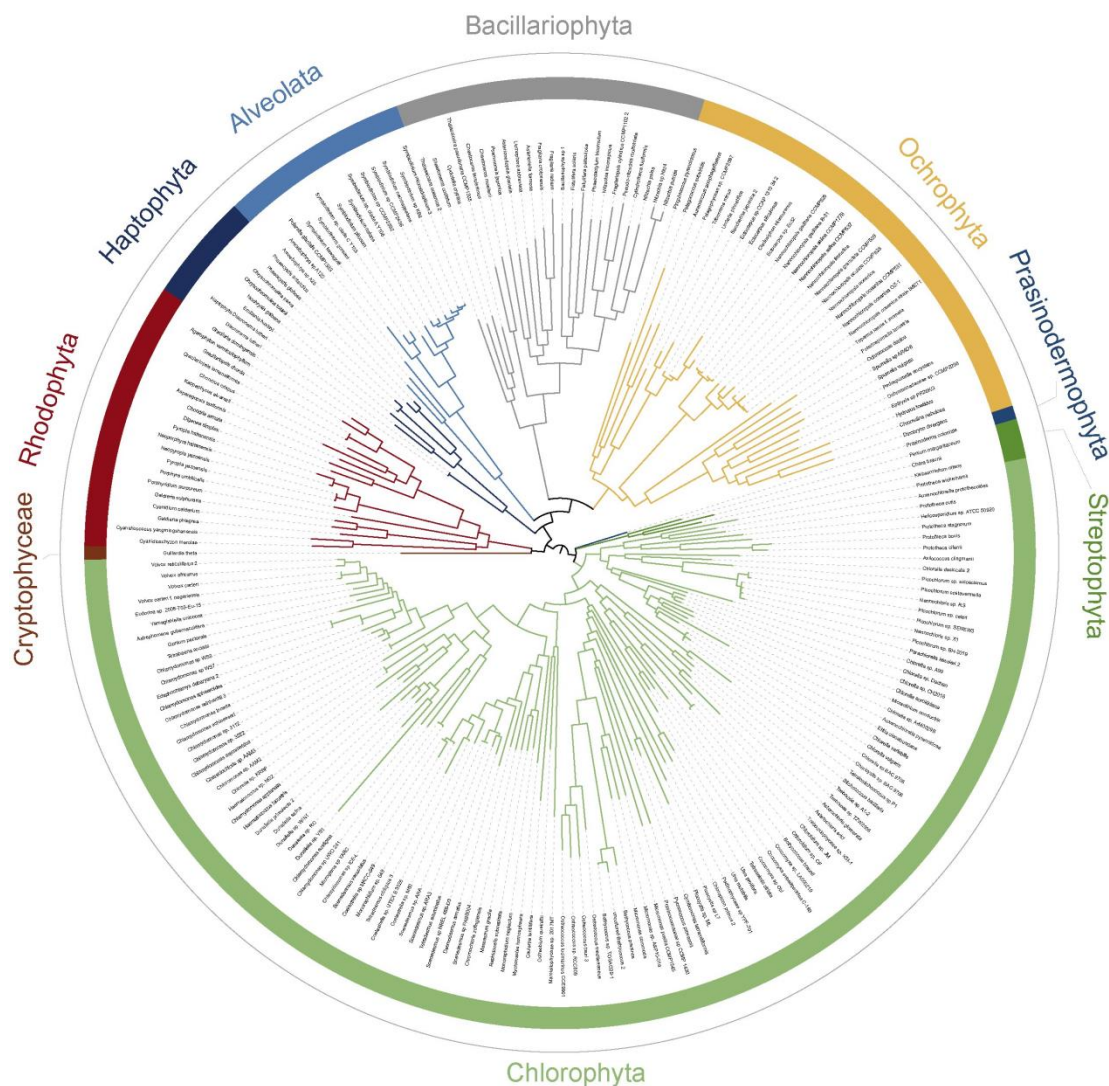


图 3.3. 藻类基因组系统发育树

通过比较这些已经公布的藻类基因组，我们发现不同藻类的基因组特征和组装质量各不相同（图 3.2a）。大多数已测序的藻类（例如绿藻门和红藻门）的基因组约 100Mb，轮藻和甲藻的基因组大于 400Mb，最大可超过 3Gb。使用 BUSCO 数据库评估基因组的完整性，发现大部分藻类基因组的真核保守基因比例在 50%~80% 范围（图 3.2b）。另外，藻类基因组组装的连续性和 GC 含量也差异较大。整体来看，藻类基因组的 Scaffold N50 和 Contig N50 在 10Kb~8Mb 范围均有分布。藻类基因组的 GC 含量整体在 40%~60% 范围。重复序列（Transposable elements, TE）是基因组的重要组成部分。鉴定到的藻类重复序列比例比陆生植物（平均 ~50%）

明显偏低，整体范围为 5% ~ 40%，其中，LINE、LTR、DNA 等类型 TE 为主要重复序列类型（图 3.2a）。此外，大部分藻类编码基因数量都在 20,000 以下，根据藻类物种中保守基因的构建系统发育树能观察到两个明显分开的藻类分支，其中一种属于红藻分支，其姐妹分支关系包括红藻、定鞭藻、甲藻、硅藻、褐藻等；另一种属于植物谱系，主要包括绿藻和轮藻，这也与藻类的演化假说基本一致（图 3.3）。

3.2 藻类基因组水平基因转移

水平基因转移（Horizontal gene transfer, HGT）是藻类基因组进化的驱动力之一。大规模基因转移（内共生）和小规模基因转移（单基因转移或片段转移）是藻类物种基因组形成丰富的遗传多样性的方式之一[1]。在红藻基因组中发现来自细菌的 HGT 基因可以帮助适应极端环境[2]。藻类基因组的增加使多物种基因组水平上的 HGT 比较研究成为可能[3]。由于藻类可能发生一种或多种内共生事件，系统性分析 HGT 转移的模式和功能多样性是一个挑战。从理论上讲，这些基因转移能明显影响它们的基因组，导致数百甚至数千个基因的转移[4]。基于上百种藻类基因组数据，我们系统检测了 169 个高质量藻类基因组中不同方向的 HGT 事件，鉴定到 6,418 个外群转移到藻类、8,874 个藻类转移到外群以及 31,923 个未知方向的 HGT 基因。平均每个物种 HGT 基因 10~100 个，占有所有基因的 0.5%~5%（图 3.4）。

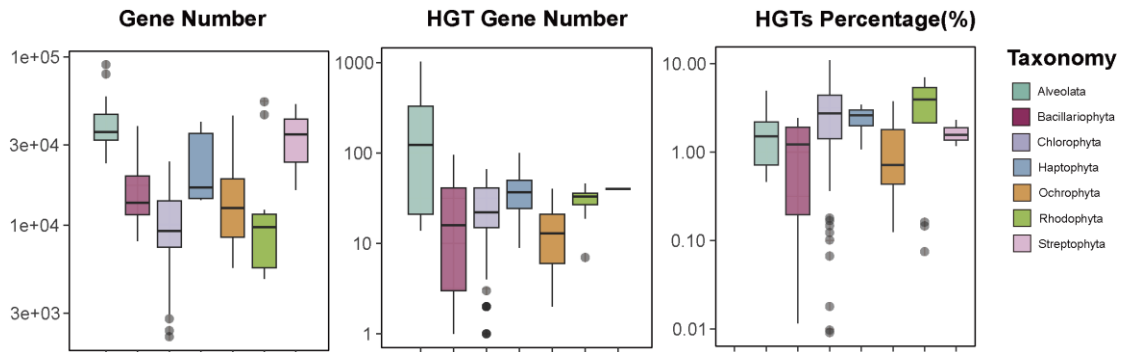


图 3.4. 藻类基因组 HGT 数量分布

外源基因通过不同供体转移进入藻类基因组，与祖先中发生的 332 个古老 HGT 事件相关，这在藻类物种分化过程中的遗传信息演变扮演着重要角色。同时藻类作为古老生物，其基因也可能通过宿主关系转移到其他高等生物进而影响宿主生物基

因组的改变。将所有鉴定到的藻类 HGT 判断转移方向和来源，发现转移到藻类的 HGT 来源多为原核生物（如变形菌门 Proteobacteria、浮霉菌门 Planctomycetes），从藻类转移到外群的受体主要为脊索动物（Chordata）（图 3.5）。在转到藻类的 HGT 中，变形菌门在藻类谱系中广泛分布，蓝细菌门则集中分布在红藻、褐藻、定鞭藻，其中绿藻门也有零星分布（图 3.6）。总的来看，红藻谱系的物种发生的 HGT 比绿藻谱系的更丰富。

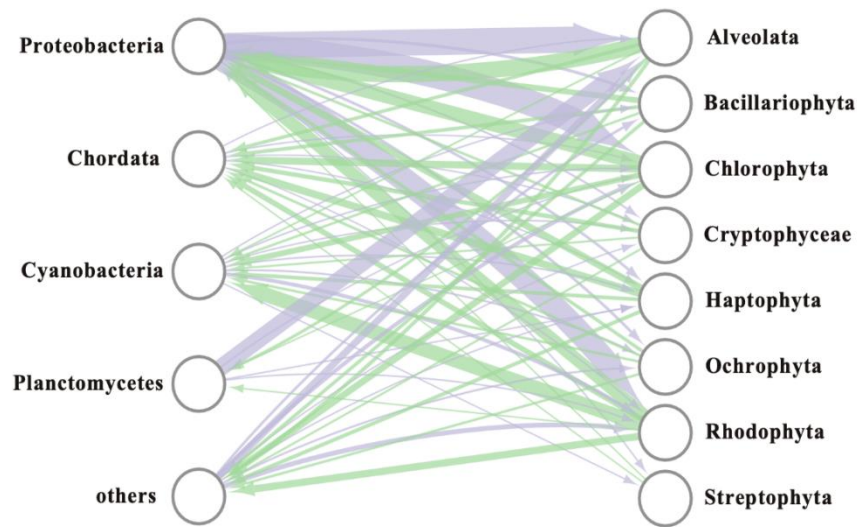


图 3.5. 藻类基因组 HGT 的转移方向分布

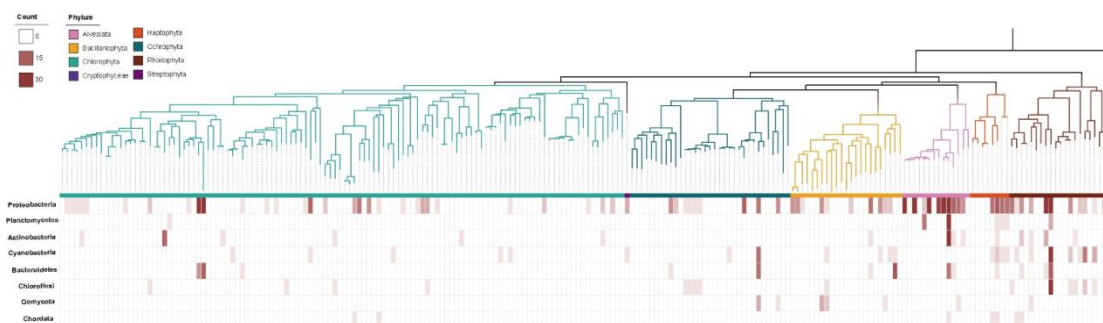


图 3.6. 藻类获得的 HGT 来源分布

对外群转移到藻类中的 HGT 基因进行聚类来寻找共同 HGT 事件。所有的 HGT 基因聚类得到超过一千个基因家族（OrthoGroups, OG）。我们定义：包含了

每个门类藻类至少半数以上纲级别种类的基因家族为祖先 HGT。在这一标准下，发现红藻绿藻谱系共同发生的 HGT 事件 4 个，红藻谱系的 HGT 事件 12 个，甲藻、硅藻与褐藻的 HGT 事件 7 个（图 3.7）。统计每个 HGT 事件的基因功能分布，发现最古老的 HGT 事件（即绿藻谱系和红藻谱系共同的 HGT），其基因功能主要分布在光合过程、蛋白转运等，这也与藻类初级内共生相验证。而红藻谱系的共同 HGT 基因功能主要参与离子代谢、信号转导，以及氮代谢调控。发生在囊泡藻、硅藻、褐藻所在分支中的共同 HGT 事件的基因的功能主要与蛋白质代谢、极端环境适应、胁迫响应以及核苷酸代谢、硫代硫酸盐合成相关。广泛的功能分布表明，红藻谱系在初级内共生事件之后进行了多次不同谱系的 HGT 事件，这也是藻类物种多样性进化的驱动力之一。

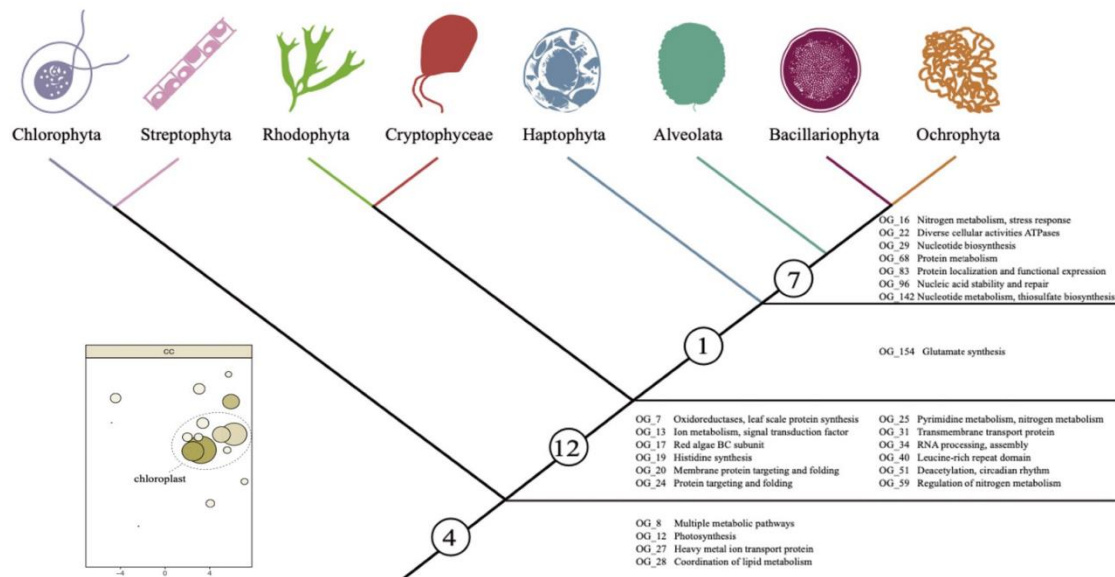


图 3.7. 祖先 HGT 事件在藻类系统分支节点的分布

3.3 藻类基因组研究热点

绿藻作为藻类中数量最大的类群，其组学研究最为广泛深入，拥有数量最多的基因组测序和科研产出。绿色植物的陆地化是一个关键的进化事件，改变了地球的生态系统。轮藻是陆地植物（有胚植物）最近源的藻类。轮藻物种形态多样，包括单细胞和结构复杂的多细胞形式，涵盖六个主要谱系：轮藻纲（Charophyceae）、绿方藻纲（Chlorokybophyceae）、双星藻纲（Zygnematophyceae）、中眼藻纲

（*Mesostigmatophyceae*）、克里藻纲（*Klebsormidiophyceae*）和鞘毛藻纲（*Coleochaetophyceae*）[5]。系统发育研究表明，双星藻纲（*Zygnematophyceae*）是离陆地植物的最近的姊妹类群。轮藻在适应陆地的过程中获得了能够抵抗陆地生活条件（如紫外线、盐度和干燥等）的相关基因。提高陆地植物对生物和非生物胁迫抗性的转录因子（Transcription factor, TF）起源于双星藻纲和胚胎植物的共同祖先，包括 *GRAS*、*HDKNOX2*、*BBR/BPC*、*NAC*、*LOB*、*bZIP*、*bHLH*、*WRKY* 和 *ERF* 等大量扩张的基因家族[6, 7]。在已测序的几个轮藻门物种中预测 TF，可以看到 *GRAS* 家族主要在 *Mesotaenium endlicherianum* 和 *Spirogloea muscicola* 中存在（图 3.8）。使用 TF 基因对 16 种代表性绿藻、轮藻、拟南芥、水稻和银杏构建的系统发育树结果显示，轮藻和陆地植物中只存在 *GRAS* 和 *NAC* 基因，*GRAS* 可以聚集到一个完整的进化枝，而 *NAC* 基因显然可以聚集到三个进化枝，表明两者对应基因家族存在的不同进化模式。

Species/TF	GRAS	bZIP	bHLH	WRKY	BBR/BPC	LOB	NAC
<i>Mesotaenium endlicherianum</i>	8	2	3	3	0	0	4
<i>Spirogloea muscicola</i>	14	8	16	9	0	4	3
<i>Mesostigma viride</i>	0	3	3	1	1	0	1
<i>Klebsormidium nitens</i>	0	7	10	2	0	5	3
<i>Chlorokybus atmophyticus</i>	0	0	0	2	0	0	0
<i>Chara braunii</i>	0	1	11	1	4	2	1

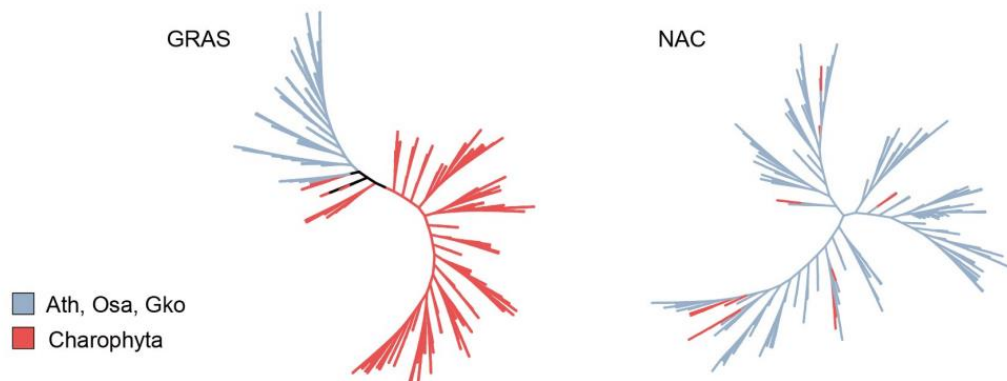


图 3.8. 轮藻基因组重要转录因子分布和 *GRAS/NAC* 系统发育树

红藻因含有藻红素使其呈现红色，作为最古老和最大的门类之一，包含了超过七千多个物种[8]。红藻分布广泛，主要在海洋中生长。红藻富含维生素和蛋白质，可以作为食物，具有很高的营养价值和经济价值，在生物技术与产业具有广泛应用。已发表的红藻基因组大小从 13Mb 到 400 Mb 不等（图 3.1）。红藻在早期进化过程中经历了大规模的基因组缩减，也没有多倍化事件[9]，因此红藻虽作为多细胞生物，其基因组大小及基因数量较其他典型多细胞真核生物低，基因组也表现出紧凑的特征，如内含子少且短、基因之间的中值距离低、基因家族拷贝数低等[10]。但某些特定基因家族的数量变化仍可以在一定程度上表征红藻基因组的进化特征。例如，碳获取途径对光合生物碳浓缩机制至关重要，细胞内碳酸酐酶（Carbonic anhydrase, CA）可逆地催化二氧化碳转化为碳酸氢盐，而红毛菜纲物种中 CA 的数量大大超过真红藻纲[11]。还有一些研究发现，大多数红藻缺乏肌球蛋白，因此有更简单的胞质分裂系统[12, 13]。红藻的环境适应性也是其研究热点。基因组比较分析发现，在红藻尤其是紫菜中，与抗氧化功能相关的基因（如 SOD）拷贝数增加，由于 SOD 对改善活性氧引起的细胞损伤和维持细胞稳态具有重要作用，因此其扩增可能与潮间带环境适应相关[11]。温泉红藻能够在高温环境中生存，基因组比较分析显示，此类红藻通过水平基因转移、亚端粒基因复制和基因/基因组丢失等三种策略驱动基因组发生结构变异和进化，由此适应高温、酸性和富含重金属的环境[14]。

硅藻和褐藻在系统分类上属于不等鞭毛门（Heterokont），作为真核藻类的重要谱系，包含超过一万多个物种。2008 年，使用二代测序技术解析了第一个硅藻——三角褐指藻基因组[15]。在硅藻组学的主要研究点包括亚热带海岸、低铁和寒冷等环境适应性[16-18]，细胞内重要代谢途径机制[19, 20]，以及不同门类硅藻的基因组结构进化特征[15, 21]等方面。大多数硅藻都能进行光合作用，但也存在异养物种。非光合作用硅藻基因组 *Nitzschia* sp. strain Nitz4 研究全面地解析了代谢向异养的转变以及对光合作用损失的限制[22]。在拟菱形藻 *Pseudo-nitzschia multistriata* 的基因组中，通过比较鉴定基因的获得和丢失、水平基因转移以及性别决定相关基因的保存和进化研究了有性繁殖期的细胞应答[23]。褐藻包括多种海藻，

可作为食物和海洋动物的潜在的栖息地。在模式物种水云褐藻 *Ectocarpus siliculosus* 的基因组中鉴定了应对潮汐环境的光捕获、色素合成基因、卤化物代谢等过程相关基因，以及与受体激酶家族关联的多细胞进化与信号转导基因[24]。在岩藻 *Cladosiphon okamuranus* 基因组中鉴定了参与硫酸岩藻聚糖和藻酸盐生物合成途径的蛋白质[25]。在海带 *Saccharina japonica* 基因组发现了与细胞壁合成、发育和免疫应答相关的基因家族发生扩张[26]。这些为改良物种的经济性状提供了宝贵的资源。此外，部分褐藻对非生物胁迫的高度耐受性遗传基础也是研究的关注点。比如，*Ectocarpus subulatus* 基因组中发现热休克蛋白家族的扩增和细胞壁多糖修饰酶的存在，这可能与胁迫耐受性相关[27]。

在所有的藻类中，甲藻是比较特殊的一种。一般认为，甲藻具有最复杂的演化历史，它与绿藻/红藻发生次级内共生，并与其他藻类发生三级内共生事件。在系统发育关系中出现在红藻分支，但也可能与绿藻门物种混杂。甲藻的基因组较其他藻类更大，最小的 1Gb，最大的可达 200Gb[28]。目前已测序的甲藻物种基因组大小一般在 5G 以下。考虑到生态位的重要性，关注度最高的甲藻是珊瑚生态系统中的共生藻属（虫黄藻）。由于它们特殊的生存方式，与宿主和谐共生背后的分子机制成为备受瞩目的科学问题。通过基因组测序，研究人员在虫黄藻基因组中观察到了与基因结构相关的特征（例如串联基因簇和 TTTT 盒的转录机制）、共生藻物种功能生物学的具体遗传特征（例如参与碳氮代谢的跨膜转运蛋白基因家族扩展等），以及珊瑚和共生藻细胞之间交换或补偿的共生相关基因[29]。此外，甲藻巨大的基因组（>30Gb）是另一个有趣的关注点。在物种形成的进化史上，这一特征必然与它们的生理生命活动密切相关。在大部分植物中，高比例的重复序列尤其反转录转座子 LTR 是大基因组的显著特征。因此，我们也怀疑甲藻基因组中存在大量重复序列。已发表的基因组较小的甲藻，其重复序列低于 30%[30]，而在基因组稍大（>3Gb）的冰川极地甲藻 *Polarella glacialis* 中重复序列比例高达 68%，其中简单重复占 37%[31]。重复序列影响基因组结构的进化和编码基因的结构和表达，物种主要重复序列类型的不同可能预示着不同的基因组进化模式。因此，重复序列元件

的比较分析，包括数量和类型的差异、爆发时间及其调控作用，也是今后甲藻基因组学研究中必不可少的内容。

3.4 讨论与展望

藻类通常与细菌共生，因此在藻类基因组测序中，污染问题是高质量组装的阻碍之一。目前常规的方法是将共生体的所有序列与 NR 库比对，通过鉴定序列所属物种类型将藻类与污染分离[10, 11, 13, 32, 33]。如果能够在测序前通过实验降低污染，或者针对性开发组装算法，可以一定程度提升组装准确度，同时后续分析的假阳性结果将大大减少。目前有很多软件，例如 EukRep[34]、Kraken2[35]和 k-means[36]，根据真核生物和原核生物的序列特征（GC 含量、K-mer 频率等）的差异区分藻类和污染。另外，我们可以把“如何分离污染的藻类基因组”的问题转变为“这些序列属于什么物种，藻类或者细菌”，即将藻类共生体当作宏基因组，通过分箱策略或 16S/18S 鉴定，来同时获得同一栖息地中存在的藻类和多种细菌的物种信息，在此基础上展开藻类共生生物的研究，探索藻类与共生细菌之间遗传交换的全貌。

值得一提的是，转录组学通常与比较基因组学一起用来探索适应性进化过程中基因组残留的证据进而挖掘其潜在的调控机制。例如，不同盐度实验条件下的绿藻 *Picochlorum renovo* 的比较转录组学分析证实了已报道的参与高盐响应的脯氨酸代谢基因[37]。结合 microRNA 系统的转录调控网络分析阐明了虫黄藻 *F. kawagutii* 和珊瑚基因组之间的生化互补性，并提出了共生和物质运输的机制模型[38]。近年来，单细胞转录组技术被广泛应用于细胞动态发育的探索中。在软珊瑚 *Xenia* 内共生细胞谱系的研究中，通过对不同细胞阶段的共生和非共生细胞进行单细胞测序，构建了珊瑚和藻类的内共生动态演化系统[39]。这为藻类的共生以及藻类与有害生态现象藻华等领域研究提供了新的思路和技术参考。因此，集成基因组学、转录组、单细胞的多组学分析的多平台综合应用将是藻类基因组学研究趋势。

本章作者：

施成成、贾利敏、韩凯、彭玲、刘小川、李良伟、葛奇金（青岛华大基因研究院）

本章参考文献

1. Suzuki, K., K. Moriguchi, and S. Yamamoto, Horizontal DNA transfer from bacteria to eukaryotes and a lesson from experimental transfers. *Research in Microbiology*, 2015. **166**(10): p. 753-763.
2. Chen, R., et al., Adaptive innovation of green plants by horizontal gene transfer. *Biotechnology Advances*, 2021. **46**: p. 107671.
3. Fan, X., et al., Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions. *Science advances*, 2020. **6**(18): p. eaba0111.
4. Bux, F. and Y. Chisti, *Algae biotechnology: products and processes*. 2016: Springer.
5. McCourt, R.M., C.F. Delwiche, and K.G. Karol, Charophyte algae and land plant origins. *Trends Ecol Evol*, 2004. **19**(12): p. 661-6.
6. Cheng, S., et al., Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. *Cell*, 2019. **179**(5): p. 1057-1067 e14.
7. Jiao, C., et al., The *Penium margaritaceum* Genome: Hallmarks of the Origins of Land Plants. *Cell*, 2020. **181**(5): p. 1097-1111 e12.
8. Schoch, C.L., et al., NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, 2020. **2020**.
9. Lee, J., et al., Analysis of the draft genome of the red seaweed *Gracilariopsis chorda* provides insights into genome size evolution in Rhodophyta. *Molecular biology and evolution*, 2018. **35**(8): p. 1869-1886.
10. Collén, J., et al., Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences*, 2013. **110**(13): p. 5247-5252.
11. Wang, D., et al., *Pyropia yezoensis* genome reveals diverse mechanisms of carbon acquisition in the intertidal environment. *Nature Communications*, 2020. **11**(1): p. 1-11.
12. Matsuzaki, M., et al., Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, 2004. **428**(6983): p. 653-657.
13. Brawley, S.H., et al., Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proceedings of the National Academy of Sciences*, 2017. **114**(31): p. E6361-E6370.

14. Cho, C.H., et al., Genome-wide signatures of adaptation to extreme environments in red algae. *Nature Communications*, 2023. **14**(1): p. 10.
15. Bowler, C., et al., The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, 2008. **456**(7219): p. 239-44.
16. Mock, T., et al., Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature*, 2017. **541**(7638): p. 536-540.
17. Nelson, D.R., et al., Potential for Heightened Sulfur-Metabolic Capacity in Coastal Subtropical Microalgae. *iScience*, 2019. **11**: p. 450-465.
18. Lommer, M., et al., Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol*, 2012. **13**(7): p. R66.
19. Traller, J.C., et al., Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnol Biofuels*, 2016. **9**: p. 258.
20. Tanaka, T., et al., Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell*, 2015. **27**(1): p. 162-76.
21. Armbrust, E.V., et al., The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*, 2004. **306**(5693): p. 79-86.
22. Pendergrass, A., et al., The genome of the nonphotosynthetic diatom, *Nitzschia* sp.: Insights into the metabolic shift to heterotrophy and the rarity of loss of photosynthesis in diatoms. 2020.
23. Basu, S., et al., Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol*, 2017. **215**(1): p. 140-156.
24. Cock, J.M., et al., The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature*, 2010. **465**(7298): p. 617-21.
25. Nishitsuji, K., et al., A draft genome of the brown alga, *Cladosiphon okamuranus*, S-strain: a platform for future studies of 'mozuku' biology. *DNA Res*, 2016. **23**(6): p. 561-570.
26. Ye, N., et al., Saccharina genomes provide novel insight into kelp biology. *Nat Commun*, 2015. **6**: p. 6986.
27. Dittami, S.M., et al., The genome of *Ectocarpus subulatus* - A highly stress-tolerant brown alga. *Mar Genomics*, 2020. **52**: p. 100740.

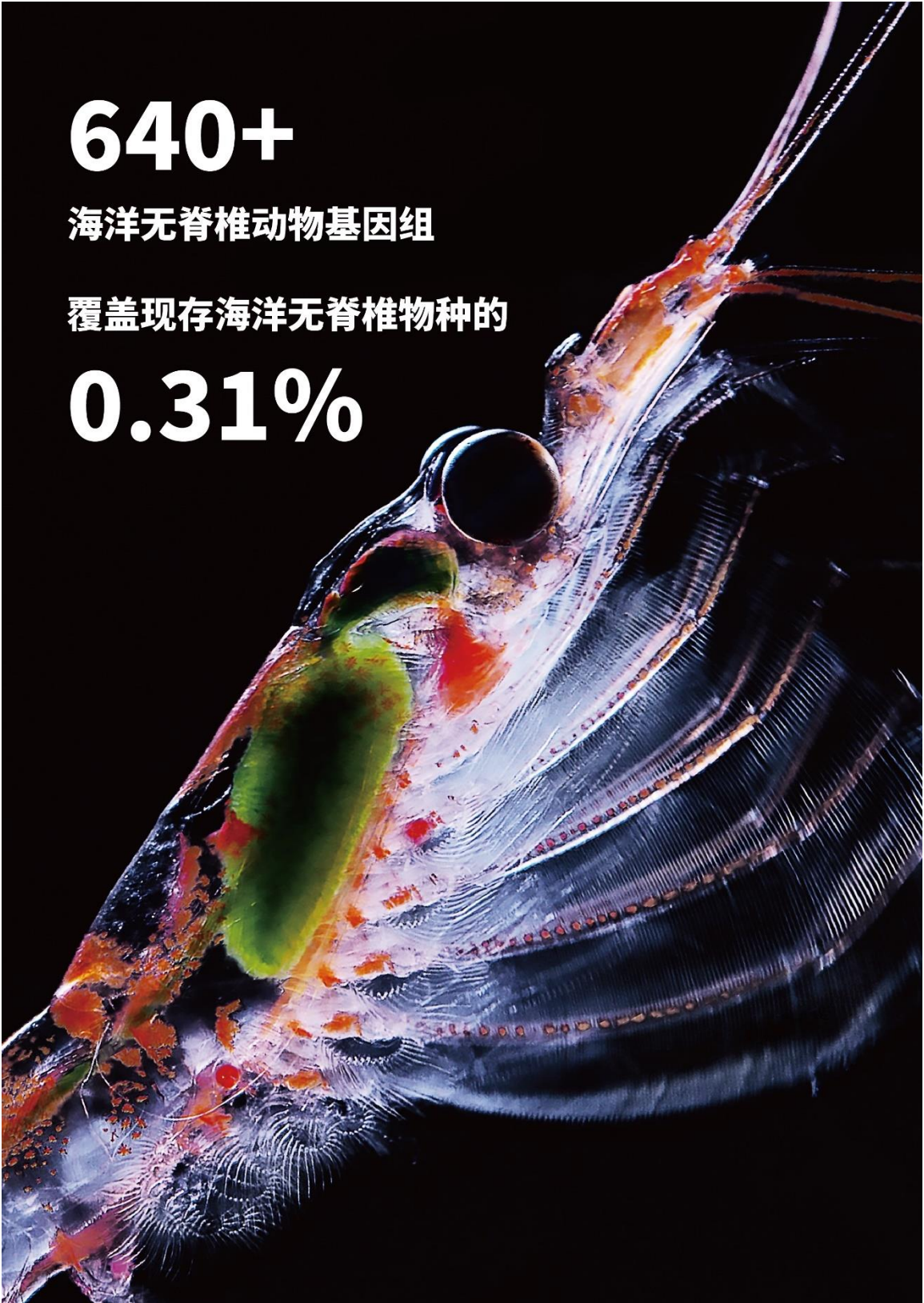
28. Lin, S., Genomic understanding of dinoflagellates. *Research in Microbiology*, 2011. **162**(6): p. 551-569.
29. Bi, Y., F. Wang, and W. Zhang, Omics Analysis for Dinoflagellates Biology Research. *Microorganisms*, 2019. **7**(9): p. 288.
30. Liu, H., et al., Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Commun Biol*, 2018. **1**: p. 95.
31. Stephens, T.G., et al., Genomes of the dinoflagellate *Polarella glacialis* encode tandemly repeated single-exon genes with adaptive functions. *BMC Biology*, 2020. **18**(1): p. 56.
32. Rossoni, A.W., et al., The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions. *Elife*, 2019. **8**: p. e45017.
33. Cao, M., et al., A chromosome-level genome assembly of *Pyropia haitanensis* (Bangiales, Rhodophyta). *Molecular ecology resources*, 2020. **20**(1): p. 216-227.
34. West, P.T., et al., Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome research*, 2018. **28**(4): p. 569-580.
35. Wood, D.E., J. Lu, and B. Langmead, Improved metagenomic analysis with Kraken 2. *Genome biology*, 2019. **20**(1): p. 257.
36. MacQueen, J. Some methods for classification and analysis of multivariate observations. in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. 1967. Oakland, CA, USA.
37. Dahlin, L.R., et al., Development of a high-productivity, halophilic, thermotolerant microalga *Picochlorum renovo*. *Commun Biol*, 2019. **2**: p. 388.
38. Lin, S., et al., The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*, 2015. **350**(6261): p. 691-4.
39. Hu, M., et al., Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*. *Nature*, 2020. **582**(7813): p. 534-538.

640+

海洋无脊椎动物基因组

覆盖现存海洋无脊椎物种的

0.31%



第四章 海洋无脊椎动物

无脊椎动物，因背侧没有脊柱而得名，包含广泛的类群和物种，其在地球上的多样性非常丰富，约占所有动物总数的 95% [1]，覆盖 32 个门，其中分布在海洋的物种占据 31 个门。海洋无脊椎动物目前有 207,974 个现存物种记录（数据来源于 2023 年 10 月 4 日的《世界海洋物种名录》（WORMS）） [2]，包括栉水母、海绵、水螅、珊瑚、水母、海星、海蛞蝓、虾、螃蟹、乌贼和牡蛎等，在生态系统中扮演着重要的角色（图 4.1）。其中十个主要的门：包括节肢动物门、软体动物门、脊索动物门、环节动物门、扁形动物门、刺胞动物门、多孔动物门、棘皮动物门、苔藓动物门和线虫动物门，约占据所有海洋无脊椎动物物种的 97.59%。

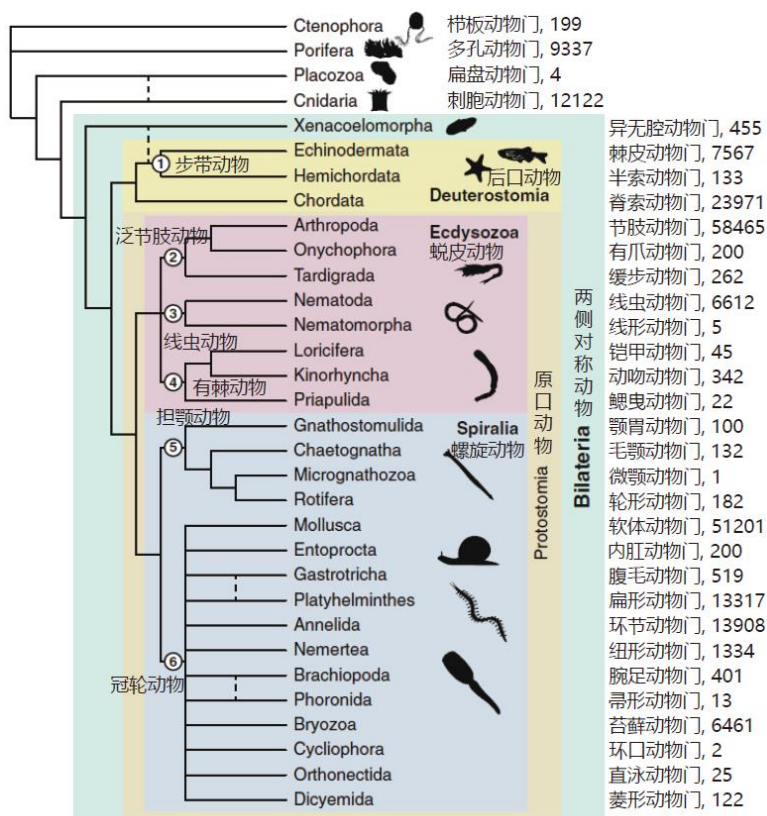


图 4.1. 现存的海洋无脊椎动物种数及在每个门中的分布，每个动物门后边数字代表现存记录物种数，图中系统发育树引自“Old Questions and Young Approaches to Animal Evolution”。

海洋无脊椎动物因其丰富性、多样性以及对生态系统和人类经济的重要影响，长期以来一直吸引着研究人员的浓厚兴趣，其多样性和丰富性使其成为生态学、生物学和环境科学等领域的研究重点。了解海洋无脊椎动物的分类、分布、生态和适应性机制，能够揭示海洋生态系统的复杂性，评估物种多样性的变化和动态，并提供可持续管理和保护海洋生态系统的科学依据。作为海洋渔业的重要资源，无脊椎动物提供了丰富的食物来源，如贻贝、虾和甲壳类动物。同时，一些无脊椎动物对于生物医学也具有重要价值。因此，深入了解海洋无脊椎动物的生态学和生物学特征，能够促进可持续的海洋资源利用和保护及推动相关产业的发展。尽管海洋无脊椎动物具有极高的重要性，与脊椎动物和植物相比，无脊椎动物物种相对较少受到关注[3]，常被认为是“较低级”或“更原始”的生物形式[4]。

4.1 海洋无脊椎动物测序进展

在过去的二十年里，日益发展的基因组测序技术推动了重要经济物种（如鱼类、家畜和植物）的研究进展，为经济物种的改良育种提供了坚实的数据基础，也为探究生命演化提供了前所未有的视角。相比之下，无脊椎动物基因组研究主要采用了更“传统”的方法，重点关注形态学和有限的分子或基因，并不能为系统地理解生命演化提供强有力的证据。全球无脊椎动物基因组联盟（GIGA）[5]旨在通过全球范围内的合作，促进无脊椎动物基因组学的发展。青岛华大基因研究院等机构已发起“万种软体动物基因组计划”等，致力于推动无脊椎动物基因组研究。海洋无脊椎动物的全基因组测序和解读将帮助我们识别和理解其重要特性相关的基因、代谢途径和生物过程，有助于揭示它们在生态系统中的角色和功能，以及生物之间的相互作用，更好地评估其对生态系统功能、物种多样性和生态稳定性的影响，有助于开发更精确、可持续的保护和管理策略，以减少对海洋无脊椎动物的负面影响并最大限度地提供生态系统和人类生活的积极效益。我们对已测序的海洋无脊椎动物基因组及研究热点进行了总结和展望。

截至 2023 年 6 月 21 日，共有 640 个海洋无脊椎物种被全基因组测序和组装（数据主要来源于 NCBI）（图 4.2），涵盖了 22 门、56 纲、154 目。节肢动物

（114个物种）、软体动物（143个物种）和刺胞动物（139个物种）占据了一半以上（约 60.74%），其次是线虫动物（60个物种）、棘皮动物（47个物种）、环节动物（30个物种）、扁形动物（30个物种）和脊索动物（24个物种）。2005年，首个海洋无脊椎动物紫海胆（*Strongylocentrotus purpuratus*）[6]被测序。然而，接下来的几年中，海洋无脊椎动物的全基因组测序进程几乎停滞。2010年前后，新一代测序技术的发展和测序成本的降低为海洋无脊椎动物基因组测序研究提供了新的机遇。此后，海洋无脊椎动物的全基因组测序进入爆发期（图 4.3）。

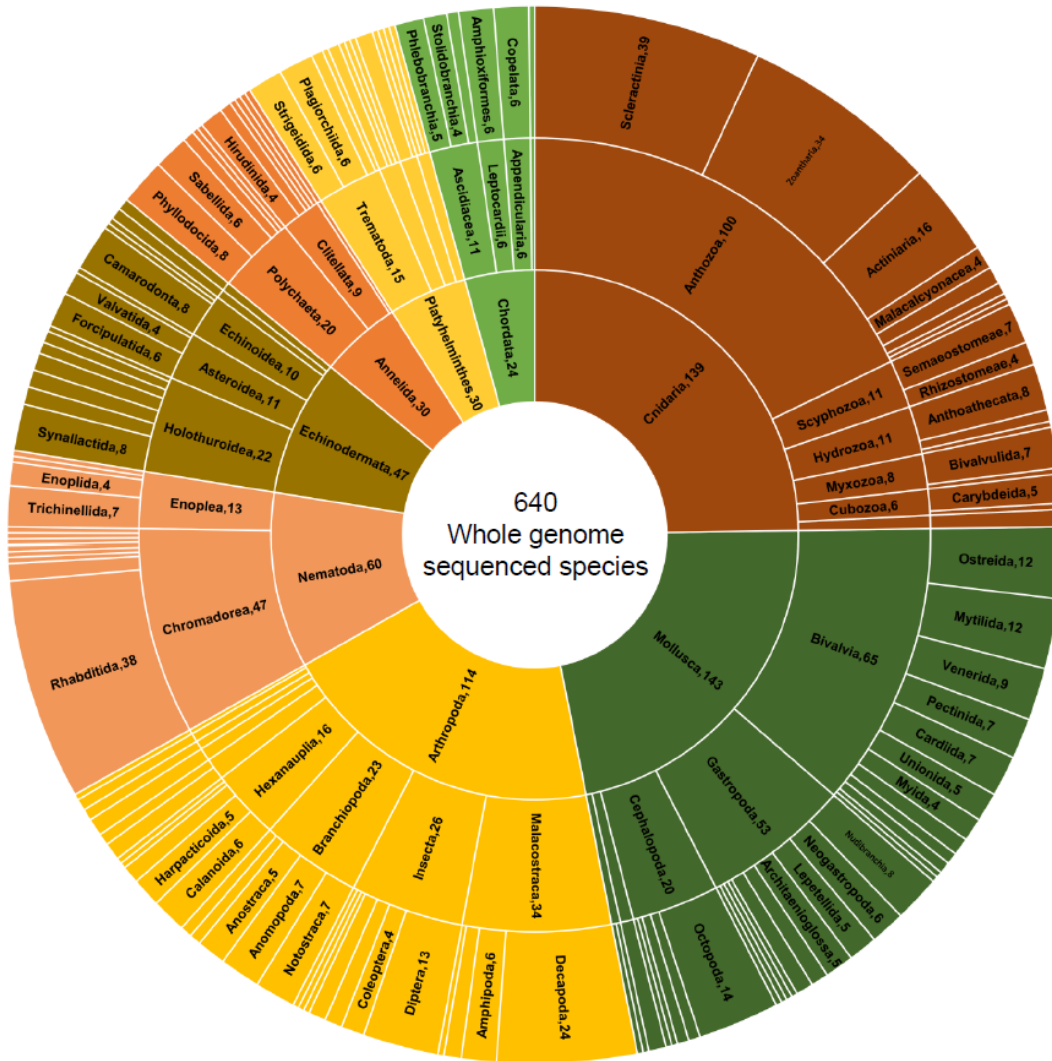


图 4.2. 已测序的海洋无脊椎动物在代表性动物门的数量分布

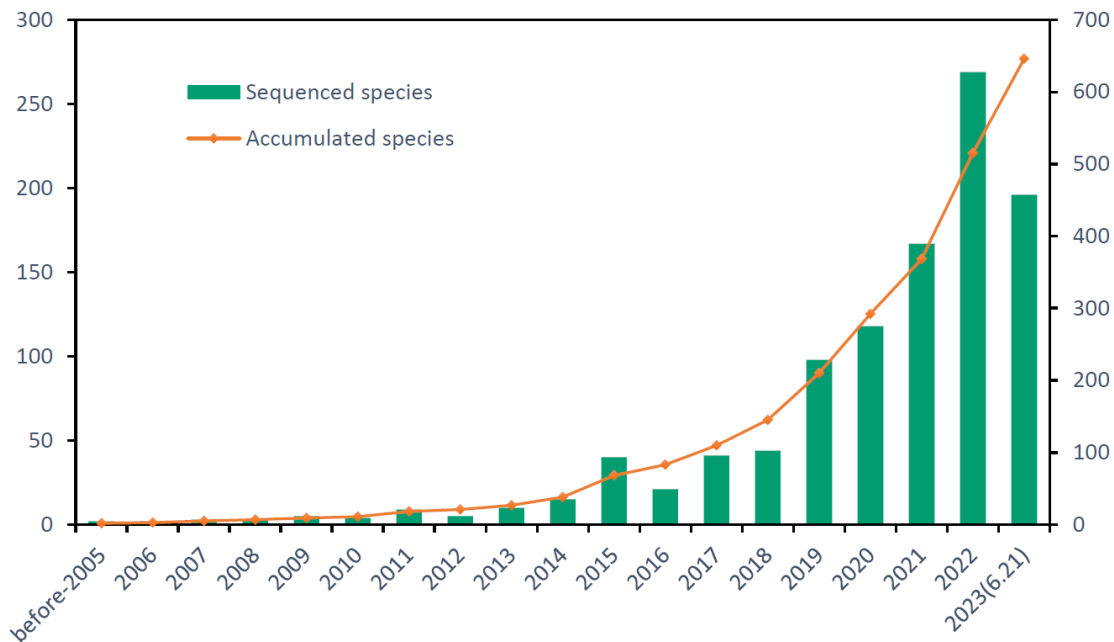


图 4.3. 历年测序的海洋无脊椎动物物种及累计数目

4.2 海洋无脊椎动物基因组特征

海洋无脊椎动物的高多样性首先体现在广泛的基因组大小分布。约 93.74% 的已公布海洋无脊椎动物物种的基因组小于 2 Gb（图 4.4），其中 194 个物种的基因组小于 300 Mb（最小的脊椎动物基因组河豚约 360 Mb [7]），67 个基因组小于 100 Mb。只有 15 个物种（9 个节肢动物和 6 个软体动物）的基因组大于 3 Gb，其中，南极磷虾（*Euphausia superba*）基因组约为 48.01Gb，是已测序的最大动物基因组，其次是脊尾白虾（*Palaemon carinicauda*, 6.70Gb）、詹姆斯具足虫（*Bathynomus jamesi*, 5.89Gb）、夏威夷短尾鱿鱼（*Euprymna scolopes*, 5.28Gb）及红螯螯虾（*Cherax quadricarinatus*, 5.24Gb）。

脊椎动物基因组大小与重复序列含量之间通常有较强的相关性 [8]。然而，整体来看，海洋无脊椎动物基因组的大小与重复序列含量之间并没有明显的线性关系（图 4.4）。例如，棘头虫（*Pomphorhynchus laevis*）基因组大小约为 250 Mb，却含有 65.2% 的重复序列。一些物种既具有大基因组，又具有高重复序列含量。例如，扁足拟石蟹（*Paralithodes platypus*）的基因组大小为 4.8 Gb，含有 77.2% 的重复序

列。另外，有 9 个物种的基因组重复序列含量小于 10%。重复序列亚型的比例也各不相同，约 90.53% 的物种具有低比例的 SINEs（约小于 3%）。海洋无脊椎动物基因组的 GC 含量差异也较大，平均 GC 含量约为 37.35%，略低于海洋哺乳动物（38.2%），但高于鱼类（32.8%）。GC 含量超过 45% 的基因组有 31 个，如 *Bunonema sp.* (54.99%)，但也有 43 个物种的 GC 含量低于 30%（图 4.5）。基因组特征显示出海洋无脊椎动物基因组的复杂性，可能会导致组装困难。在 640 个基因组中，只有 152 个组装到染色体水平。其中，约 37.66% 的基因组（241 个）的 Contig N50 小于 20 Kb，有 187 个小于 10 Kb，143 个小于 5 Kb，表明这些基因组存在大量未组装区域，并且基因的完整性也可能较低。

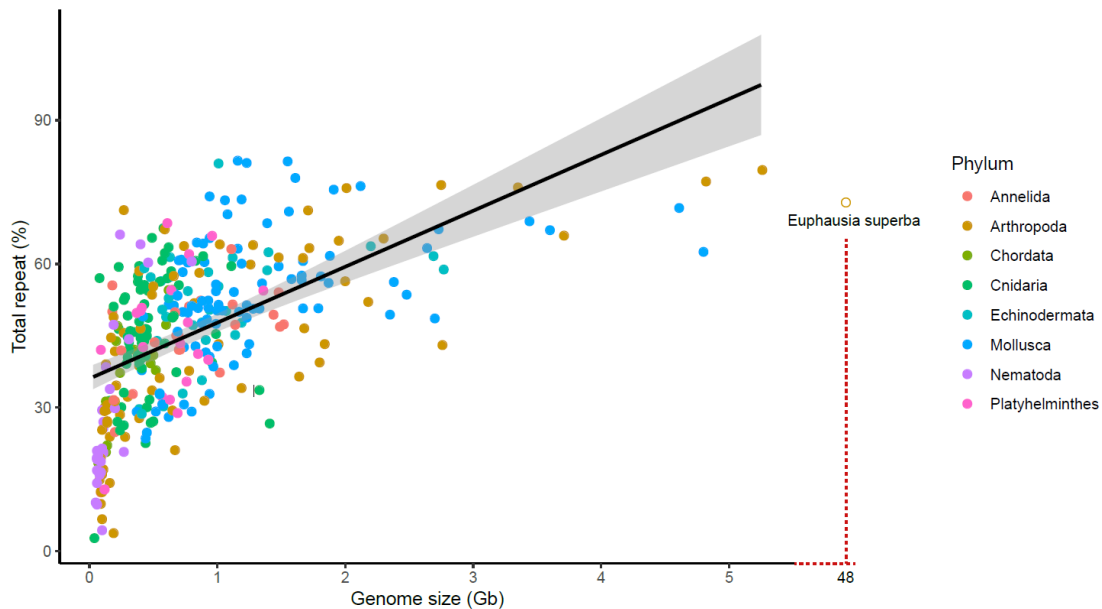


图 4.4. 代表性海洋无脊椎动物门基因组大小与重复序列含量关系。

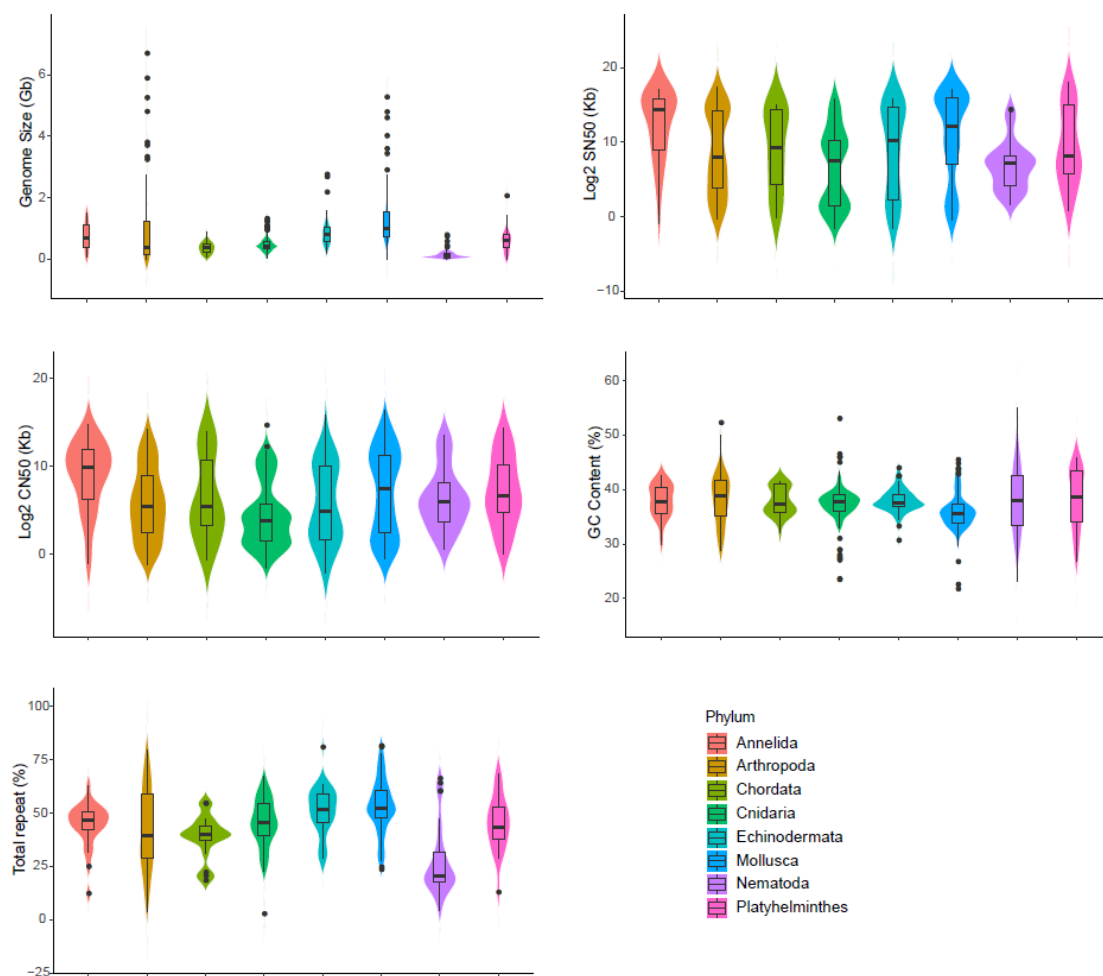


图 4.5. 代表海洋无脊椎动物基因组大小、Contig N50、Scaffold N50、GC 及重复序列含量。

4.3 系统发生关系研究

利用这些基因组数据，我们对无脊椎动物的系统发生关系进行了分析。首先，我们选择了代表性的基因组，选取的基因组满足 Scaffold N50 大于 20 Kb 和 Contig N50 大于 5 Kb 的条件。接下来，我们使用 BUSCO (Benchmarking Universal Single-Copy Orthologs, v5) 和“metazoa_odb10”数据库鉴定了所选择基因组的 BUSCO 基因。根据每个门的完整 BUSCO 值及门类覆盖度，进行了严格的人工筛选。最终，所选择的物种具有 BUSCO 中位数和平均数为 90.20%和 87.00%。另外，我们还选择了四个脊索动物，包括人类、斑马鱼、佛罗里达文昌鱼和海七鳃鳗，它们的完整 BUSCO 值分别为 98.7%、97.9%、96.3%和 93.6%；以及两个外类群物种，包括

Capsaspora owczarzaki 和 *Monosiga brevicollis*。我们使用 MAFFT 对每个单拷贝 BUSCO 基因进行蛋白质序列比对，并使用 Aliscore 和 Alicut 对比对结果进行过滤，以去除含有很少或没有系统发育信号的模糊比对区域，最终得到了一个包含 291,736 个氨基酸的比对矩阵。通过使用 IQ-TREE，并且选择最佳模型构建 974 个基因树，然后使用 ASTRAL 构建了物种树（图 4.6）。此外，我们采用 GBLOCK 对比对结果进行了更为严格的过滤，得到 178,704 个氨基酸的比对矩阵作为额外的数据集进行系统谱系重建（图 4.7）。从构建的系统发育树，我们发现：1）栉水母 Ctenophora 位于基部，是其他动物的姐妹群。2）Placozoa 和 Cnidaria 可能是姐妹群。3）Bryozoa 位于 Lophotrochozoa 的基部，而不在 Lophophorata 内部。虽然这些关键位置的系统发生关系需要进一步的化石数据进行论证，但也证明了全基因组数据是解决这些重大争议进化关系不可或缺的证据。

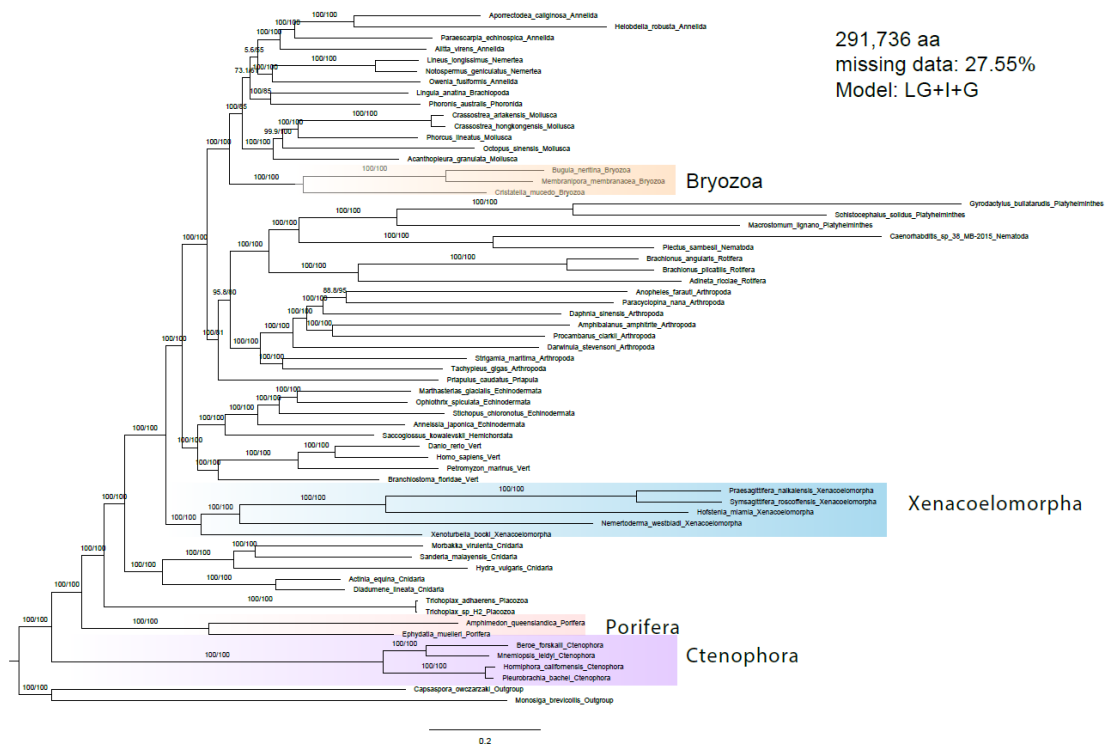


图 4.6. 基于 291736 个氨基酸位点构建的系统发生树

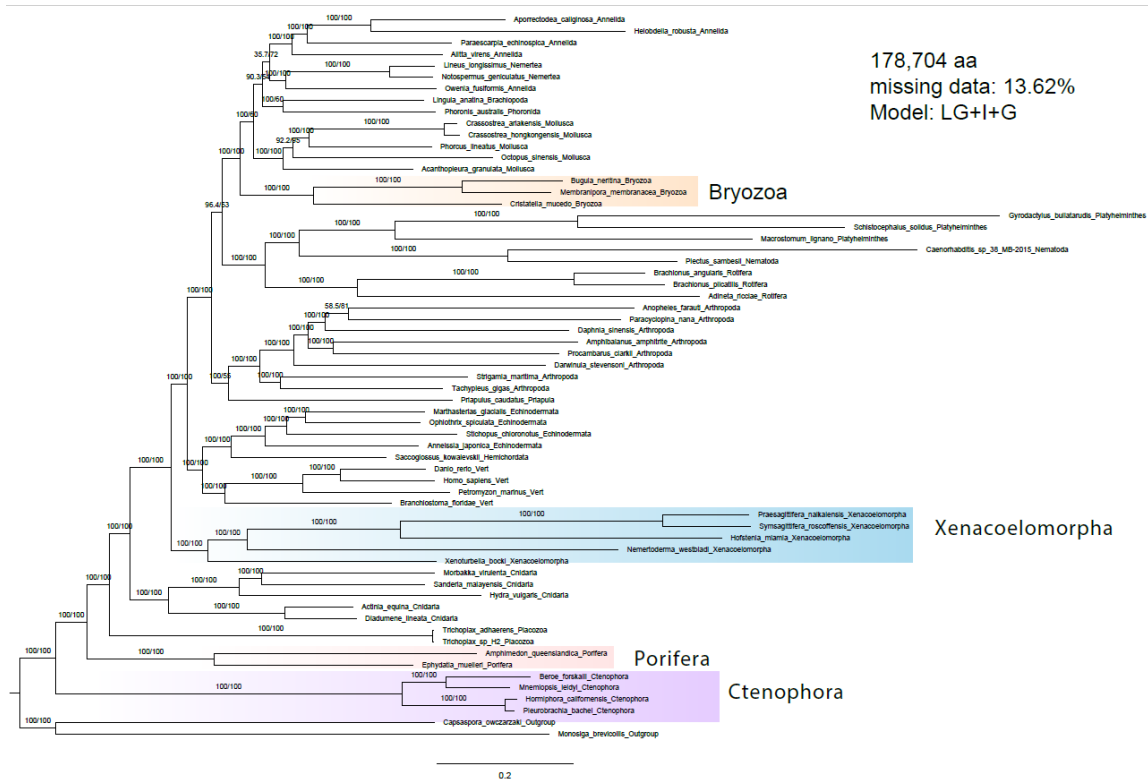


图 4.7. 基于 178704 个氨基酸位点构建的系统发生树

4.4 研究热点

4.4.1 动物发育研究

海绵被认为是地球上最早的多细胞动物之一，动物细胞信号传导和黏附基因的起源可以追溯到单细胞生物向多细胞生物转变的过程中。海绵中存在与脊椎动物蛋白酪氨酸激酶同源的基因[9]，黏附分子和黏附受体在海绵动物中已经存在了 5 亿多年，参与细胞间通讯和信号传导，以及细胞分化、组织形成和免疫应答等生物学过程 [10, 11]。对大堡礁海绵（*Amphimedon queenslandica*）基因组中基因数量和基因家族的演化研究，提出了一些动物特征的演化起源，例如神经系统、免疫系统和生殖系统等 [12]。珊瑚转录组分析发现早期后生动物中存在一系列维持其重要生命活动的关键基因，揭示了珊瑚钙化、共生和昼夜节律的分子基础，为进一步研究珊瑚的生态和保护提供了重要的基础[13]。对半索动物的基因组研究发现，半索动物基

基因组中存在着与“后口”（早期咽部“鳃”缝）发育有关的基因，对深入了解后口动物的起源和演化具有重要意义[14]。

通过对 3 种螺旋动物的全基因组测序及比较基因组学分析，为研究两侧对称动物的起源提供了重要信息，并揭示了谱系特异性的基因组进化[15]。环节动物的功能基因组学研究揭示了两侧对称动物生命周期的起源，头部和躯干形成的时性是两侧对称动物幼虫进化的原因[16]。虾夷盘扇贝（*Patinopecten yessoensis*）的基因组分析发现其染色体与两侧对称动物的假定祖先存在显著的对应关系，或许会成为理解早期两侧对称动物进化的宝贵生物学资源[17]。

头足纲动物拥有复杂的器官结构和神经系统及高度发达的智力，对其基因组和转录组分析，发现了调节神经发育的原钙粘蛋白（protocadherins）和 C2H2-ZNFs 基因家族的显著扩张 [18]；原钙粘蛋白和 GPCR 在章鱼特定神经细胞中共表达[19]；此外，研究发现，章鱼大脑中存在大量特定的 microRNA，这些 microRNA 与神经发育和神经元通信等过程密切相关；并且，这些 microRNA 在章鱼和其他无脊椎动物中具有显著的区别 [20]，揭示了 microRNA 与章鱼复杂大脑之间的紧密联系，为无脊椎动物智能的起源和演化提供了新的线索。此外，单细胞 RNA 测序和 RNA 荧光原位杂交技术，揭示了章鱼 6 种主要的神经元细胞以及非神经元和未成熟神经元群体，并发现了这些神经元的亚型和详细的层组织结构，为进一步研究章鱼视觉系统的发育和功能以及视觉处理的进化提供了基础[21]。椎实螺（*Lymnaea stagnalis*）中枢神经系统的发育在不同年龄阶段可能发生重要变化，通过对其不同年龄阶段的中枢神经系统进行了转录组分析发现，椎实螺中枢神经系统的发育过程存在着显著的基因表达变化，涉及包括神经发育、突触形成和神经递质信号传导等，为理解其神经系统的发育和功能提供了重要线索[22]。Martina Rosato 等首次全面鉴定了海百合纲棘皮动物中的神经肽，研究了这些神经肽在不同发育阶段、不同组织器官的表达模式，为了解海百合类棘皮动物中的神经肽信号传导提供了新的视角和基础，也为研究这些动物的生理和行为提供了重要线索 [23]。鹦鹉螺眼睛发育相关基因家族如光感受器、视黄醛代谢和神经元发育等的研究，有助于我们更好地理解眼睛的起源和进化 [24]。Yu-Qing Wang 等发现厚壳贻贝（*Mytilus coruscus*）幼体利用免疫系

统分子来感知和响应外部信号，支持了免疫和神经内分泌系统在其变态发育中的作用[25]。

4.4.2 免疫功能研究

对感染了十足目虹彩病毒 1（decapod iridescent virus 1）的日本对虾（*Marsupenaeus japonicus*）的转录组分析，发现其免疫应答主要涉及免疫信号通路、炎症反应、细胞凋亡和抗病毒反应等关键生物学过程，并发现了包括 miRNA-基因互作、miRNA-转录因子-基因等潜在调控网络 [26]。对草虾（*Penaeus monodon*）全基因组序列分析发现，其基因组中存在一个新的 Infectious Hypodermal and Hematopoietic Necrosis Virus（IHHNV）的碎片化病毒（EVE）序列，对了解对虾与病毒之间的相互作用以及对虾病毒病的发生机制具有重要意义[27]。比较基因组和进化分析发现，石珊瑚（*Dendrophyllia cribrosa*）在无共生体的情况下，酰辅酶 A 代谢、碳水化合物转运、免疫和钙化相关基因家族都发生了扩张，为了解无共生石珊瑚的生态和进化提供了新的见解 [28]。获得海星衰亡综合征的海星，免疫系统受到损害，其细胞凋亡和炎症反应显著增加，相关通路激活，同时许多抗菌肽和抗微生物蛋白等免疫相关基因表达明显减少 [29]。中华鲎的先天免疫系统发达，IMD 和 JAK-STAT 通路基因存在扩张现象，包括抗菌肽、凝集素、补体系统、细胞因子等基因，在抵抗病原体入侵和维持免疫平衡方面发挥着重要作用 [30]。海胆基因组中先天免疫基因具有多样性，而且包括大量人类疾病相关基因，可以为研究人类健康提供重要资源 [31]。

4.4.3 环境适应性分子机制研究

珊瑚作为共生生物，拥有丰富的微生物菌群。鹿角珊瑚（*Acropora*）多样性较高，主要通过基因家族扩张和基因重组来适应环境变化。二甲基磺酰丙氨酸裂解成二甲硫醚的酶是鹿角珊瑚祖先中重复最多的基因，可对抗温室气体；此外，抗菌肽和共生相关基因在鹿角珊瑚均受到正向选择作用，或许有助于其适应不同的环境[32]。软珊瑚（*Xenia*）基因组及单细胞测序探究了其内共生细胞的起源和发展过程，表

明内共生细胞是一种在软珊瑚组织中存在的特殊细胞类型，与珊瑚的生存和功能密切相关 [33]。三种珊瑚基因组测序发现 HSP20 蛋白的数量和胁迫耐受性之间存在明显的相关性，揭示了不同珊瑚谱系的进化轨迹，也揭示了不同珊瑚谱系的遗传基础和适应能力 [34]。珊瑚成体及精子的测序表明干细胞、体细胞和生殖细胞具有不同的突变率和突变模式，对于理解动物中生殖细胞系的进化以及珊瑚如何在应对气候变化中产生适应性多样性具有重要意义 [35]。管状蠕虫 (*Paraescarpia echinospica*) 基因组测序表明，与底物转运、先天免疫、抑制细胞凋亡、溶酶体消化、氧化应激保护相关的基因家族显著扩张，与其环境适应相关 [36]。入侵物种具有很强的环境适应能力，斑马贻贝 (*Dreissena polymorpha*) 基因组中富含与黏附、免疫应答和适应环境等过程相关的基因，并且存在大量的反转录转座子和缺失区域，可能与其适应性进化有关 [37]。红沼虾基因组中存在大量的基因变异及基因家族扩张，为其在适应寒冷环境方面提供了遗传基础 [38]。海鞘是一种分布于全球的入侵性被囊动物，其锌指超家族的基因数量差异最为显著，可能与物种分化和环境适应有关 [39]。牡蛎 (*Saccostrea glomerata*) 通常具有出色的环境适应能力，基因组中多种复杂的免疫受体为其提供了识别入侵微生物的独特能力 [40]。深海海参 (*Chiridota heheva*) 基因组中具有适应深海生活的多个特化基因，其基因组进化速度较慢，可能与其在深海环境中较低的代谢率和繁殖速率有关 [41]。条纹叉状贻贝 (*Mytilisepta virgata*) 中 MvUSP15 基因中的同义突变可以影响基因表达，并与微生物异质性相互作用，可影响热阻力，揭示了适应性同义突变作为对热应激的一种被低估的遗传缓冲的重要性 [42]。洗衣板蚌 (*Megalonaias nervosa*) 基因组分析表明，基因复制是其近期适应性提升的原因之一，揭示在最近突然面临环境挑战的物种中基因复制的重要性 [43]。

鳞足蜗牛基因组中存在大量与生物矿化相关的基因，包括胶原蛋白、蛋白聚糖、壳蛋白等，控制着其生物矿化壳的形成和发育 [44]，生物矿化基因在鹦鹉螺壳的形成和修复中具有重要作用 [24]。马氏珠母贝基因组中存在大量与无机物质运输、细胞外基质合成和细胞分化等相关的基因，以及新的与生物矿化相关基因和分子信号通路，可以影响贝壳和珍珠的形成 [45]。蛤蜊基因组数据表明，壳的颜色受到酪氨

酸酶及 MITF 基因家族的调控[46,47]。基因家族的扩张可能是珍珠贝适应固着生活方式的关键因素之一[48]。牡蛎基因组表明，一组复杂的壳形成相关基因的表达模式和变异可以影响壳的性状、结构和颜色 [49]。苹果螺科包括水生和两栖螺，在两栖的物种中，新发现了一个卵神经毒素和一个用于制造钙质卵壳的蛋白质，可能是其从水下到陆地产卵过渡的关键适应基础[50]。

海胆基因组中细胞凋亡和自噬相关的基因可能影响海胆的衰老过程[51]。在永生的腔肠动物中发现了与 DNA 复制、DNA 修复、端粒维持、氧化还原、干细胞群体和细胞间通信相关基因的变异和扩增，此外多能性诱导因子的沉默和多能性靶点的激活或许在其复苏过程中起关键作用[52]。摇蚊（*Clunio marinus*）的繁殖受昼夜节律和月相节律影响，与昼夜节律最密切相关的基因组区域表现出钙/钙调蛋白依赖性激酶 II.1（CaMKII.1）的特异性差异；此外，可变剪接可能是其调节昼夜节律的遗传基础[66]。南极磷虾基因组及转录组数据显示，昼夜节律相关的部分基因在夏季和冬季显示出不同的表达水平，表明南极磷虾已经进化出由昼夜节律系统控制的 身体适应和行为模式，有助于它们在低温和剧烈变化的光照条件下保存能量并生存[146]。

4.4.4 生态学研究

不同的共生生物群落通过介导光合作用、碳、氮循环等过程来增加海绵的代谢能力[53-58]。Acropora 基因组的研究为珊瑚共生的分子基础和对环境变化的反应提供了重要的见解[59]。珊瑚的先天复杂免疫系统可以更好地应对环境压力和病原体，但是很难确定石珊瑚的鹿角珊瑚基因组在多大程度上反映了珊瑚的一般特征，或者它在多大程度上与其他珊瑚的基因组不同[60]。通过对柱头珊瑚（*Stylophora pistillata*）和鹿角珊瑚（*Acropora digitifera*）基因组的比较，发现其核心保守蛋白集中在刺胞—鞭毛共生过程中。在这两个物种中都发现了与藻类共生、先天免疫和应激反应有关的基因的独立、不均匀扩展，这表明它们是截然不同的基因组 [61]。对来自澳大利亚和冲绳的棘鲈（*Acanthaster planci*）的全基因组测序 [62]，揭示了与珊瑚聚集活性相关的物种特异性交流因子的关键基因和生物网络调控模型。

4.4.5 遗传改良研究

梨形环棱螺 (*Bellamya purificata*) 的首个染色体基因组的组装和分析, 为其遗传改良及田螺科的生态和进化研究提供了重要的基因组资源[63]。虾夷扇贝 (*Patinopecten yessoensis*) 和栉孔扇贝 (*Chlamys farreri*) 的全基因组测序加速了我国重要养殖贝类的育种进程[17,64]。通过对日本对虾 (*Marsupenaeus japonicus*) 和草虾 (*Penaeus monodon*) 两种经济对虾的基因组数据分析, 鉴定出对其体型具有重要影响的基因, 为对虾的选择性育种和一些生物学特性的研究提供了宝贵的资源[65]。

4.4.6 系统发生研究

水螅 *Hydroidolina* 基因组的研究为水螅纲的系统发生提出新的假说提供了有力证据, 其结果表明 *Hydroidolina* 属于一个单系群, 与其他水螅纲的物种有着相对较远的亲缘关系, 并将 *Hydroidolina* 分为两个亚目[68]。栉水母、海绵及近源单细胞物种的基因组结构分析表明, 栉水母是所有其他动物的姐妹群[69]。扁盘动物基因组的分析, 揭示了扁盘动物的本质特征和演化地位[70]。多个文昌鱼基因组的比较研究, 探究了脊椎动物的起源和进化, 表明脊索动物的基因组演化速度较快, 且存在更高的基因组变异和基因家族扩张[71]。

4.5 讨论与展望

尽管基因组测序已经取得了一些进展, 但因为无脊椎动物的多样性极其庞大, 未被记载的无脊椎动物可能仍有数百万种, 未来仍要针对其重要研究方向, 开展广泛的基因组学研究。

冠轮动物超门拥有包含软体动物、环节动物和扁形动物等在内的 12 个动物门类, 现有记录物种超过 8 万。冠轮动物分化较早、进化时间跨度大, 在生命演化、神经、免疫系统发育等方面具有极为重要的科学意义。2021 年 10 月, 青岛华大基因研究院联合浙江大学、中国海洋大学、中国科学院海洋研究所、美国新泽西州立

大学、维也纳大学等单位共同发起“万种软体动物基因组计划（M10K+ project）”，计划分三个阶段，历时十年，绘制贝类和其他冠轮动物等的万种基因组图谱，涵盖整个冠轮动物超门，建立一个大规模、高质量的基因组数据库，为动物的系统演化等重要科学问题提供有效和可靠的数据基础。该计划鼓励国内及国际合作，欢迎来自不同地区的科学家能够共同参与基因组测序和分析工作、积极推进资源共享、技术互助，加强科学交流，从而促进生命演化、海洋资源利用和生物医学、生物多样性保护和可持续发展等方面的研究。

在方法上，要采用更先进的测序技术、开发针对性的分析方法，以提高数据产出的效率和准确性。此外，需整合基因组数据和其他组学数据信息，如单细胞（时空）组学、表观遗传学、蛋白质组学等，这种综合分析将更加深入地揭示无脊椎动物基因组的功能和调控机制。了解不同无脊椎动物之间的遗传关系和进化历史，全面地了解生命的演化。

本章作者：

张要磊、王佳昊、吕美琪、高昊阳（青岛华大基因研究院）

本章参考文献

- 1 May, R. M. How many species are there on earth? *Science* **241**, 1441-1449 (1988).
- 2 Horton, T. *et al.* (WoRMS Editorial Board, 2020).
- 3 Eisenhauer, N., Bonn, A. & C, A. G. Recognizing the quiet extinction of invertebrates. *Nat Commun* **10**, 50, doi:10.1038/s41467-018-07916-1 (2019).
- 4 Darwin, C. *The origin of species*. (PF Collier & son New York, 1909).
- 5 Scientists, G. C. o. *et al.* The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. *J Hered* **105**, 1-18, doi:10.1093/jhered/est084 (2014).
- 6 Sea Urchin Genome Sequencing, C. *et al.* The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941-952, doi:10.1126/science.1133609 (2006).
- 7 Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301-1310 (2002).
- 8 Song, Y. *et al.* Diversity, function and evolution of aquatic vertebrate genomes. *bioRxiv*, 2021.2010. 2029.466026 (2021).
- 9 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).
- 10 Müller, W. E. 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).
- 11 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* **103**, 12451-12456 (2006).
- 12 Srivastava, M. *et al.* The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720-726 (2010).
- 13 Han, T. *et al.* Full-length transcriptome maps of reef-building coral illuminate the molecular basis of calcification, symbiosis, and circadian genes. *International Journal of Molecular Sciences* **23**, 11135 (2022).

- 14 Simakov, O. *et al.* Hemichordate genomes and deuterostome origins. *Nature* **527**, 459-465 (2015).
- 15 Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526-531 (2013).
- 16 Martín-Zamora, F. M. *et al.* Annelid functional genomics reveal the origins of bilaterian life cycles. *Nature* **615**, 105-110, doi:10.1038/s41586-022-05636-7 (2023).
- 17 Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature ecology & evolution* **1**, 0120 (2017).
- 18 Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220-224 (2015).
- 19 Jiang, D. *et al.* The gold-ringed octopus (*Amphioctopus fangsiao*) genome and cerebral single-nucleus transcriptomes provide insights into the evolution of karyotype and neural novelties. *BMC Biology* **20**, doi:10.1186/s12915-022-01500-2 (2022).
- 20 Zolotarov, G. *et al.* (2022).
- 21 Songco-Casey, J. O. *et al.* Cell types and molecular architecture of the Octopus bimaculoides visual system. *Current Biology* **32**, 5031-5044.e5034, doi:10.1016/j.cub.2022.10.015 (2022).
- 22 Rosato, M., Hoelscher, B., Lin, Z., Agwu, C. & Xu, F. Transcriptome analysis provides genome annotation and expression profiles in the central nervous system of *Lymnaea stagnalis* at different ages. *BMC genomics* **22**, 1-16 (2021).
- 23 Aleotti, A. *et al.* Discovery and functional characterization of neuropeptides in crinoid echinoderms. *Frontiers in Neuroscience* **16**, doi:10.3389/fnins.2022.1006594 (2022).
- 24 Zhang, Y. *et al.* The genome of *Nautilus pompilius* illuminates eye evolution and biomineralization. *Nature Ecology & Evolution* **5**, 927-938 (2021).
- 25 Wang, Y.-Q. *et al.* Stage-Specific Transcriptomes of the Mussel *Mytilus coruscus* Reveals the Developmental Program for the Planktonic to Benthic Transition. *Genes* **14**, 287, doi:10.3390/genes14020287 (2023).
- 26 He, Z. *et al.* Integrated analysis of mRNA-Seq and miRNA-Seq reveals the molecular mechanism of the intestinal immune response in *Marsupenaeus japonicus* under decapod iridescent virus 1 infection. *Frontiers in immunology* **12**, 807093 (2022).

- 27 Huerlimann, R. *et al.* Genome assembly of the Australian black tiger shrimp (*Penaeus monodon*) reveals a novel fragmented IHNV EVE sequence. *G3* **12**, jkac034 (2022).
- 28 Kim, J. *et al.* Comparative genome and evolution analyses of an endangered stony coral species *Dendrophyllia cribrosa* near Dokdo Islands in the East Sea. *Genome Biology and Evolution* **14**, evac132 (2022).
- 29 Ruiz-Ramos, D. V., Schiebelhut, L. M., Hoff, K. J., Wares, J. P. & Dawson, M. N. An initial comparative genomic autopsy of wasting disease in sea stars. *Molecular Ecology* **29**, 1087-1102 (2020).
- 30 Zhou, Y. *et al.* The draft genome of horseshoe crab *Tachypleus tridentatus* reveals its evolutionary scenario and well-developed innate immunity. *BMC genomics* **21**, 1-15 (2020).
- 31 Sodergren, E. *et al.* The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941-952 (2006).
- 32 Shinzato, C. *et al.* Eighteen Coral Genomes Reveal the Evolutionary Origin of Acropora Strategies to Accommodate Environmental Changes. *Mol Biol Evol* **38**, 16-30, doi:10.1093/molbev/msaa216 (2021).
- 33 Hu, M., Zheng, X., Fan, C.-M. & Zheng, Y. Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*. *Nature* **582**, 534-538 (2020).
- 34 Ying, H. *et al.* Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. **19**, 175, doi:10.1186/s13059-018-1552-8 (2018).
- 35 López-Nandam, E. H., Albright, R., Hanson, E. A. & Sheets, E. A. Mutations in coral soma and sperm imply lifelong stem cell renewal and cell lineage selection. **290**, 20221766, doi:10.1098/rspb.2022.1766 (2023).
- 36 Sun, Y. *et al.* Genomic Signatures Supporting the Symbiosis and Formation of Chitinous Tube in the Deep-Sea Tubeworm *Paraescarpia echinospica*. *Molecular Biology and Evolution* **38**, 4116-4134, doi:10.1093/molbev/msab203 (2021).
- 37 McCartney, M. A. & Auch, B. The genome of the zebra mussel, *Dreissena polymorpha*: a resource for comparative genomics, invasion genetics, and biocontrol. **12**, doi:10.1093/g3journal/jkab423 (2022).
- 38 Sato, D. X. *et al.* Genomic adaptive potential to cold environments in the invasive red swamp crayfish. *iScience* **26**, 107267, doi:10.1016/j.isci.2023.107267 (2023).

- 39 Zhang, J., Wei, J., Yu, H. & Dong, B. Genome-Wide Identification, Comparison, and Expression Analysis of Transcription Factors in Ascidian *Styela clava*. **22**, doi:10.3390/ijms22094317 (2021).
- 40 Powell, D. *et al.* The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. *DNA research : an international journal for rapid publication of reports on genes and genomes* **25**, 655-665, doi:10.1093/dnares/dsy032 (2018).
- 41 Zhang, L. *et al.* The genome of an apodid holothuroid (*Chiridota heheva*) provides insights into its adaptation to a deep-sea reducing environment. **5**, 224, doi:10.1038/s42003-022-03176-4 (2022).
- 42 Tan, Y., Ma, C.-Y., Li, X.-X., Han, G.-D. & Dong, Y.-W. Genome-wide sequencing identifies a thermal-tolerance related synonymous mutation in the mussel, *Mytilisepta virgata*. *Communications Biology* **6**, doi:10.1038/s42003-022-04407-4 (2023).
- 43 Rogers, R. L., Grizzard, S. L., Garner, J. T. & Larracuente, A. Strong, Recent Selective Sweeps Reshape Genetic Diversity in Freshwater Bivalve *Megaloniais nervosa*. *Molecular Biology and Evolution* **40**, doi:10.1093/molbev/msad024 (2023).
- 44 Sun, J. & Chen, C. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. **11**, 1657, doi:10.1038/s41467-020-15522-3 (2020).
- 45 Du, X. *et al.* The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *GigaScience* **6**, 1-12, doi:10.1093/gigascience/gix059 (2017).
- 46 Wei, M. *et al.* Chromosome-Level Clam Genome Helps Elucidate the Molecular Basis of Adaptation to a Buried Lifestyle. *iScience* **23**, 101148, doi:10.1016/j.isci.2020.101148 (2020).
- 47 Yan, X. *et al.* Clam Genome Sequence Clarifies the Molecular Basis of Its Benthic Adaptation and Extraordinary Shell Color Diversity. *iScience* **19**, 1225-1237, doi:10.1016/j.isci.2019.08.049 (2019).
- 48 Takeuchi, T. *et al.* Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Lett* **2**, 3, doi:10.1186/s40851-016-0039-2 (2016).
- 49 Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49-54, doi:10.1038/nature11413 (2012).

- 50 Sun, J. *et al.* Signatures of Divergence, Invasiveness, and Terrestrialization Revealed by Four Apple Snail Genomes. *Mol Biol Evol* **36**, 1507-1520, doi:10.1093/molbev/msz084 (2019).
- 51 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* **8**, 260-271, doi:10.18632/aging.100889 (2016).
- 52 Pascual-Torner, M. & Carrero, D. Comparative genomics of mortal and immortal cnidarians unveils novel keys behind rejuvenation. **119**, e2118763119, doi:10.1073/pnas.2118763119 (2022).
- 53 Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. *Nature reviews. Microbiology* **10**, 641-654, doi:10.1038/nrmicro2839 (2012).
- 54 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. *Mol Ecol* **24**, 1263-1274, doi:10.1111/mec.13107 (2015).
- 55 Reiswig, H. M. Partial Carbon and Energy Budgets of the Bacteriosponge *Verohgia fistularis* (Porifera: Demospongiae) in Barbados. *Marine Ecology* **2**, 273-293, doi:<https://doi.org/10.1111/j.1439-0485.1981.tb00271.x> (1981).
- 56 Schläppy, 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).
- 57 Taylor, M. W., Radax, R., Steger, D. & Wagner, M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and molecular biology reviews : MMBR* **71**, 295-347, doi:10.1128/mmb.00040-06 (2007).
- 58 Webster, N. S. & Taylor, M. W. Marine sponges and their microbial symbionts: love and other relationships. *Environmental microbiology* **14**, 335-346, doi:10.1111/j.1462-2920.2011.02460.x (2012).
- 59 Shinzato, C. *et al.* Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320-323, doi:10.1038/nature10249 (2011).
- 60 Voolstra, C. R. & Li, Y. Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals. **7**, 17583, doi:10.1038/s41598-017-17484-x (2017).

- 61 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).
- 62 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).
- 63 Jin, W. *et al.* Chromosome-level genome assembly of the freshwater snail *Bellamya purificata* (Caenogastropoda). *Zoological research* **43**, 683-686, doi:10.24272/j.issn.2095-8137.2022.118 (2022).
- 64 Li, Y. *et al.* Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nature communications* **8**, 1721 (2017).
- 65 Yuan, J. *et al.* Genomic resources and comparative analyses of two economical penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*. *Marine Genomics* **39**, 22-25, doi:<https://doi.org/10.1016/j.margen.2017.12.006> (2018).
- 66 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).
- 67 Shao, C. *et al.* The enormous repetitive Antarctic krill genome reveals environmental adaptations and population insights. *Cell* **186**, 1279-1294. e1219 (2023).
- 68 Bentlage, B. & Collins, A. G. Tackling the phylogenetic conundrum of Hydroidolina (Cnidaria: Medusozoa: Hydrozoa) by assessing competing tree topologies with targeted high-throughput sequencing. *PeerJ* **9**, e12104 (2021).
- 69 Schultz, D. T. *et al.* Ancient gene linkages support ctenophores as sister to other animals. *Nature*, 1-8 (2023).
- 70 Srivastava, M. *et al.* The Trichoplax genome and the nature of placozoans. *Nature* **454**, 955-960 (2008).
- 71 Huang, S. *et al.* Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nature communications* **5**, 5896 (2014).

533个

鱼类基因组

涵盖现存鱼类物种的

1.78%



第五章 鱼类

现存鱼类数量超过三万种，可分为两大类：无颌鱼类（Agnatha）（如海七鳃鳗）和有颌鱼类（Gnathostomata）（分别为硬骨鱼类和软骨鱼类）[1]。硬骨鱼包括肉鳍鱼类（Lobe-finned fish）和辐鳍鱼类（Ray-finned fish），它们在 4.3 亿多年前分化[2-4]。目前最大的类群（超过 95% 的鱼类物种）是辐鳍鱼，这其中就包括真骨鱼。自二叠纪真骨鱼开始出现以来[5, 6]，真骨鱼已经进化为数量最多、生物多样性最丰富的类群。它们的领地已从其祖先主要占据的浅水环境扩展到覆盖全球大部分水体，从高原到马里亚纳海沟，从淡水到海洋。真骨鱼的起源以及是什么遗传创新使它们成功进化仍然是鱼类进化的未解之谜。真骨鱼类的适应性辐射的演化方式给系统发育推断带来了巨大挑战。

在过去的 50 年中，分子系统学的出现和利用大规模基因组数据进行的系统发育学研究在解决这些难题方面取得了巨大进步[2, 4, 7-11]。关于鱼类生命树的开创性资料是 Nelson 的《Fishes of the World》[12]，该书主要根据形态学证据划分系统发育关系。尽管分子证据对鱼类系统发育学的贡献已得到重视，并且在 2017 年 Betancur-R 依靠分子和基因组学数据对近 2,000 种硬骨鱼类进行了分类和系统谱系分析[13]，但鱼类的系统发生关系在不同分类水平上仍未得到很好的解决，鱼类共性和特异性的遗传特征，如鱼鳔和运动器官（胸鳍）的进化，在很大程度上仍是未解之谜。

近年来高通量测序技术和分析方法的发展促进了更加细致精确的基因组得以构建，而所需的成本和时间仅为 20 年前人类基因组测序所需的一小部分[14-19]。高质量的参考基因组为大规模和全面的功能基因组学研究铺平了道路。脊椎动物基因组计划（Vertebrate Genome Project）、B10K[20, 21]、Genome10K 计划[22]和地球生物基因组计划（EBP）[23]等为脊椎动物基因组的大规模测序做出了大量努力。然而许多鱼类的基因组测序工作仍具有挑战性：它们的基因组复杂多样，许多品系都含有多倍体或大基因组；而且样本采集困难，尤其是那些生存在独特栖息地的物

种[24-27]。尽管在对大基因组物种和稀有鱼类的测序方面取得了一定的进展，如 40 GB 的肺鱼基因组[3, 4]和生活在马里亚纳海沟 8,000 米深处的狮子鱼基因组[28]，但迄今为止基因组研究仍没有覆盖所有鱼类的多样性[29]。

为了探索鱼类的多样性并揭示不同鱼类的基因组多样性，华大主导发起的万种鱼基因组计划（Fish10K），对超过 87%的现存鱼目水平的代表性物种基因组进行了收集、测序、注释和系统评估，以增加对鱼类基因组多样性的了解。利用该项目第一阶段获得的前所未有的鱼类基因组数据，建立了迄今为止最全面的系统发生树，并系统地探索了鱼类的多样性和适应性演化。

5.1 全面的鱼类基因组数据集

Fish10K 第一阶段共收集了 533 个鱼类基因组，涵盖 87%的鱼现存的目（71/82）[12]（图 5.1）。其中有 116 个物种（包括 95 种硬骨鱼类和 21 种软骨鱼类）利用 stLFR 测序技术[30]和长读长测序技术进行了该物种的首次基因组测序，产生总数据量约 16.1Tb。这些新测序基因组包括以前未测序的 8 个目，其中 6 个是软骨鱼纲，2 个是硬骨鱼纲（图 5.1），将软骨鱼类物种基因组数量增加了三倍。所有数据均可从中国国家基因库数据库（<https://db.cngb.org>；accession number: DATAani16）获取。在新组装的 116 个基因组中，硬骨鱼和软骨鱼的平均 Scaffold N50 分别为 4.12Mb 和 1.99Mb。大约 65.51%的基因组（76/116）的 Scaffold N50 大于 1.00Mb（图 5.2），其中大多数基因组包含超过 85.00%的完整的 BUSCO 基因，BUSCO 评估基因组完整度中位值为 96%，以上指标表明新组装的鱼类基因组具有很高的连续性和完整性。

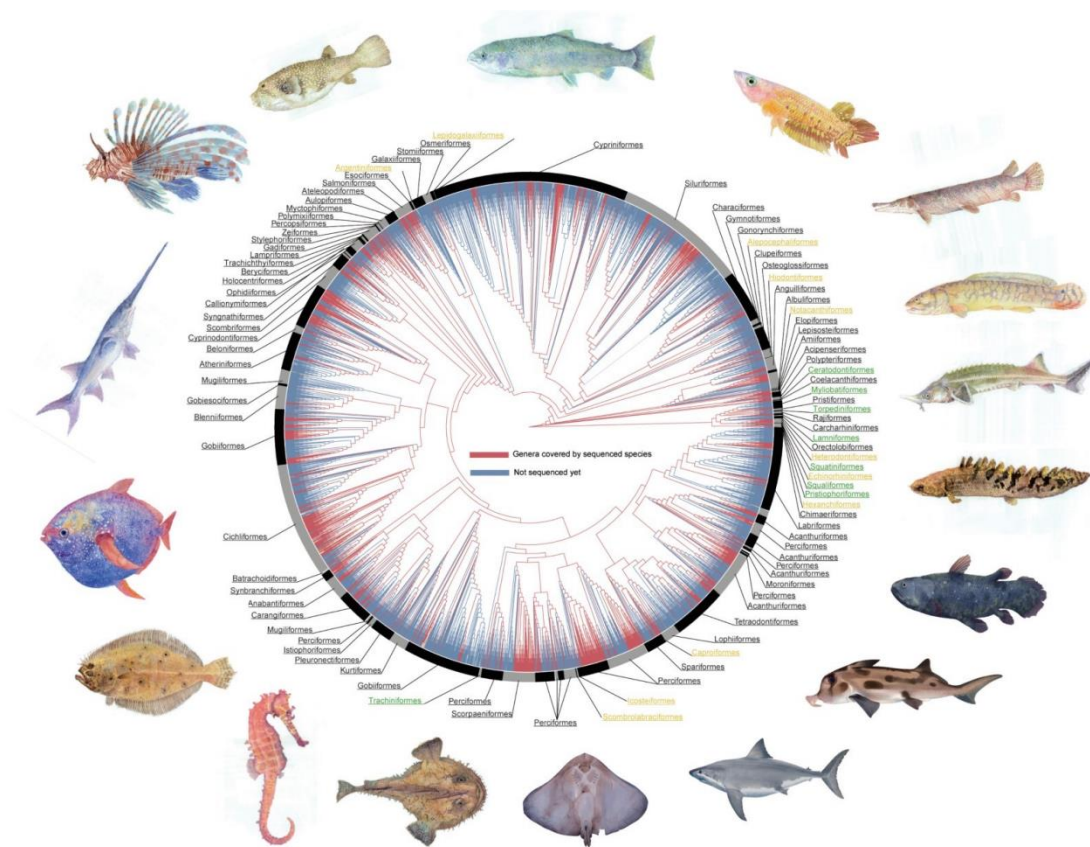


图 5.1. 现存鱼类 82 个目的概述。迄今为止已经记录了超过 30,000 种鱼类物种，然而只有小部分具有基因组序列（来自 FishBase 网站的物种列表，<https://www.fishbase.se/search.php>）。该草图进化树是通过将 Time Tree 网站的进化信息与使用 MiToFish 数据库中的线粒体遗传数据推断出的进化树相结合而产生的（<http://mitofish.aori.u-tokyo.ac.jp/>）。该树显示了已测序和未测序物种的分布，分别用红色和蓝色表示。已测序物种（包括本研究中新测序的 116 个物种）涵盖了 71 个目，有 11 个目（橙色）由于采样困难而尚未被测序。在这些已测序的目中，八个目（绿色）是新测序的。

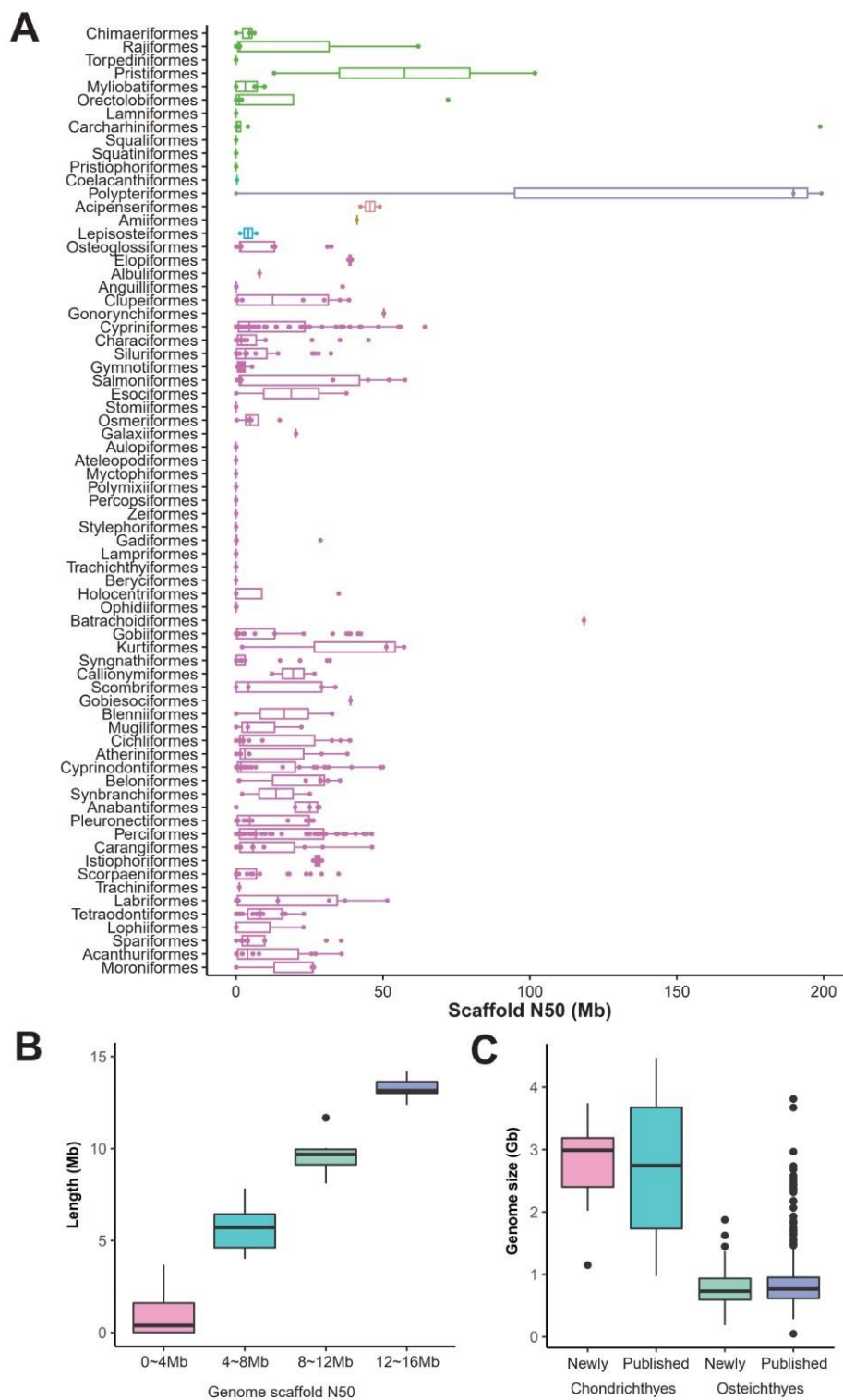


图 5.2. 鱼类基因组的质量评估。(A) 来自 71 个目的 533 个鱼类物种基因组的 Scaffold N50 (Mb) 分布 (按目分组)。(B) Fish10K 项目测序的 116 个鱼类物种中, 约有 65.5% (76 个)

的 Scaffold N50 大于 1Mb。(C) 116 个新测序的鱼类物种(其中 95 个为硬骨鱼, 21 个为软骨鱼)的组装大小与其已发表的有颌类物种的基因组质量相近。

5.2 基于全基因组数据的鱼类系统发生树

真骨鱼类的起源时间和现存真骨鱼鱼类最基部的类群关系一直是鱼类系统发育学中的研究热点, 是鱼类学中不可或缺的一环, 也是揭开真骨鱼繁衍之谜的基础。此外, 骨鳔鱼类四个目之间的系统发育关系以及鲈形目是否为一个单系群仍存在争议[12, 31]。为了解决当前的这些系统发育学争论, 我们根据 BUSCO 基因集中的单拷贝直系同源基因对 533 种鱼类进行了系统发育分析。在对 BUSCO 基因进行严格筛选后(图 5.3), 基于 3,005 个辐鳍鱼和 2,501 个脊椎动物单拷贝 BUSCO 基因构建了 533 种鱼类的基于聚合的最大似然法(ML)物种树[32-34](图 5.3)。

5.2.1 早期真骨鱼的进化关系

本研究的鱼类系统发育树结果表明骨舌鱼目是最早出现的真骨鱼。这一结论与 Betancur-R 等人(2017 年)基于线粒体遗传数据[13]得出的拓扑结构不同, 后者显示海鲢目是最基部分类, 骨舌鱼目和鲱形总目是姊妹群。最近的相关研究也提出了第三个不同的观点, 即认为鲱形总群是骨舌鱼目和海鲢总目的姊妹群。为了进一步确认早期真骨鱼三大分支的进化关系, 我们使用 GGI 方法[35]对以上三种不同的拓扑结构进行了验证。支持率最高的 GGI 树将骨舌鱼目单独列为鲱形总群和海鲢总目的姐妹群(图 5.4A、B 中的节点 a)。在前两个备选假说中, H032 与最近一项研究的 ML 树一致, 该树显示海鲢总目是骨舌鱼目和鲱形总群的祖先[36]; 然而, 与拥有最高支持度的 GGI 拓扑结构(H01)相比, 该假说代表的进化关系的支持率要低得多(图 5.4)。此外, 我们 GGI 检测结果中 H03 和 H04 代表了相同的拓扑结构, 即鲱形总群是骨舌鱼目和海鲢总目的姊妹群; 我们发现该拓扑结构在 GGI 测试中的支持度与 H01 的支持度基本一致, 这说明早期真骨鱼的进化关系更可能是这两种拓扑结构中的一种。在这里我们采用 GGI 测试支持度最高的拓扑结构, 即认为骨舌鱼类是海鲢总目和鲱形总群的姊妹群。这些关键位置的系统发生关系需要进

一步的化石数据并结合表型数据进行论证，但也证明了更多更全面的基因组数据资源是解决这些重大争议进化关系不可或缺的证据。

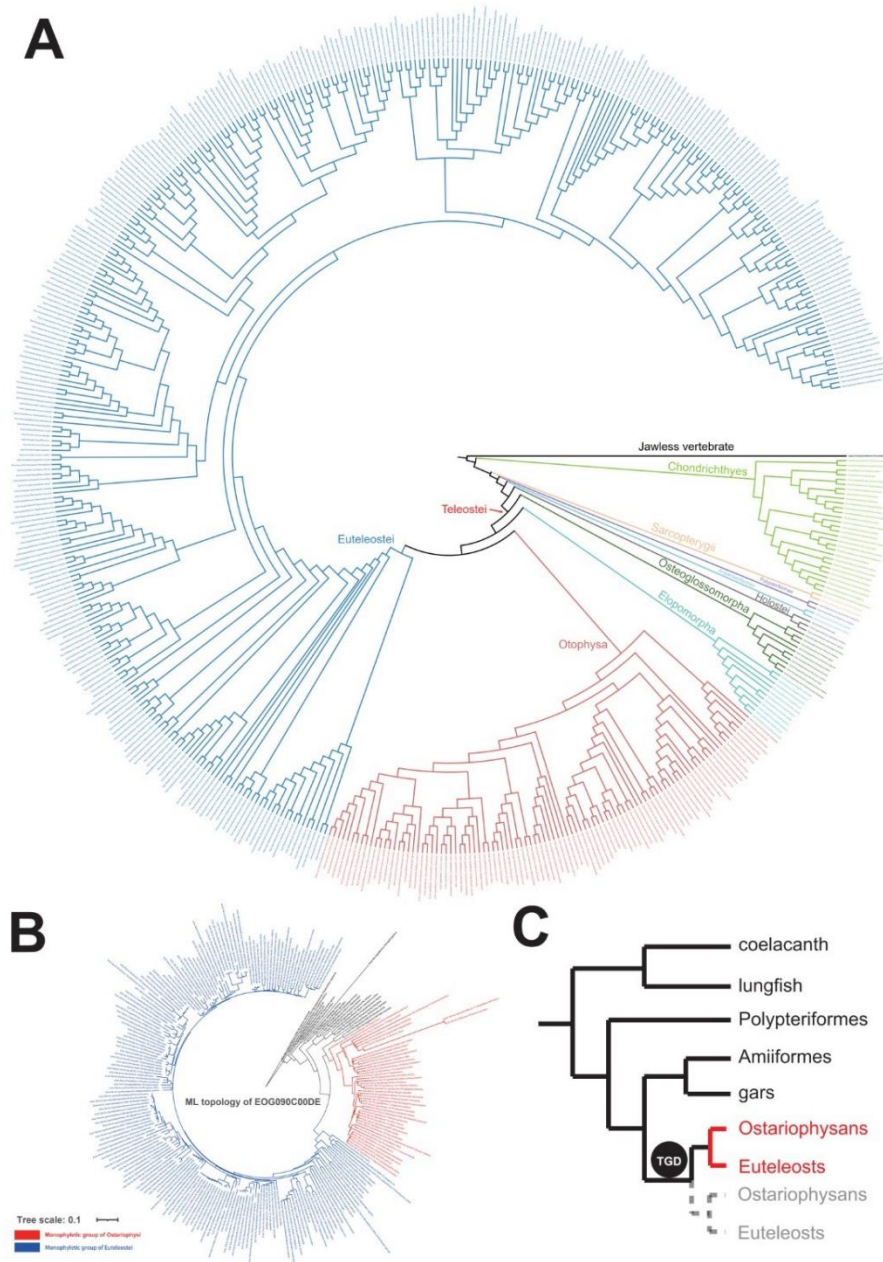


图 5.3. 鱼类物种的系统发育树。 (A) 使用 BUSCO 和 ASTRAL 软件分别生成了硬骨和软骨鱼类的物种树，并将其连接以生成一个该树。该树平均具有 97.38% 的 Bootstrap 值，表明其具有高度可靠的骨干系统发育。不同的颜色表示不同的分类群。(B) 通过 AU-test 并保留下来生成多基因物种树的基因树示例。受约束的树被用来进行旁系同源基因在硬骨鱼中由 TGD 事件产

生的拓扑结构测试，通过比较受约束的树和每个单独的非受约束 ML 基因树。这两棵树用 AU 测试进行比较。如果受约束的拓扑结构被拒绝（AU 测试 $P > 0.05$ ），这表明相对于另一个组是旁系同源基因。在 AU-test 中，1,009 个基因树的拓扑结构被约束，明显优于非约束方案（ $P < 0.05$ ），因此被视为后续分析的候选基因。在这个例子中，EOG090C00DE 的基因树将所有硬骨鱼分为两个单系群（Ostariophysi 和 Euteleostei）。（C）鱼类 TGD 事件的示意图。使用两个单系群（Ostariophysi 和 Euteleostei）来筛选出由 teleosts 中全基因组复制事件产生的潜在旁系同源基因的影响。

5.2.2 骨鳉鱼类的进化关系

骨鳉鱼类是一个淡水鱼支系，包括约 30% 的鱼类物种，其系统发育和分化过程仍存在争议[12, 37, 38]。耳鱼纲包括四个目：鲤形目、脂鲤目、电鳗目和鲇形目。我们基于共生树（图 5.4）以及之前的分子和形态学研究[13, 36]结论，认为鲤形目与其他三个目之间的姊妹关系应予支持。尽管之前的研究表明脂鲤目分为两个分支[39-43]，但我们推断的系统发生支持脂鲤目为单系类群，是电鳗目和鲇形目的祖先分支（更详细的物种系统发育树见图 5.3A）。

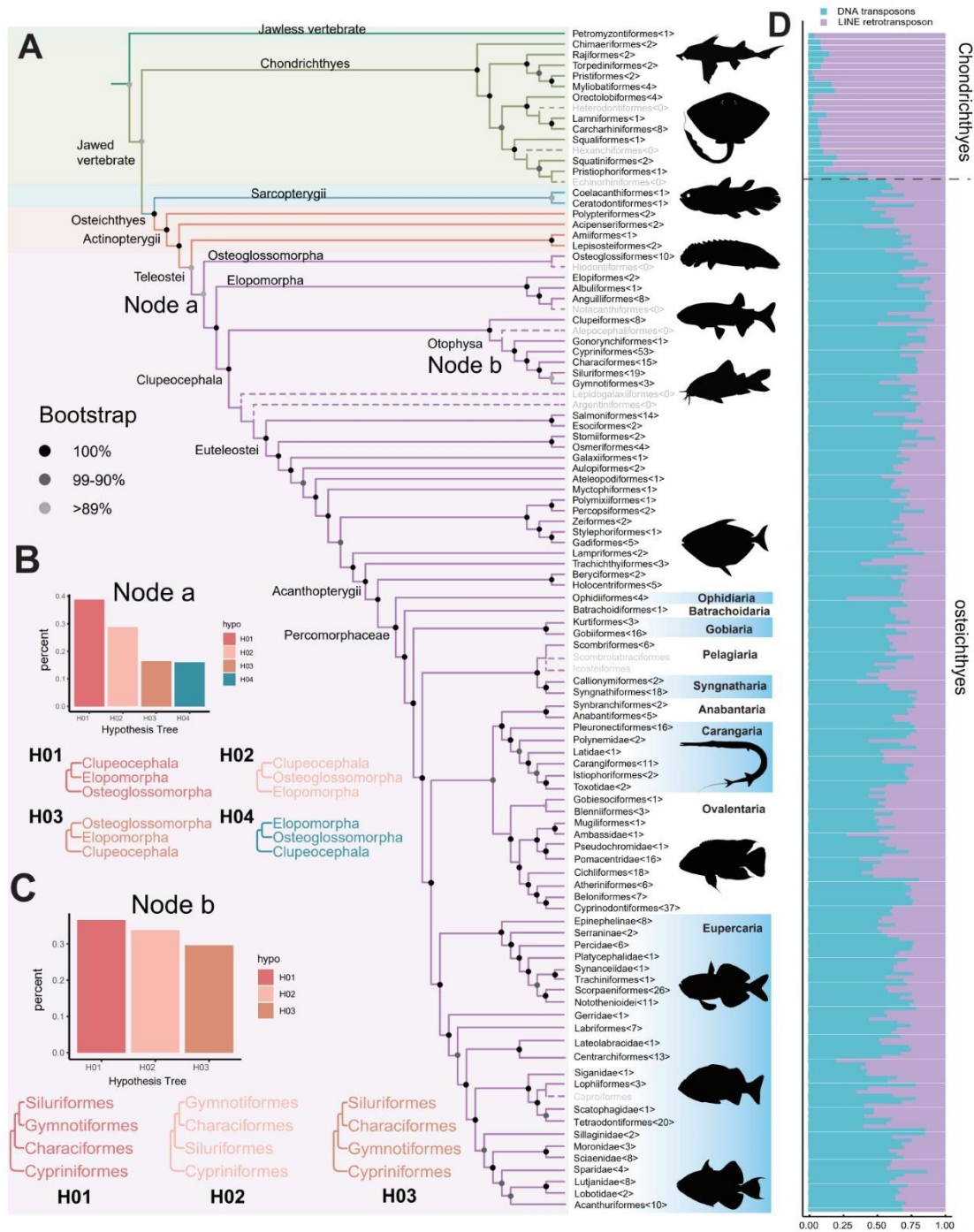


图 5.4. 现存鱼类目级别的系统发育和基因组特征。(A) 基于 4,584 个 Actinopterygii (硬骨鱼) 和 2,586 个脊椎动物 (软骨鱼) 单拷贝 BUSCO 基因的 Coalescent 最大似然物种树, 包含 533 个鱼类物种。实线表示已测序核基因组的 71 个鱼类目, 虚线显示尚未测序的 11 个目的推断位置。节点 a 和 b 是两个具有争议的进化关系的谱系。(B) 基于 Actinopterygii 的 4,584 个单拷贝 BUSCO 基因的基因谱系研究 (GGI) 分析支持 Osteoglossomorpha 作为 Clupeocephala 和

Elopomorpha 的姐妹群（拓扑结构 H01），与（A）中的 ML 树一致。（C）基因谱系研究（GGI）分析表明 Cypriniformes 是 Characiformes + Siluriformes + Gymnotiformes 的姐妹群，而 Siluriformes 是 Gymnotiformes 的姐妹群（拓扑结构 H01），与（A）中的 ML 树一致。（D）在软骨鱼类中，LINEs 是主要的 TE 成分（蓝色），而在硬骨鱼类中，DNA 转座子是主要的 TE 类型（粉色）。

5.3 鱼类中独特的转座元件分布

转座元件（Transposable elements, TE）的比例通常被认为是影响基因组大小变化的主要因素[44]。通过数据统计表明硬骨鱼类和软骨鱼类基因组大小的中位数分别为~755.05 Mb 和~2.99 Gb。我们观察到硬骨鱼类和软骨鱼类基因组大小的显著差异，这主要与 TE 的含量和多样性有关。为了进一步评估 TE 的多样性，我们对 533 个物种的数据集进行了分层取样，得到了收集到的鱼类目水平的 320 个代表性物种的基因组组装序列。在脊椎动物基因组中，LINE 类型的转座子数量比例最多，其中大部分是古老的 CR1 和 L2 家族[44]。我们发现软骨鱼类的 LINE/CR1 比例明显更高（图 5.5）。LINE/L2 是硬骨鱼的优势家族，而硬骨鱼的 LINE/CR1 比例较低。我们还观察到，基部辐鳍鱼和部分骨舌鱼类中的两个物种的 LINE/CR1 含量显著高于其他真骨鱼类（图 5.5）。这些古老的 TE 在基部辐鳍鱼上的分布与鳗鲡和斑马鱼等现代真骨鱼类的分布截然不同，但其如何在一些基部辐鳍鱼和骨舌鱼目物种中保留下来，值得进一步研究。

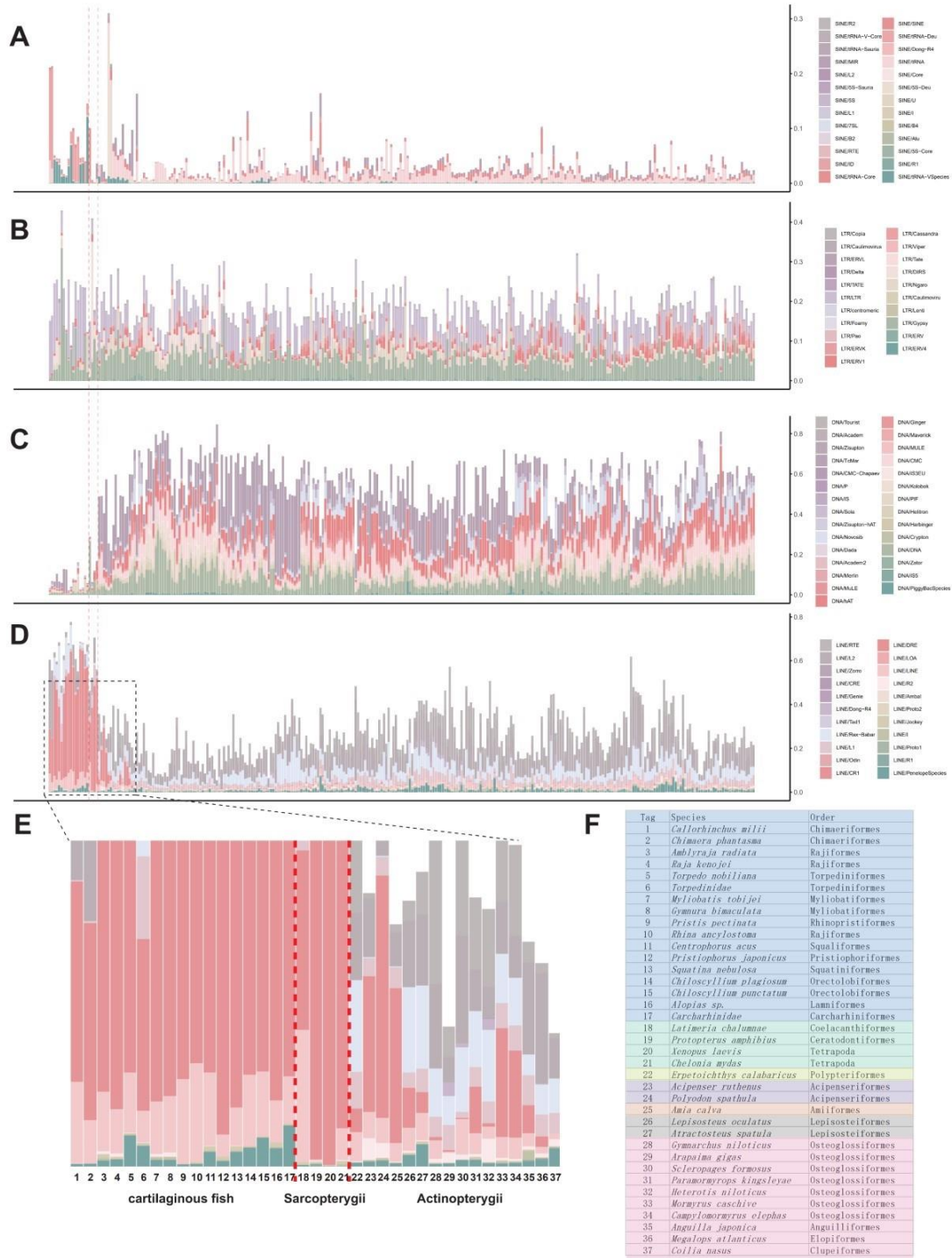


图 5.5. 鱼类转座元件的多样性。(A) SINE; (B) LTR; (C) DNA 转座子; (D) LINE; (E) 对 (D) 子图的放大视图。使用红色虚线矩形指示 Sarcopterygii 中的物种 (腔棘鱼、肺鱼和四足动物); (F) 图 E 中 x 轴上编号的物种概述。各种分类群用彩色编码。

5.4 真骨鱼特异保守调控元件

通过 Fish10K 提供的大量基因组数据集，使我们能够更准确地检测鱼类基因组中的高保守元件（highly conserved elements, HCEs）。目前已知多种类型的 HCEs，包括超保守元件（UCEs）、保守非编码元件（CNEs）和保守非外显子元件（CNEEs）；后者与 CNEs 相似，但排除了任何外显子区[45]。CNEs/CNEEs 通常聚集在基因附近或基因内，在进化过程中有助于形态的多样性[46-48]。例如，最近的一项研究发现，小鼠圆锥动脉形成所必需的一个增强子在新鳍亚纲鱼类中缺失[2]。我们发现了 549 个 HCEs，这些 HCEs 在 Fish10K 第一阶段使用的所有鱼类中普遍保守，我们还发现了鱼类特异的 HCEs，包括 346 个辐鳍鱼类特异的 HCEs 和 110 个真骨鱼类特异的 HCEs。这些 HCEs 大多极有可能在物种进化中发挥关键作用，因为一旦祖先物种获得了新的关键功能，它往往会以 HCEs 的形式保守地传递给后来的分化物种[46]。为了说明这些 HCEs 的功能作用，我们重点关注与真骨鱼的独特器官演化有关的 HCE。

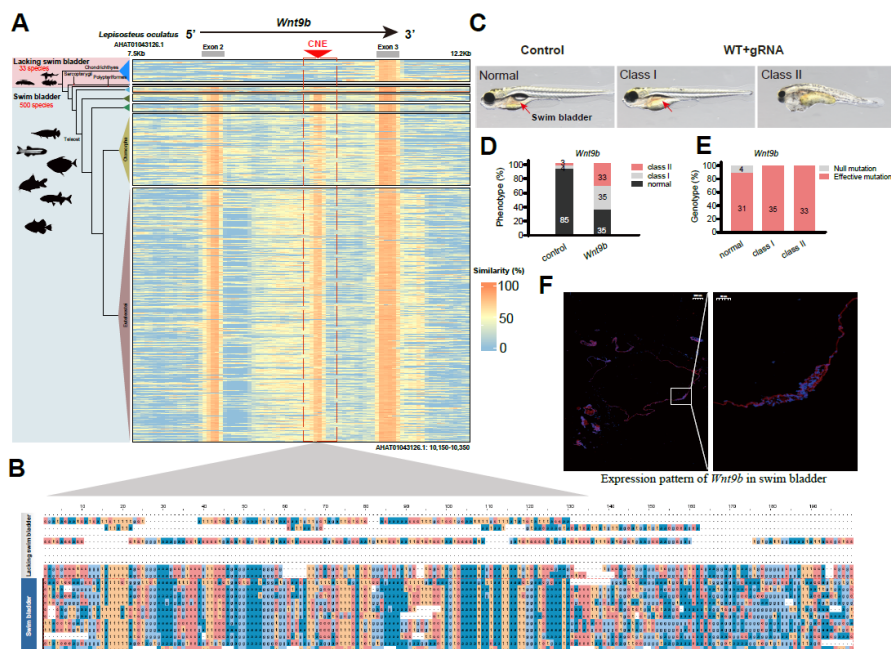


图 5.6. 鱼鳔发育相关的保守非编码元件。（A）使用 *Lepisosteus oculatus*（斑点雀鳢；版本号 LepOcu1）作为参考，对 533 个物种中的 Wnt9b 位点进行序列保守热图。红色箭头和虚线框表

示所有具有鱼鳔的硬骨鱼物种中 *Wnt9b* 内含子 2 中的保守非编码元件 (CNE)。(B) swim bladder-specific CNE 的序列比对。(C) 将三个 CRISPR-Cas9 向导 RNA 注射到一龄胚胎中靶向到 swim bladder-specific CNE 序列后获得的幼年斑马鱼表型。在受精后 120 小时 (hpf) 评估表型。正常, 野生型; I 类, 形状正常但缺乏鱼鳔; II 类, 严重畸形。(D) 堆叠条形图显示 120 hpf 时胚胎的鳔表型百分比 (见 C)。(E) 堆叠条形图显示 120 hpf 时胚胎的鳔基因型百分比。(F) 左侧图像显示正常斑马鱼 (野生型) 鳔的荧光显微镜图像 120 hpf。红色, 用 FITC-TSA 扩增的 DIG 标记的 *Wnt9b* 探针; 蓝色, DAPI (比例尺=200 μ m)。右侧图像是使用更高倍数放大观察相同组织的显微图像 (比例尺=50 μ m)。

在脊椎动物诞生之初, 鳃器官已经成为鱼类获取氧气主要的呼吸器官。硬骨鱼进化出了肺或鱼鳔, 从而使一部分物种获得了空气呼吸的能力[2, 5, 49]。例如, 陆生动物和肉鳍鱼类通过肺呼吸以及淡水中的一些原始辐鳍鱼类利用鱼鳔进行空气交换 (如弓鳍鱼) [50, 51]; 但大多数真骨鱼类的鱼鳔在进化过程中仅用于维持浮力, 在某些物种中还充当感觉器官[52]。我们发现三个 CNE 可能与鱼鳔的功能有关。这三个 CNE 中的一个位于 *Wnt9b* 的内含子, 在所有有鱼鳔的鱼类中都是保守的 (图 5.6)。我们在斑马鱼鳔的单细胞 RNA-seq 数据集[53] 或其他组织器官的 RNA-seq 数据集中都没有检测到该 CNE, 而且该 CNE 也不包含保守的开放阅读框, 表明该 CNE 并不是 *Wnt9b* 外显子。以前的研究表明, *Wnt9b* 通路对鱼鳔的发育非常重要 [54-56]。此外, 原位杂交实验也证实 *Wnt9b* 在斑马鱼的鱼鳔中表达 (图 5.6)。增强子通常存在于基因间和基因内区域[57]。因此, 我们推测 *Wnt9b*-CNE 的中断可能会破坏鱼鳔的发育或功能。为了验证这一点, 我们使用 CRISPR-Cas9 基因组编辑技术修改了斑马鱼 *Wnt9b* 三个位置的 CNE 序列。120hpf 的表型检查结果显示, 大多数突变的幼年斑马鱼表现出鳔缺失的先天性缺陷。在这些鱼鳔缺失突变体中, 约 33.98%发育为 I 类, 即鱼鳔缺失但是外观正常。我们还观察到约 32.04%的突变体发育为 II 类, 即鱼鳔缺失和畸形, 包括体型缩短和脊柱向腹部弯曲。通过测序发现, 所有幼体斑马鱼都被有效敲除 (图 5.6)。因此, 实验结果提示该段 CNE 是与鱼鳔的发育有关的重要调控元件。

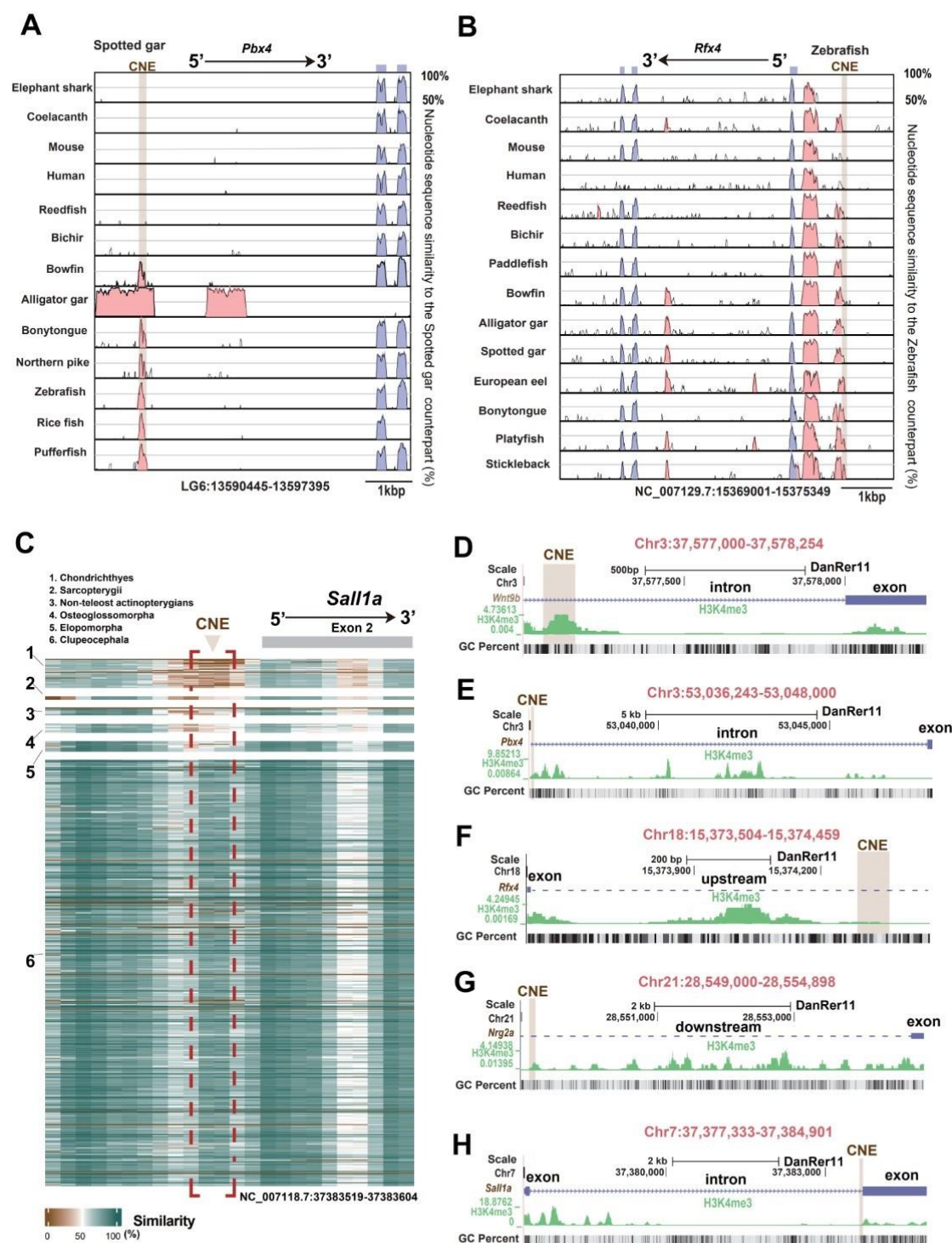


图 5.7. 鉴定到的特定谱系保守非编码元件。(A-B) 不同代表性物种中 *Pbx4* 和 *Rfx4* 基因座的 VISTA 序列保守图，分别以斑点雀鳝 (*Lepisosteus oculatus*, 组装版本 LepOcu1) 和斑马鱼

(*Danio rerio*, 组装版本 GRCz11) 为参考。图中突出显示的黄色柱状图代表 CNE 的位置。每个窗口中的曲线表示与参考基因组的一致性值。如果一致性值大于 70%, 则曲线以颜色填充。外显子以蓝色填充, 内含子以红色填充。(C) 533 个物种中附近 *Sall1a* 基因的 CNE, 以斑马鱼 (*Danio rerio*, 组装版本 GRCz11) 为参考的热图。图中箭头表示 CNE 的位置。(D-H) 序列分析表明, 鉴定的 CNE 与 DANIO-CODE 数据库和 UCSC 浏览器中的 H3K4me3 的 ChIP-seq 数据峰值区域重叠。

此外, 我们还发现另外两个 CNE 与关联基因有很强的相互作用, 一个位于 *Pbx4* 的基因间区, 另一个位于 *Rfx4* 的上游 (图 5.7)。序列分析表明, 这两个 CNE 与组蛋白修饰区域重叠。我们根据 UCSC 浏览器的 H3K4me3 功能[58] 检索了 *Pbx4*-CNE、*Rfx4*-CNE 相关区域组蛋白标签的富集水平, 发现这些 CNE 与 DANIO-CODE 数据库[59]中 ChIP-seq 数据得到的峰值区域重叠 (图 5.7)。据报道, 组蛋白 H3K4me3 与转录起始、延伸和 RNA 剪接等多种过程有关[60]。这类组蛋白修饰通常会激活基因表达[61]。与转录活性相关的修饰存在于多种生物体中[62-64]。根据 H3K4me3 数据分析, 我们认为这些 CNE 可能通过影响基因启动子来调控其附近基因的表达。先前的研究表明, *Rfx4* 基因突变会导致鱼鳔无法充气[65], 敲除 *Pbx4* 基因会导致鱼鳔发育异常[66]。因此, 一系列特异性调控元件的出现有可能促进了鱼鳔的进化。通过控制鱼鳔的充放气, 鱼体产生了不同的浮力, 这有助于真骨鱼在不同的水层中停留, 扩大了鱼类在水体中的活动范围。

5.5 鱼类进化过程中新基因的出现

新基因指在某一品系或物种进化过程中独特出现的基因 (即可能不是最近才出现的), 通常包括通过 *de novo* 方式 (从以前的非基因或内含子区域)、DNA 复制 (从基因片段到染色体或整个基因组) 以及 RNA (如逆转录转座) 产生的基因[67-69]。真骨鱼的特征是其祖先在大约 300 Mya (285.8-333.0 Mya) 发生了全基因组复制, 随后在二叠纪从祖先辐鳍鱼类开始了爆炸性辐射[5, 70], 导致它们成功适应了从深海到高原淡水池塘的各种水生生态系统。一些真骨鱼类也经历了第四轮基因组复制 (例如鲑科鱼类, 约 54 Mya) [70]。全基因组复制事件产生了大量新基因, 使自然

选择得以发挥作用。为了了解在这一节点上进化的新基因的来源和作用，本研究在鱼类系统发育的关键分支上确定了远缘物种基因的起源时间。

根据本研究中最新的鱼类系统发生树，在现代真骨鱼我们发现了 709 个新基因（包括 255 个 DNA 介导的基因、5 个 RNA 介导的基因和 62 个新基因）。在基部真骨鱼通过 DNA 介导产生的新基因中，74.51%（190/255）属于三刺鱼的 TGD 同源区块，而三刺鱼的谱系没有经历过额外的全基因组复制事件。因此，这 190 个基因是在约 300 Mya 的全基因组复制事件中出现的。有观点认为，全基因组复制事件可能需要数千万年才能显现 [71]。对于真骨鱼类来说，这意味着相关的重复基因组在物种分化事件发生很久之后才开始多样化[46, 71]。

对这些新基因的功能富集分析，发现它们在心脏发育相关和骨骼相关的 GO 通路（如心肌肥大的调节（GO:0010611）、正调控骨矿化（GO:0030501）、调控骨矿化（GO:0030500）、调控骨重塑（GO:0046850））中有显著的富集，这表明与真骨鱼特有的特殊心脏（如球动脉，维持血液向鳃毛细血管的稳定动脉压）和骨骼（部分或完全骨化）相关的基因开始出现在真骨鱼的根部，为它们形成适应不同环境的独特循环系统和更精细的骨骼结构奠定了遗传基础（图 5.8）。我们还发现，在骨舌鱼类中出现了一些不同其他分支的特定基因功能，如与骨骼骨化和成骨细胞发育相关的独特 GO 特征（GO:0045778, GO:0033688, GO:0033690）。以往的研究表明，狮子鱼特定的骨骼结构会影响其对环境的适应性[28]。在我们的研究中，总共发现了 30 个与骨骼相关的基因新出现在原始真骨鱼中，提示这些新产生的骨发育基因促进了演化过程中骨骼的形成。在这些新产生的骨骼相关基因中，编码跨膜蛋白的 *Tmem119a* 和 *Tmem64* 基因已被报道与骨化有关[72, 73]。此外，*Atraid* 基因也被报道与小鼠的骨骼发育有关[74]。

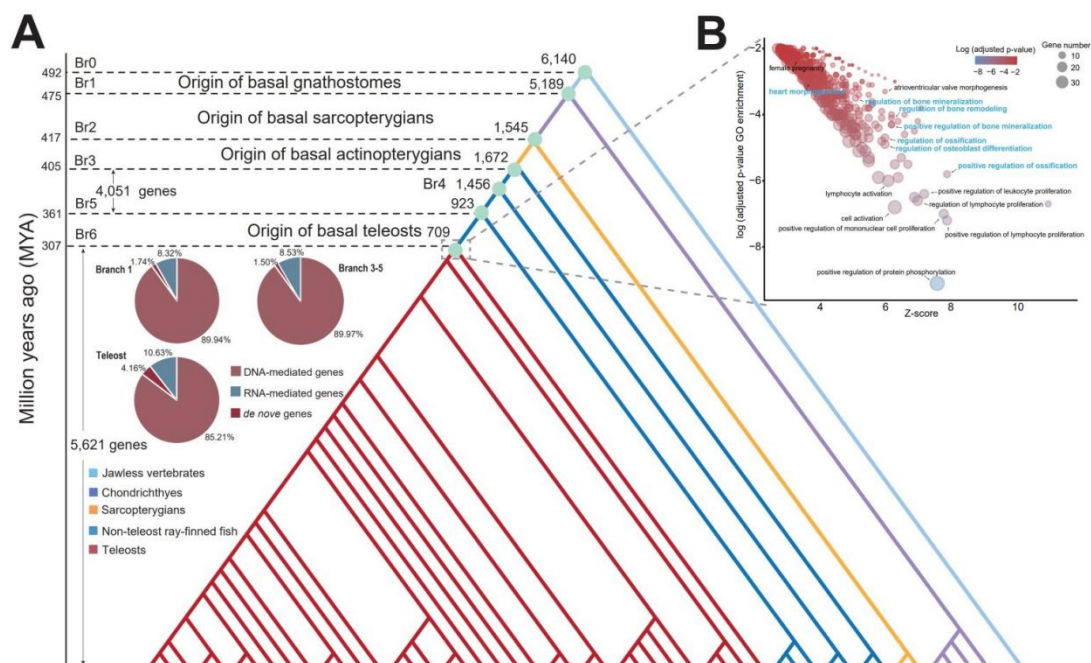


图 5.8. 硬骨鱼新基因的演化。（A）三刺鱼（*Gasterosteus aculeatus*）蛋白编码基因的年龄分布。物种之间的分歧时间是数百万年。在鱼类演化的关键时间点上，有六个不同的分支（Br0-6）。使用 49 个鱼类物种和 *G. aculeatus* 作为参考物种预测了新基因。预测了三种新的基因类别：*de novo*、DNA 介导和 RNA 介导。在各个节点显示新基因的总数。饼图中显示预测的新基因类别的比例。多样性的时间从 Near 等（2012）和 Cheng 等（2021）研究中获得[78, 88]。（B）功能富集分析表明，在硬骨鱼（Br6）中产生的新基因主要涉及心脏发育相关和骨骼重建（GO terms 以粗体显示）。

5.6 讨论与展望

自 4.3 亿年前祖先鱼类出现以来，真骨鱼已经进化出了丰富的生物多样性，并占据几乎所有水生栖息地。在本研究中，我们报告了利用 Fish10K 第一阶段产生的丰富数据资源，建立了目前覆盖最多目水平的鱼类系统发育树，并探究了鱼类学研究之前存在的拓扑学争议。通过比较基因组学分析，我们鉴定到了可能有助于鱼类适应性进化的调控元件和基因，如鱼鳔和胸鳍的发育。同时发现进化过程中承受选择性压力的调控元件经常出现在器官发育相关调控基因的附近，而这些元件之间的上下游调控关系已在其他研究中得以证实[75-77]。功能分析显示，一部分 CNE 起着顺式调控元件的作用，如驱动相关基因组织特异性表达的增强子[78, 79]。已

发现的 CNE 揭示了这些调节因子在真骨鱼性状变化和遗传创新中的作用。尽管 Fish10K 第一阶段已经取得了一些进展，但面对丰富多彩的鱼类生物资源多样性，未来还需要开展广泛的基因组学研究。

鱼类作为地球上最古老的脊椎动物，拥有多样化的生态习性和生理功能，研究其基因组对于理解生物演化、人类健康以及水生生态系统的维护具有重要意义。万种鱼基因组计划，简称 Fish10K（Fish 10,000 Genomes Project）于 2019 年 9 月 21 日在“2019 年国际海洋基因组会议（ICG-Ocean 2019）”上正式对外发布，由青岛华大基因研究院联合中国科学院水生生物研究所、西北工业大学、中国科学院海洋研究所等多家单位共同发起，旨在绘制万种代表性鱼类基因组图谱，涵盖世界各地鱼类的所有目和代表性科，建立一个大规模，高质量的鱼类基因组数据库，聚焦鱼类基因组研究。该项目计划将通过 10 年的时间，以开放的国际联盟运行形式，致力于推进 10,000 种鱼类基因组的研究，拟解决鱼类的一系列生物学问题，包括鱼类的系统发育、起源与大陆板块漂移事件；鱼类基因组复制、基因组大小与染色体进化的关系；极端环境鱼类适应性进化；鱼类体型及特殊器官发育形成；重要进化节点鱼类基因组染色体构象与进化等重要科学问题。该计划期待国内及国际合作，欢迎来自不同地区的科学家能够共同参与鱼类基因组研究、积极推进资源共享、技术互助，加强科学交流，为鱼类的物种多样性保护和经济鱼类的育种提供分子基础，同时为水产养殖和海洋生态保护及进化研究提供重要基础。

本章作者：

宋跃、袁增宝、张蘇予、刘美如、于丽丽（青岛华大基因研究院）

本章参考文献

1. R. Fricke, W. Eschmeyer, R. Van der Laan, Catalog of fishes: genera, species, references. *California Academy of Sciences, San Francisco, CA, USA*, (2018).
2. X. Bi *et al.*, Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes. *Cell* **184**, 1377-1391. e1314 (2021).
3. A. Meyer *et al.*, Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**, 284-289 (2021).
4. K. Wang *et al.*, African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**, 1362-1376. e1318 (2021).
5. D. Davesne *et al.*, Fossilized cell structures identify an ancient origin for the teleost whole-genome duplication. *Proceedings of the National Academy of Sciences* **118**, e2101780118 (2021).
6. S. Hao *et al.*, African Arowana genome provides insights on ancient teleost evolution. *Science* **23**, 101662 (2020).
7. B. Star *et al.*, The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207-210 (2011).
8. B. Venkatesh *et al.*, Elephant shark genome provides unique insights into gnathostome evolution. **505**, 174 (2014).
9. K. Reichwald *et al.*, Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell* **163**, 1527-1538 (2015).
10. Z. Liu *et al.*, The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nature communications* **7**, 1-13 (2016).
11. W. C. Warren *et al.*, Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nature ecology & evolution* **2**, 669-679 (2018).
12. J. S. Nelson, T. C. Grande, M. V. Wilson, *Fishes of the World*. (John Wiley & Sons, 2016).
13. R. Betancur-R *et al.*, Phylogenetic classification of bony fishes. *BMC evolutionary biology* **17**, 1-40 (2017).
14. Q. Lin *et al.*, The seahorse genome and the evolution of its specialized morphology. *Nature* **540**, 395-399 (2016).
15. P. Xu *et al.*, Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature genetics* **46**, 1212-1219 (2014).
16. A. Ali, G. H. Thorgaard, M. Salem, PacBio Iso-Seq improves the rainbow trout genome annotation and identifies alternative splicing associated with economically important phenotypes. *Frontiers in Genetics* **12**, (2021).
17. L. Zhao *et al.*, Chromosome-Level Genome Assembly and Annotation of a Sciaenid Fish, *Argyrosomus japonicus*. *Genome biology and evolution* **13**, evaa246 (2021).
18. B. Chen *et al.*, The sequencing and de novo assembly of the *Larimichthys crocea* genome using PacBio and Hi-C technologies. *Scientific data* **6**, 1-10 (2019).

19. S. Lien *et al.*, The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200-205 (2016).
20. S. Feng *et al.*, Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252-257 (2020).
21. G. Zhang *et al.*, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311-1320 (2014).
22. K.-P. Koepfli, B. Paten, G. K. C. o. Scientists, S. J. OBrien, The Genome 10K Project: a way forward. *Annu. Rev. Anim. Biosci.* **3**, 57-111 (2015).
23. H. A. Lewin *et al.* (National Acad Sciences, 2022), vol. 119, pp. e2115635118.
24. K. Du *et al.*, The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nature ecology & evolution* **4**, 841-852 (2020).
25. S. Nowoshilow *et al.*, The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50-55 (2018).
26. D. E. Soltis, P. S. Soltis, Polyploidy: recurrent formation and genome evolution. *Trends in ecology & evolution* **14**, 348-352 (1999).
27. L. Ma *et al.*, A hypomorphic cystathionine β -synthase gene contributes to cavefish eye loss by disrupting optic vasculature. *Nature communications* **11**, 1-15 (2020).
28. K. Wang *et al.*, Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation. *Nature ecology & evolution* **3**, 823-833 (2019).
29. G. Fan *et al.*, Fish10K Pilot Project Data GigaScience Database. *GigaScience Database*, (2020).
30. O. Wang *et al.*, Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome research* **29**, 798-808 (2019).
31. R. Betancur-R *et al.*, Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes. *Evolution* **73**, 329-345 (2019).
32. R. M. Waterhouse *et al.*, BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**, (2017).
33. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. J. B. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. **31**, 3210-3212 (2015).
34. K. Khalturin *et al.*, Polyzoa is back: The effect of complete gene sets on the placement of Ectoprocta and Entoprocta. *Sci Adv* **8**, eabo4400 (2022).
35. D. Arcila *et al.*, Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nat Ecol Evol* **1**, 20 (2017).
36. L. C. Hughes *et al.*, Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences* **115**, 6249-6254 (2018).
37. R. Fricke, W. Eschmeyer, R. Van der Laan, Catalog of fishes: genera, species, references. *California Academy of Sciences, San Francisco, CA, USA* <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>, (2018).

38. R. Van Der Laan, W. N. Eschmeyer, R. Fricke, Family-group names of recent fishes. *Zootaxa* **3882**, 1–230–231–230 (2014).
39. P. Chakrabarty *et al.*, Phylogenomic systematics of ostariophysan fishes: ultraconserved elements support the surprising non-monophyly of Characiformes. *Systematic Biology* **66**, 881–895 (2017).
40. W. J. Chen, S. Lavoué, R. L. Mayden, Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). *Evolution* **67**, 2218–2239 (2013).
41. M. Nakatani, M. Miya, K. Mabuchi, K. Saitoh, M. Nishida, Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaeon origin and Mesozoic radiation. *BMC evolutionary biology* **11**, 1–25 (2011).
42. S. V. Fink, W. L. Fink, Interrelationships of the ostariophysan fishes (Teleostei). *Zoological Journal of the Linnean Society* **72**, 297–353 (1981).
43. J. A. Alves-Gomes, The mitochondrial phylogeny of the South American electric fish (Gymnotiformes) and an alternative hypothesis for the otophysan historical biogeography. *Gonorynchiformes and ostriophysan relationships: A comprehensive review*, 517–565 (2010).
44. C. G. Sotero-Caio, R. N. Platt, A. Suh, D. A. Ray, Evolution and diversity of transposable elements in vertebrate genomes. *Genome biology and evolution* **9**, 161–177 (2017).
45. S. V. Edwards, A. Cloutier, A. J. Baker, Conserved Nonexonic Elements: A Novel Class of Marker for Phylogenomics. *Syst Biol* **66**, 1028–1044 (2017).
46. V. Ravi, B. Venkatesh, The Divergent Genomes of Teleosts. *Annu Rev Anim Biosci* **6**, 47–68 (2018).
47. D. Polychronopoulos, J. W. D. King, A. J. Nash, G. Tan, B. Lenhard, Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res* **45**, 12611–12624 (2017).
48. T. B. Sackton *et al.*, Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
49. S. Longo, M. Riccio, A. R. McCune, Homology of lungs and gas bladders: insights from arterial vasculature. *Journal of Morphology* **274**, 687–703 (2013).
50. A. W. Thompson *et al.*, The bowfin genome illuminates the developmental evolution of ray-finned fishes. *Nat Genet* **53**, 1373–1384 (2021).
51. C. Farmer, Evolution of the vertebrate cardio-pulmonary system. *Annual review of physiology* **61**, 573–592 (1999).
52. B. Pelster, Using the swimbladder as a respiratory organ and/or a buoyancy structure-Benefits and consequences. *J Exp Zool A Ecol Integr Physiol* **335**, 831–842 (2021).
53. Jiang, Mengmeng, et al. "Characterization of the zebrafish cell landscape at single-cell resolution." *Frontiers in cell and developmental biology* **9** (2021): 743421
54. Y. Wu, W. Li, M. Yuan, X. Liu, The synthetic pyrethroid deltamethrin impairs zebrafish (Danio rerio) swim bladder development. *Science of the Total Environment* **701**, 134870 (2020).
55. Y. Yang *et al.*, Comparative transcriptome analysis reveals conserved branching morphogenesis related genes involved in chamber formation of catfish swimbladder. *Physiol Genomics* **50**, 67–76 (2018).

56. A. Yin, S. Korzh, C. L. Winata, V. Korzh, Z. Gong, Wnt signaling is required for early development of zebrafish swimbladder. *PLoS One* **6**, e18431 (2011).
57. A. Panigrahi, B. W. O'Malley, Mechanisms of enhancer action: the known and the unknown. *Genome Biol* **22**, 108 (2021).
58. J. Navarro Gonzalez *et al.*, The UCSC genome browser database: 2021 update. *Nucleic Acids Research* **49**, D1046-D1057 (2021).
59. H. Tan, D. Onichtchouk, C. Winata, DANIO-CODE: toward an encyclopedia of DNA elements in zebrafish. *Zebrafish* **13**, 54-60 (2016).
60. J. R. Davie, W. Xu, G. P. Delcuve, Histone H3K4 trimethylation: dynamic interplay with pre-mRNA splicing. *Biochemistry and Cell Biology* **94**, 1-11 (2015).
61. A. Barski *et al.*, High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).
62. B. E. Collins, C. B. Greer, B. C. Coleman, J. D. Sweatt, Histone H3 lysine K4 methylation and its role in learning and memory. *Epigenetics & chromatin* **12**, 1-16 (2019).
63. S. Park, G. W. Kim, S. H. Kwon, J. S. Lee, Broad domains of histone H3 lysine 4 trimethylation in transcriptional regulation and disease. *The FEBS journal* **287**, 2891-2902 (2020).
64. Y. Ding, M. Fromm, Z. Avramova, Multiple exposures to drought'train'transcriptional responses in Arabidopsis. *Nature communications* **3**, 1-9 (2012).
65. S. Irina *et al.*, Zebrafish Rfx4 controls dorsal and ventral midline formation in the neural tube. *Developmental dynamics : ,* 247 (2018).
66. P. H. Teoh, A. C. Shu-Chien, W. K. Chan, Pbx1 is essential for growth of zebrafish swim bladder. *Developmental dynamics: an official publication of the American Association of Anatomists* **239**, 865-874 (2010).
67. H. Kaessmann, Origins, evolution, and phenotypic impact of new genes. *Genome research* **20**, 1313-1326 (2010).
68. H. Li *et al.*, Pattern of New Gene Origination in a Special Fish Lineage, the Flatfishes. *Genes* **12**, 1819 (2021).
69. Y. Ding, Q. Zhou, W. Wang, Origins of new genes and evolution of their novel functions. *Annual Review of Ecology, Evolution and Systematics* **43**, 2012 (2012).
70. T. J. Near *et al.*, Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences* **109**, 13698-13703 (2012).
71. F. M. Robertson *et al.*, Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol* **18**, 111 (2017).
72. K.-i. Tanaka *et al.*, Interaction of Tmem119 and the bone morphogenetic protein pathway in the commitment of myoblastic into osteoblastic cells. *Bone* **51**, 158-167 (2012).
73. H. Kim *et al.*, Tmem64 modulates calcium signaling during RANKL-mediated osteoclast differentiation. *Cell metabolism* **17**, 249-260 (2013).
74. L. E. Surface *et al.*, ATRAID regulates the action of nitrogen-containing bisphosphonates on bone. *Science translational medicine* **12**, eaav9166 (2020).

- 75. P. J. Wittkopp, B. K. Haerum, A. G. Clark, Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature genetics* **40**, 346-350 (2008).
- 76. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36 (2008).
- 77. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**, 206-216 (2007).
- 78. P. Navratilova *et al.*, Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. *Developmental biology* **327**, 526-540 (2009).
- 79. A. Visel *et al.*, Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature genetics* **40**, 158-160 (2008).

198个

海洋四足动物基因组

覆盖现存海洋四足动物物种的

17.55%



第六章 海洋四足动物

6.1 海洋四足动物简介

海洋四足动物包括海洋哺乳动物、海鸟、海洋爬行动物和海洋两栖动物，由约 27 个目组成，包含约 1,128 个物种或亚种（表 6.1）[1]。海洋四足动物由不同的陆地祖先进化而来[2]，不同类别之间具有显著的身体形态差异。它们依靠海洋生态环境生存，从海洋中获取大部分或全部营养物质，通常位于海洋食物链的顶端。除了进行交配、繁殖、蜕皮等一系列生命活动时需返回到岸边或沙滩，大部分时间生活在水中。

海洋哺乳动物不代表传统的系统进化分类单元，而仅仅是根据它们的生活环境而分类的一个群体，主要包含 3 个目：鲸目、海牛目和食肉目，包含大约 180 个物种或亚种，比如海狮、海豹、海豚和海牛（此处统计包含少数重要的淡水水生哺乳类，如白鱔豚等）（表 6.1）。在入水之后，不同的分支经历了相似的趋同进化。在海洋哺乳动物中，有一个比较特殊的物种——北极熊，它需要漂浮的海冰来维持生存。

海鸟也根据它们的生活环境进行的一个比较粗略的分类，因此不同海鸟之间区别很大。为了适应海洋环境，不同海鸟进化出了一些相似的特点。因此，海鸟的趋同进化也特别明显。海鸟由 19 个目组成，包含大约 841 物种或亚种（表 6.1）。在海鸟中，也有一个比较特殊的物种——企鹅，它极端依赖南极极端环境。

海洋爬行动物在中生代时期开始适应海洋环境[3]，主要包含 4 个目：蜥蜴目、鳄目、有鳞目和龟鳖目，由大约 106 个物种组成，如海龟、海蛇、海鬣蜥等（表 6.1）。

海洋两栖类大多数依赖淡水环境，仅有一个物种——食蟹蛙，又称海蛙（*Fejervarya cancrivora*）可以称为海洋两栖类。因为其生活在近海且通过特殊的渗透压调节机制，食蟹蛙可以承受高盐离子环境[4]。

表 6.1 海洋四足动物统计*

类别	目	科	属	种（亚种）
海洋哺乳动物（Mammalia）	3(3)	24(18)	73(36)	180(42)
海鸟（Aves）	19(18)	69(54)	276(102)	841(142)
海洋爬行动物（Reptilia）	4(3)	13(8)	49(12)	106(14)
海洋两栖动物（Amphibia）	1(0)	1(0)	1(0)	1(0)
总计	27(24)	107(80)	399(149)	1128(198)

*物种数据来源于 WoRMS 数据库，并经过人工筛选。已测序物种来源于 NCBI 和已发表的文章。表中括号外数字表示各分类阶元的现存数量；括号内表示已被全基因组测序的数量。

6.2 海洋四足动物基因组研究基本情况

尽管海洋四足动物具有重要的进化地位和生态价值，但是其基因组研究仍偏少。2008 年，科学家完成了第一个海洋四足动物宽吻海豚全基因组测序，但直到 2012 年才完成第二个海洋四足动物北极熊的全基因组测序[5]。2013 年，由于环境破坏被宣告功能性灭绝的物种白鱀豚和极端濒危物种绿海龟完成了全基因组测序[6, 7]，人们开始意识到采用全基因组数据进行物种保护的重要性。2014 年，鸟类基因组里程碑研究——48 种鸟类基因组学分析在《Science》以专刊形式发表，该研究包含 12 种海鸟，极大促进了海洋四足动物基因组学研究。近年来，随着 MGI stLFR、PacBio HiFi-CCS、Hi-C、Nanopore ultra-long 测序技术等测序技术的发展和成本的降低，海洋四足动物全基因组研究开始迅速发展，一些海洋四足动物得到了较高质量的基因组。2019 年，青岛华大基因研究院等机构利用 MGI stLFR 和 Hi-C 技术，获得了世界首个染色体水平的抹香鲸基因，进而研究了祖先染色体构建，染色体融合和断裂等染色体进化过程[8]。同济大学等研究机构利用 PacBio 三代测序技术、Hi-C 技术，获得了染色体级别的青环海蛇与平颞海蛇的基因组版本，在连续性上较之前版本提高了 197 倍，进而开展了基因组结构变异和比较毒素组学研究[9]。

已测序海洋四足动物物种数目每年都在稳定增长（图 6.1）。目前，共有 198 个物种/亚种被全基因组测序和组装，约占所有海洋四足动物的 17.55%（表 6.1 和图 6.1）。从基因组质量上看，只有 57 个物种（占 28.79%）被组装到染色体水平；仍有 52 个物种（占 26.26%）的 Scaffold N50 < 1 Mb，为基因组草图（图 6.2），如南极小须鲸[10]、侏露脊鲸[11]等，这些基因组组装的片段较短，将会影响后续的比较基因组学分析。因此，亟需利用新技术对基因组进行升级、提高参考基因组质量。此外，构建端粒到端粒（T2T）基因组逐渐成为研究热点，T2T 基因组具有高度准确性、连续性、完整性，但海洋四足动物尚无 T2T 水平的基因组发表；同时，泛基因组（Pangenome）可以覆盖更多的变异，但海洋四足动物也尚未发表过泛基因组相关研究。

同时需要注意的是，海洋四足动物已被基因组测序的物种仍然偏少，还有 249 个属（占 62.66%）未有物种被测序，亟需获得至少一个物种的完整基因组序列来保存该分支的遗传信息。因此，在基因组的数量和质量上来看，海洋四足类研究仍处于较低水平。根据现有组装完成的基因组统计，海洋哺乳类基因组大小为 2.01-3.16Gb，平均为 2.50 Gb；海鸟类基因组大小为 0.94-1.46 Gb，平均为 1.18Gb；海洋两栖动物基因组大小为 1.26-2.35 Gb，平均为 1.90 Gb。整体来讲，海洋四足动物的基因组大小呈现相对恒定的趋势，同时，GC 含量变化不大，平均为 41.95%。

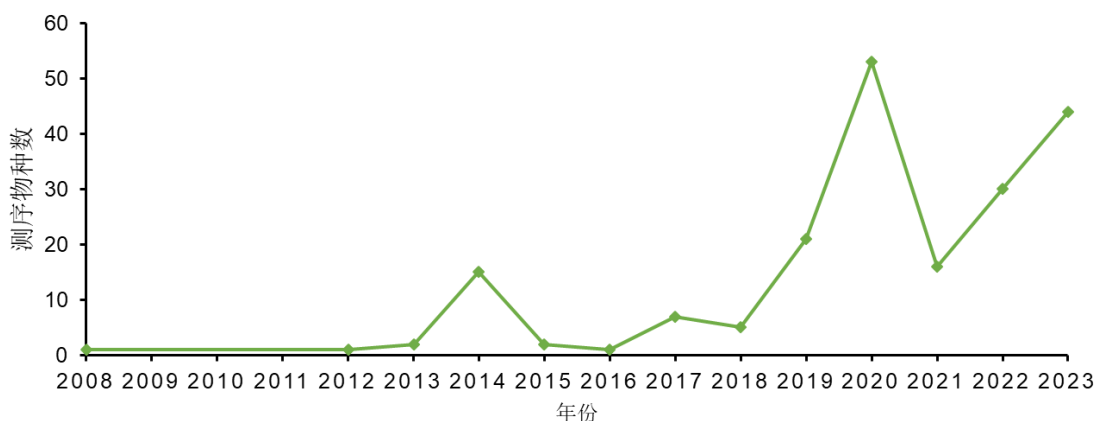


图 6.1 不同年份已发表基因组的物种数统计

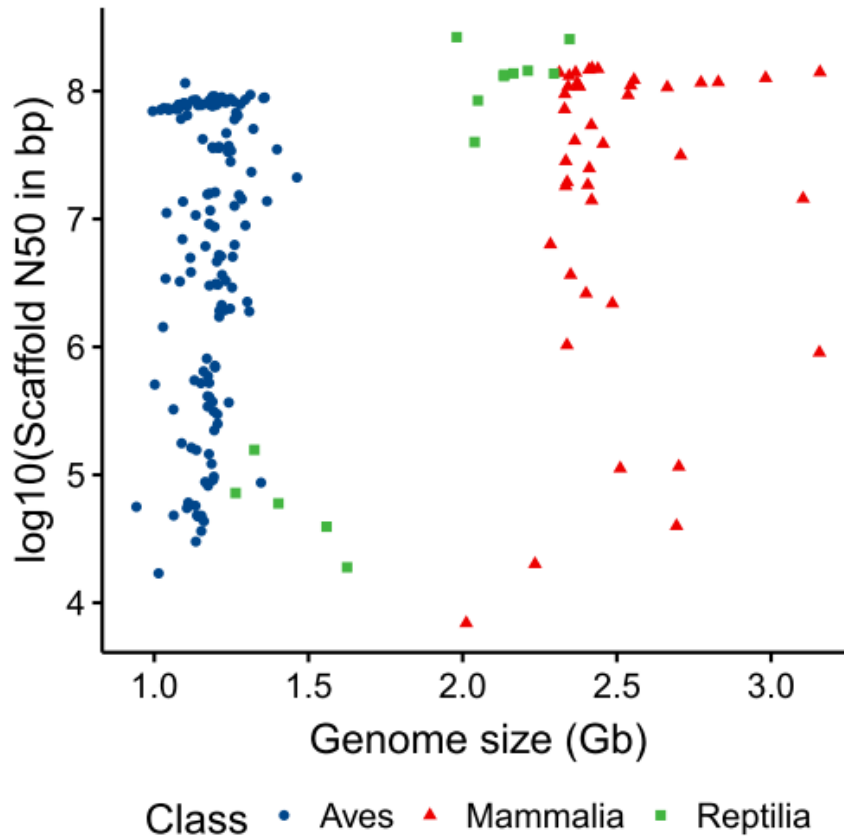


图 6.2 已发表基因组的大小和质量（连续性）统计

同时需要注意的是，海洋四足动物已被基因组测序的物种仍然偏少，还有 249 个属（占 62.66%）未有物种被测序，亟需获得至少一个物种的完整基因组序列来保存该分支的遗传信息。因此，在基因组的数量和质量上来看，海洋四足类研究仍处于较低水平。

根据现有组装完成的基因组统计，海洋哺乳类基因组大小为 2.01-3.16Gb，平均为 2.50 Gb；海鸟类基因组大小为 0.94-1.46 Gb，平均为 1.18Gb；海洋两栖动物基因组大小为 1.26-2.35 Gb，平均为 1.90 Gb。整体来讲，海洋四足动物的基因组大小呈现相对恒定的趋势，同时，GC 含量变化不大，平均为 41.95%。

6.3 海洋四足类基因组研究热点

海洋四足动物主要研究热点主要集中在：基于基因组数据构建大规模的种系发育树，重构系统演化发育过程；探讨为适应海洋环境的所发生的趋同

演化的分子机制；基于群体基因组学的种群遗传学研究，以及应用于濒危物种的保护。

6.3.1 种系发育研究

中科院深海所、西北工业大学、青岛华大基因研究院等单位，从科级水平出发对 17 个海洋哺乳动物物种进行了全基因组测序及组装，并结合已发表的海洋哺乳动物基因组数据，重建了基于全基因组数据有史以来的最全面的海洋哺乳动物系统发生树[12]。深圳华大生命科学研究院等单位，构建了全部 24 种现存企鹅和近代灭绝的 3 种企鹅的高质量基因组数据集，解析了企鹅内部六大支脉之间的关系，重建了企鹅的演化历史，进而解析了企鹅形态向海洋发生二次适应的演化规律[13]。

突变率也是遗传学和基因组学中的一个关键参数，用于确定系统发育进化和适应率。美国得克萨斯理工大学、密西西比州立大学等单位组装获得了咸水鳄基因组，并研究了其基因组的进化速率，包括碱基替换率、转座元件含量及跳跃、基因家族的进化和基因组共线性等，发现咸水鳄基因组进化较慢[14]。荷兰格罗宁根大学、美国海岸研究中心等单位基于系谱法推断鲸鱼的突变率，结果表明，鲸鱼的突变率与人类、猿类和海豚等小型哺乳动物的谱系中的突变率相似。相比之下，与这些较小的哺乳动物相比，使用系统发育方法对鲸鱼的早期估计要低得多[15]。

6.3.2 环境适应性研究

作为原始生命的孕育场所，海洋环境对地球上生命的起源与演化起着至关重要的作用。目前已知生活于海洋之中的物种数约 24 万种，其中大部分物种是从海洋中演化而来的，但是部分物种却是从陆地环境次生性返回海洋并逐渐适应生存下来的，哺乳动物中的鲸类便是最为人所熟知的类群。尽管不同支系的海洋四足动物具有完全不同的祖先和不同的演化历程，但是由于生活的海洋环境的相似性，这些动物在低氧耐受、体型和骨骼发育、嗅觉、听觉、味觉等感觉器官、渗透压调节、体

温维持等相关的基因上却发生了一致性改变，呈现对海洋环境的趋同演化的适应性演化机制。

在白鬻豚[6]、小须鲸[16]和南极小须鲸[10]基因组中，嗅觉相关基因家族明显收缩。同时，在这些基因组中，一些与味觉和视觉相关基因/家族经历收缩或假基因化失去功能，甚至完全丢失。特别地，相对于陆生四足动物来说，海洋哺乳类听觉及回声定位相关的基因经历了显著的快速进化。最后，对于海洋哺乳动物来说，趋同进化也已经被广泛研究，比如齿鲸的回声定位系统及适应性分子趋同与表型趋同的联系等[2, 17-19]。另外，一些与水生适应性相关的基因或基因家族如 *PRDX*、*OGT*、*SLC16A1*、*PRDX*、*OGT*、*SLC16A1*、*DAG1* 和 *BTN1A1* 等经历了正选择或者扩张，从而适应了海洋低氧、高渗透压、深潜[16]及极端环境的压力[5]。此外，北极露脊鲸基因组的研究发现了一些与 DNA 修复、细胞周期、肿瘤及老化相关基因的获得或者丢失，很可能与其长寿相关[20]。更大规模的海洋哺乳动物基因组研究，如对海洋哺乳动物三个主要支系（鲸类、鳍足类、海牛类）从陆地重返海洋的分子适应机制进行了全面分析和探究[12]。研究发现海洋哺乳动物从产热和散热两个方面，通过双重调控的方式维持体温恒定：通过 *NFIA* 和 *UCP1* 基因来调控棕色脂肪细胞的合成和利用从而控制产热的变化，通过 *SMEA3E* 基因使血管系统发生适应性改变以调节热量的散失。同时，*XYLT1* 和 *FMA20B* 两个关键基因在海洋哺乳动物中发生了特异性改变，可能会影响到海洋哺乳动物骨骼形态的变化：前肢变成鳍状肢，身体呈流线型或者纺锤形。除此，还发现了与低氧耐受、回声定位、深潜及视力相关基因在海洋哺乳动物中发生了分子层面的适应性演化。

大规模企鹅基因组学研究，揭示了企鹅二次适应海洋生态环境的基因遗传学基础。研究解析了一系列与体温调节、味觉偏好、肢体发育、骨骼生长和免疫系统等相关的受到正选择、快速进化、家族扩张或假基因化的基因，揭示了企鹅对寒冷环境适应以及水下生活的潜在分子机制[13]。比如，光敏感相关的 *TEEM30A*，*KCNV2* 等基因突变加强了企鹅在水下及暗光环境下敏锐的视觉；血红蛋白和肌红蛋白相关的 *TNMD*，*TBXT* 等基因突变使得企鹅能够更高效地利用血液中的氧气。

海洋爬行动物绿海龟、咸水鳄的基因组测序，为研究爬行类到鸟类的进化提供了基础，并且通过研究胚胎发育和形体发育，鉴定出了一些与肢体发育相关的基因（如 *wnt5a*）[7]。

6.3.3 基于全基因组数据的海洋四足动物保护

根据 2018 年国际自然保护联盟（IUCN2018）统计，目前有对应数据的约 261 个物种处于极度濒危（critically endangered, CR）、濒危（endangered, EN）、脆弱（vulnerable, VU）和近危（near threatened, NT）的状态，这些物种的保护应该引起人们的注意。目前，物种保护首要做的是增强人们对自然保护的认识，采取一系列保护措施，如限制人为的野蛮捕获、减少对其栖息地的破坏、尽量减少各种无意识的污染并进行广泛的野外搜救行动等。

基因组数据还可以为海洋生物多样性的管理和保护提供良好的分子基础。通过基因组数据，我们可以研究物种在应对自然环境压力和人为威胁（如污染）时的适应机制和恢复能力，这有利于设定保护优先级和种群恢复策略。此外，全基因组学研究可以揭示物种面对某些生物威胁（如疾病和有毒食物）时所产生的特征变化。例如，通过对存活海豚和死亡海豚种群之间全基因组差异的比较发现，两者之间等位基因频率发生了大量变化，这可能是由于有害的藻类大量繁殖所致，此项研究有助于鉴定出抵抗甲藻毒素的候选基因[21, 22]。中科院深海所、青岛华大基因研究院等单位等完成了中华白海豚染色体水平基因组及群体遗传学研究，揭示了中华白海豚的演化历史和特点，提示中华白海豚面临着灭绝的高风险，亟需加强保护[23]。奥斯陆大学、苏黎世联邦理工学院等单位通过对大西洋海雀 12 个种群的 72 个个体的参考基因组和全基因组重测序数据进行分析，确定了四个遗传上不同的大集群，发现了一系列复杂的当代生物因素阻碍了基因在不同空间尺度上的流动，证明了全基因组数据在揭示洄游海鸟意外种群结构及其在海鸟分类、进化和保护方面的价值[24]。

全基因组数据还可以揭示某些濒危物种不能适应选择压力的分子机制。例如，免疫相关基因遗传多样性的缺失和毒素降解相关基因的有害突变富集导致了朱鹮和

其它濒危物种的主要遗传缺陷的形成[25]。中科院深海所、青岛华大基因研究院等单位对虎鲸群体进行全基因组重测序，揭示了近亲繁殖导致的长纯合性片段对虎鲸群体的影响，对海洋哺乳类的遗传保护起到重要的理论支撑作用。

历史种群研究是另外一个被广泛研究的热点[6, 16, 26-28]，这些研究可以估计在应对自然环境变化或人为压力时物种种群大小及数量的变化，从而有助于理解物种生存所需要的最佳环境及为物种保护提供了良好的理论基础。

6.4 讨论与展望

总之，获得更多、更全面、更高质量的全基因组数据，有助于了解物种本身的遗传基础及应对疾病、压力等的适应性变化。再结合进化基因组学、转录组学、群体遗传学等研究，为保护濒危物种提供良好的数据基础。为了更好地利用基因组数据对海洋四足动物进行保护，我们计划分三个阶段对这些物种进行全基因组测序。第一阶段：CR 和 EN 类物种；第二阶段：VU 和 NT 类物种；第三阶段：对每个科至少进行一个物种测序。获得这些基因组数据后，我们将能够整合更全面的数据来进行海洋四足动物的保护和管理。

本章作者

孙帅、张要磊、刘雅琳、孟亮（青岛华大基因研究院）

本章参考文献:

1. WoRMS. World Register of Marine Species. 2023; Available from: <http://www.marinespecies.org/index.php>.
2. Foote, A.D., et al., Convergent evolution of the genomes of marine mammals. *Nature genetics*, 2015. 47(3): p. 272.
3. Williston, S.W., *Water reptiles of the past and present*. 1914: University of Chicago Press.
4. Shao, Y., et al., Transcriptomes reveal the genetic mechanisms underlying ionic regulatory adaptations to salt in the crab-eating frog. *Scientific Reports*, 2015. 5: p. 17551.
5. 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*, 2012. 109(36): p. E2382-E2390.
6. Zhou, X., et al., Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nature communications*, 2013. 4: p. 2708.
7. 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*, 2013. 45(6): p. 701.
8. Fan, G., et al., The first chromosome-level genome for a marine mammal as a resource to study ecology and evolution. *Molecular Ecology Resources*, 2019. 19(4): p. 944-956.
9. Li, A., et al., Two reference-quality sea snake genomes reveal their divergent evolution of adaptive traits and venom systems. *Molecular Biology and Evolution*, 2021. 38(11): p. 4867-4883.
10. Kishida, T., et al., Aquatic adaptation and the evolution of smell and taste in whales. *Zoological letters*, 2015. 1(1): p. 1.
11. Wolf, M., et al., The genome of the pygmy right whale illuminates the evolution of rorquals. *BMC biology*, 2023. 21(1): p. 79.
12. Yuan, Y., et al., Comparative genomics provides insights into the aquatic adaptations of mammals. *Proceedings of the National Academy of Sciences*, 2021. 118(37): p. e2106080118.
13. Cole, T.L., et al., Genomic insights into the secondary aquatic transition of penguins. *Nature communications*, 2022. 13(1): p. 3912.
14. Green, R.E., et al., Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science*, 2014. 346(6215): p. 1254449.
15. Suárez-Menéndez, M., et al., Wild pedigrees inform mutation rates and historic abundance in baleen whales. *Science*, 2023. 381(6661): p. 990-995.
16. Yim, H.-S., et al., Minke whale genome and aquatic adaptation in cetaceans. *Nature genetics*, 2014. 46(1): p. 88.
17. Liu, Y., et al., Convergent sequence evolution between echolocating bats and dolphins. *Current Biology*, 2010. 20(2): p. R53-R54.
18. Shen, Y.-Y., et al., Parallel evolution of auditory genes for echolocation in bats and toothed whales. *PLoS Genetics*, 2012. 8(6): p. e1002788.

19. Parker, J., et al., Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*, 2013. 502(7470): p. 228.
20. Keane, M., et al., Insights into the evolution of longevity from the bowhead whale genome. *Cell reports*, 2015. 10(1): p. 112-122.
21. Cammen, K.M., et al., 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*, 2015. 67(2): p. 125-133.
22. Cammen, K.M., et al., Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*). *Molecular ecology*, 2015. 24(18): p. 4697-4710.
23. Zhang, P., et al., An Indo-Pacific humpback dolphin genome reveals insights into chromosome evolution and the demography of a vulnerable species. *Iscience*, 2020. 23(10).
24. Kersten, O., et al., Complex population structure of the Atlantic puffin revealed by whole genome analyses. *Communications Biology*, 2021. 4(1): p. 922.
25. Li, S., et al., Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome biology*, 2014. 15(12): p. 557.
26. Árnason, Ú., et al., Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science advances*, 2018. 4(4): p. eaap9873.
27. Zhou, X., et al., Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nature communications*, 2018. 9(1): p. 1276.
28. Foote, A.D., et al., Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature communications*, 2016. 7: p. 11693.

456万 个单细胞

46 个水生生物

24 个组织器官



第七章 水生生物细胞组学

7.1 水生生物细胞组学研究概述

细胞基因的选择性表达决定了细胞的蛋白质组成和表型。尽管生物体的所有细胞都含有相同的遗传物质成分，但不同的细胞表现出不同的基因表达偏好，从而导致表型的多样性，进一步表现为细胞异质性[1]。细胞异质性是各类复杂生物系统的基本特征，例如发育中的胚胎、发育中的器官和肿瘤等。单细胞测序(scRNA-seq)和空间转录组学技术分析基因表达的高分辨率特征，为构建各种器官和生物体的进化和发育图谱提供了新的契机和切入点[2-4]。通过表征特定的细胞类型，scRNA-seq 实现了对单个细胞精确基因表达模式的快速确定，而空间转录组学通过保留空间信息来识别细胞是如何在空间上组织分布的[5]。随着单细胞和空间组学测序技术的蓬勃发展，大量的单细胞和空间组学数据不仅提升了我们对细胞类型和细胞命运选择机制的了解，还促进了我们在细胞水平上对生物体发育过程、物种进化和疾病发生发展过程的理解[6]。

浩瀚而神秘的海洋是一个复杂的水生生态系统，覆盖了地球表面 71% 的面积，具有极高的生物多样性。与陆地物种不同，海洋生物生活在高盐度、高压、光照不足、低温、缺氧等极端环境中，表现出独特的发育特征和生物多样性[7]。广泛而多样的海洋物种为探索各种科学问题提供了巨大的资源，例如生命的起源和进化、发育生物学以及对复杂环境的适应等。快速的技术革命为探索这些问题提供了前所未有的机会。

单细胞和空间组学通量的急剧增长让研究者们能够实现以单细胞分辨率在各种代表性水生生物中构建细胞景观，通过单细胞测序和空间转录组学技术研究众多具有进化重要性和代表性的水生物种，这些研究通过重塑我们对各种水生生物和器官中细胞类型的认识，加深了我们对进化和发育生物学基本问题的理解。

7.2 水生生物细胞图谱数据库构建

水生细胞组学数据具有其独特性和重要性，为了便于搜索、访问、可视化、聚

焦和拓展应用水生生物的单细胞和空间组学数据，我们开发了一个新的门户网站，即水生生物细胞图谱数据库（https://www.bgiocean.com/white_paper/）（图 7.1）。通过调研和总结数百份聚焦于进化上重要的、具有代表性意义的水生物种的细胞组学研究文献，我们收集和分析了单细胞和空间组学数据，收录在数据库中。同时此数据库收录了青岛华大基因研究院自主测序的肺鱼、领航鲸等重要水生物种的单细胞和空间组学数据，后续陆续将有一系列青岛华大基因研究院自主测序的水生细胞组学数据持续上传更新数据库，便于研究者访问。

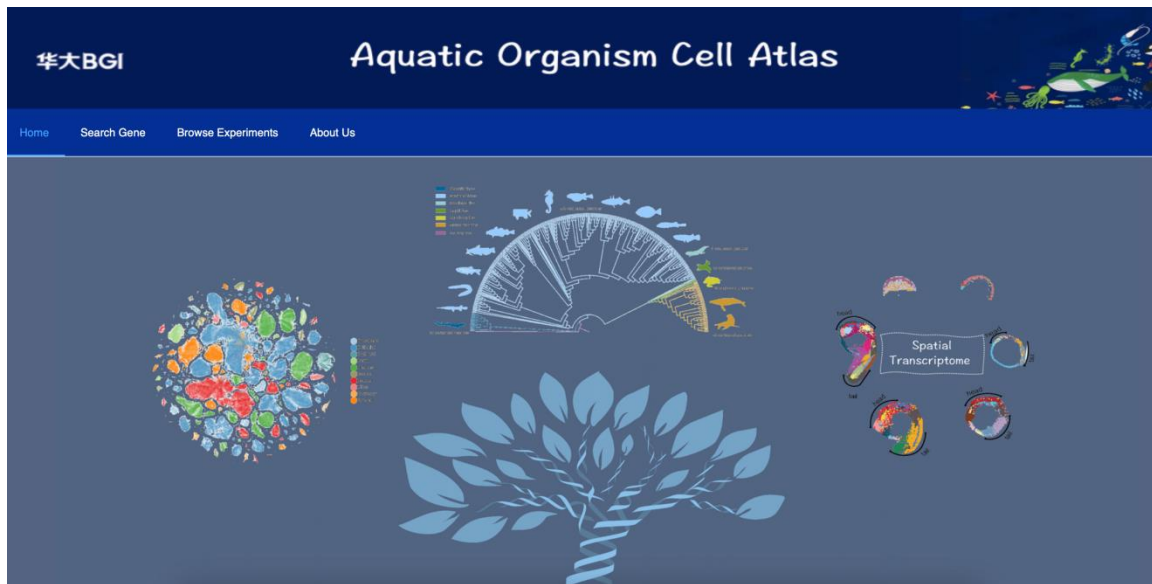


图 7.1 水生生物细胞图谱数据库主页

7.2.1 数据库数据量统计

水生生物细胞图谱数据库包含 456 万个单细胞，描绘了 46 个物种的 24 个器官，涵盖了水生物种进化中的 13 个代表性门（表 7.1），总结收集了 113 个针对进化节点水生物种的研究著作，这些研究涉及多种代表性物种，如模式物种斑马鱼（*Danio rerio*）、墨西哥钝口螈（*Ambystoma mexicanum*）和非洲爪蟾（*Xenopus laevis*）等（表 7.1）。

水生生物单细胞和空间转录组学研究关注的科学热点之一是发育生物学，旨在阐明胚胎细胞谱系从多能性到成体细胞的发育过程[8]。调研结果显示，多数水生单细胞研究集中在描述代表性水生物种的胚胎发育细胞谱系特征以及分子表达变化

特征[9-15]（图 7.2）。同时，时空组学研究实现了斑马鱼胚胎发育的基因表达景观和发育轨迹的时空映射[16]。一个全面而准确的胚胎细胞谱系参考图谱对于理解胚胎发育的体内生物学过程至关重要。除此之外，包括大脑、心脏、四肢和肠道在内的重要组织器官在水生生物单细胞组学中也受到了相当大的关注[17-25]（图 7.2）。值得注意的是，研究人员利用 MGI Stereo-seq 建立了蝾螈大脑的空间景观，并解码了室管膜胶质细胞的潜在神经发生过程，以促进发育和再生[26]。这些单细胞和空间组学数据将有助于了解海洋生物器官系统的发展、调节和进化。

7.2.2 细胞图谱重建

利用这些可访问的测序数据，我们重新分析了 293 万个单细胞，并在水生生物细胞图谱数据库中可视化了 24 个器官和 25 个物种的 42 个细胞图谱，包含所注释的细胞类型（图 7.3）。水生生物细胞图谱包括以下代表性物种的细胞景观：文昌鱼（*Branchiostoma floridae*），章鱼（*Octopus vulgaris*），玻璃海鞘（*Ciona robusta*, *Ciona intestinalis*, *Ciona savignyi*），斑马鱼（*Danio rerio*），水母（*Clytia medusa*），海葵（*Nematostella vectensis*），栉水母（*Mnemiopsis leidyi*）和海绵动物（*Amphimedon queenslandica*, *Spongilla lacustris*）等。

表 7.1 水生生物单细胞测序研究汇总

Phylum	Species	Tissue
Annelida	<i>Platynereis dumerilii</i>	whole-body
Apicomplexa	<i>Plasmodium berghei</i>	whole-body
Arthropoda	<i>Litopenaeus vannamei</i> , <i>Penaeus monodon</i>	hepatopancreas, blood
	<i>Alligator mississippiensis</i> , <i>Ambystoma mexicanum</i> , <i>Anolis carolinensis</i> , <i>Aspidelaps lubricus</i> , <i>Astyanax mexicanus</i> , <i>Branchiostoma floridae</i> , <i>Branchiostoma japonicum</i> , <i>Chrysemys picta bellii</i> , <i>Ciona intestinalis</i> , <i>Ciona robusta</i> , <i>Ciona savignyi</i> , <i>Cynoglossus semilaevis</i> , <i>Danio rerio</i> , <i>Epinephelus coioides</i> , <i>Gadus morhua</i> , <i>Lates calcarifer</i> ,	limb, brain, immune, tail, liver, lung, gill, testis, fin, vascular, spinal cord, embryo, intestine, whole-body,

	<i>Oreochromis niloticus</i> , <i>Petromyzon marinus</i> , <i>Phallusia mammillata</i> , <i>Pogona vitticeps</i> , <i>Salmo salar</i> , <i>Trachemys scripta</i> , <i>Xenopus laevis</i> , <i>Xenopus tropicalis</i>	kidney, pancreatic islet, lateral line, thyroid gland, ovarian, venom, pancreas islets, heart
Cnidaria	<i>Xenia</i> sp., <i>Stylophora pistillata</i> , <i>Clytia hemisphaerica</i> , <i>Nematostella vectensis</i> , <i>Hydra vulgaris</i>	whole-body
Echinodermata	<i>Strongylocentrotus purpuratus</i> , <i>Lytechinus variegatus</i>	whole-body, embryo
Mollusca	<i>Yesso scallops</i> , <i>Dreissena rostriformis</i> , <i>Octopus vulgaris</i> , <i>Crassostrea hongkongensis</i>	muscle, whole-body, brain, blood
Nematoda	<i>Caenorhabditis elegans</i>	whole-body, embryo, brain
Platyhelminthes	<i>Schistosoma mansoni</i> , <i>Schmidtea mediterranea</i>	whole-body
Porifera	<i>Spongilla lacustris</i>	whole-body
Xenacoelomorpha	<i>Isodiametra pulchra</i>	whole-body
Ctenophora	<i>Mnemiopsis leidyi</i>	whole-body
Placozoa	<i>Trichoplax adhaerens</i>	whole-body

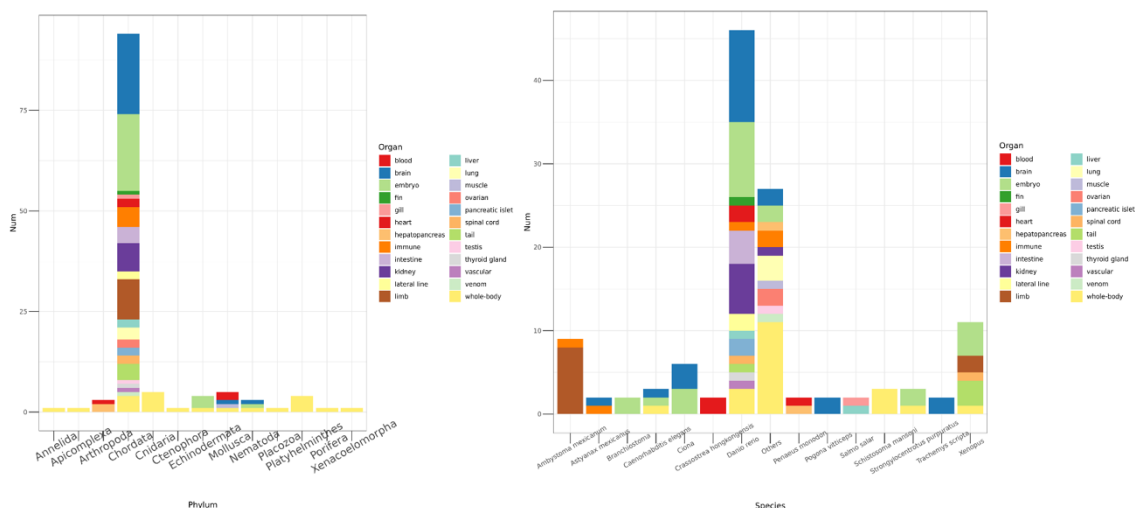


图 7.2. 水生生物单细胞和空间图谱数据库中总结整理的单细胞和空间转录组的研究。

图左：按门分类的数量统计；图右：按物种分类的数量统计。不同颜色表示研究的不同器官类型。

这些物种涵盖了海洋生物进化和发育的多个代表性分支，用户可以在水生生物细胞图谱数据库中轻松搜索和浏览这些可视化的细胞图谱，以查找感兴趣的物种。值得关注的是，Porifera、Placozoa、Ctenophora、Coelenterata、Platyhelminthes、Mollusca、Echinodermata 和 Chordata 等代表性门的水生物种均已进行了单细胞或空间组学测序（表 7.1 和图 7.3）。数据库中，脊索动物门的细胞测序数量占据绝对优势（87.13%；约 250 万个细胞）。这些进化上重要分支的单细胞和空间转录组学数据为从细胞角度研究海洋生物进化和发育的基础科学问题提供了重要机会。

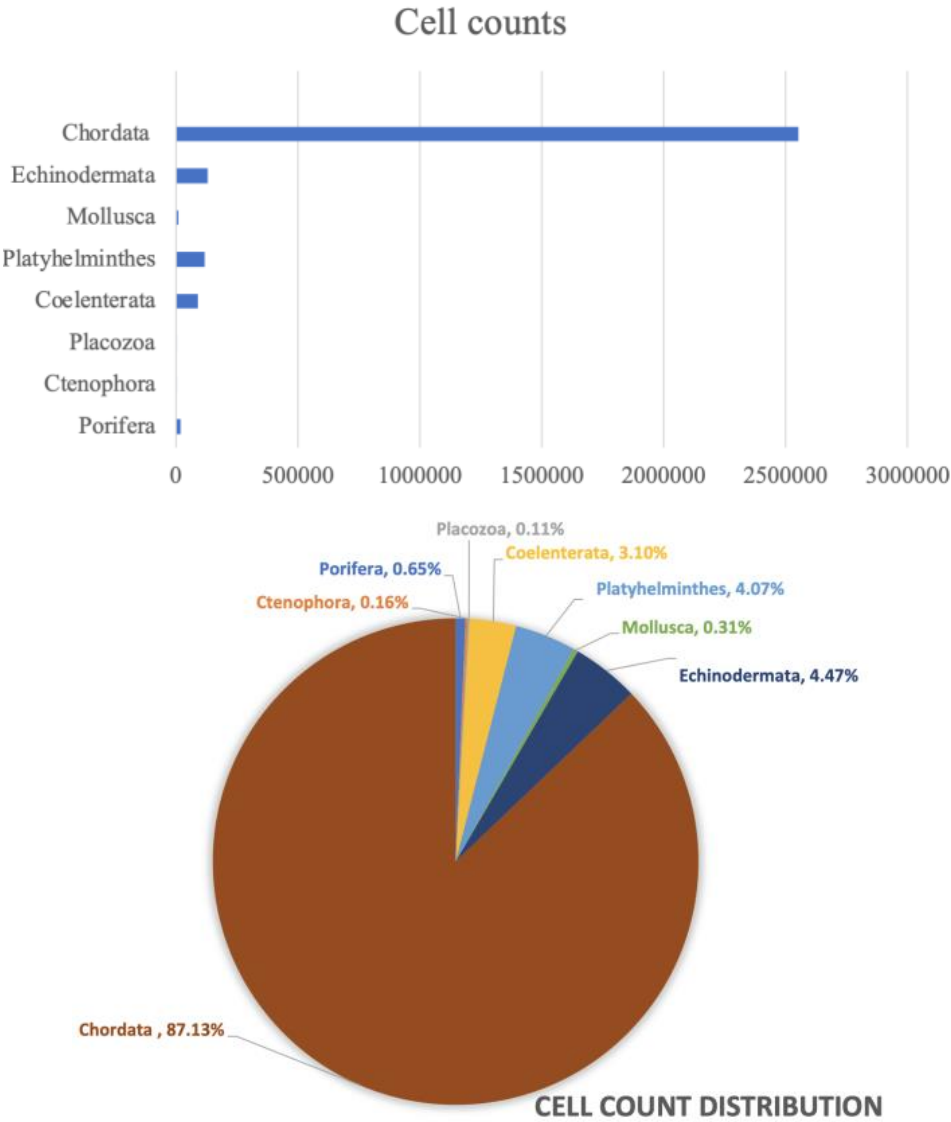


图 7.3. 水生单细胞和空间图谱数据库中的水生生物研究细胞数及占比统计

7.3 水生生物细胞图谱数据库应用

庞大的数据资源势必蕴含着丰富的信息，但是信息的无序性和复杂程度又造成归纳整理的时间成本和数据存储的硬件成本。此外，在当今单细胞转录组测序和时空组技术的成本仍具备一定门槛的情况下，借助于公开发表的数据库对自己的科研项目进行补充和升华是提升研究深度、丰富研究内容、节约研究经费的不二之选。尽管基于常见模式物种（如人、鼠、斑马鱼等）的数据库构建和应用已经屡见不鲜[27-30]，单细胞组学相关的研究成果也相继涌现[31-32]；而针对水生生物这一具有广泛地理分布和丰富物种数量的类群而言，针对其细胞组学数据的整合和应用仍处于初期，还有待进一步探索，我们所开发和维护的水生生物数据库在一定程度上填补了这一空白。在此，结合本数据库中收录的数据资源和近十多年受到广泛关注的进化发育生物学和神经细胞生物学课题，我们通过具体的数据分析结果对数据库中相关数据集的综合应用实例和前景进行讨论。

7.3.1 在胚胎发育研究中的应用

胚胎发育是生物个体形成所必须经历的生命阶段，涵盖了胚胎构造由简到繁，个体由细胞到组织再到器官的发展变化。其中细胞作为生命活动的基本单位，也经历了由干细胞到成熟细胞的逐级分化过程。得益于单细胞组学数据高通量和高分辨率的特点，对于不同胚胎发育时期的细胞类型的分化关系可以借助成熟的计算和分析软件进行深入的认识和探究。

以水生生物细胞图谱数据库中收录的绿海胆[33]和斑马鱼胚胎发育[11]的数据为例，利用 TOME 算法，对其胚胎发育各个阶段的细胞类型相关性进行整体水平的认识和评估（图 7.4 和图 7.5）。其中在绿海胆原文中构建了 18 个受精后不同时期的共 50,935 个细胞数的细胞图谱，但其文章内容多为对关键细胞类型和分子调控的研究，尽管也涉及对分化轨迹的探究，但是只选用了 4 个受精后的时间点进行讨论，并未形成全局层面的认识。利用更新的 TOME 算法[34]，可对胚胎发育的各个时期相关性进行系统化认识，从我们的结果来看，在 6 hpf（hour post-fertilization，受精后小时数）之前，基于原文的注释结果，胚胎细胞并未表现出明显的特异性，

在 6 hpf 时期，内胚层和中胚层细胞开始出现且向后延续，并可直观地得出内胚层祖细胞为 5 hpf 时期的 S2 细胞类群。同时，PGC（primordial germ cells，原始生殖细胞）在 6 hpf 时期依然存在，并在 7 hpf 分化为 PMC（primary mesenchyme cells，原代间充质细胞/成骨细胞）和 NSM（non-skeletogenic mesenchyme，非成骨细胞）。

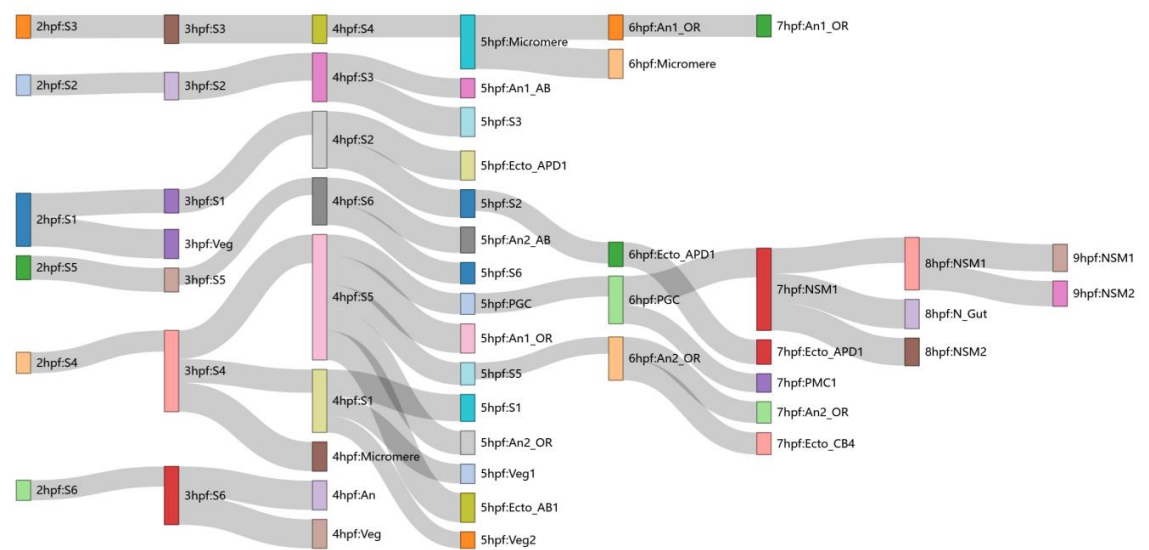


图 7.4. 绿海胆从 2hpf 到 9hpf 的发育谱系。

对斑马鱼胚胎发育的研究，文章选用了受精 24 小时内的 7 个时间节点，共涉及 92,000 个细胞。我们对其发育各阶段的细胞分化关系进行重新归纳整理之后可看出，随着胚胎发育的进行，细胞类型的复杂程度也在迅速增加（图 7.5）。得益于斑马鱼作为典型模式物种扎实的研究基础、较大规模的数据量和精细的细胞类型划分，其胚胎发育阶段的细胞类型谱系关系也更加明晰。在 6 hpf 中注释到的胚胎边缘位置的细胞类群与 8 hpf 的中胚层及 10 hpf 以后进一步分化的血液、皮肤、泌尿系统、生殖系统具有连续一致的相关性，对于中胚层的发育阶段进行了系统化的梳理。此外，胚胎发育过程中也伴随着细胞凋亡，在 6 hpf 时期的凋亡细胞与 8 hpf 时期的凋亡细胞具有连续性并且止步于下一发育阶段。



图 7.5. 斑马鱼从 4hpf 到 24hpf 的发育谱系。

除了利用单个物种的多个发育阶段数据进行解读以外，基于数据库中收集整理多个物种门类的胚胎发育数据，还可通过跨物种研究进一步对其发育过程中的进化特征进行探究。本文选取绿海胆（*Lytechinus variegatus*）[33]、玻璃海鞘（*Ciona intestinalis*）[9]、文昌鱼（*Branchiostoma floridae*）[35]、斑马鱼（*Danio rerio*）[11]和爪蟾（*Xenopus tropicalis*）[36]单细胞数据中的最后一个发育阶段，利用 SAMap[37]工具进行跨物种横向比较，从细胞组学水平揭示其进化的关联性（图 7.6）。图中显示，绿海胆的囊胚结构与玻璃海鞘和文昌鱼的脊索、斑马鱼和爪蟾的神经基板显示出保守性；从绿海胆到文昌鱼的内胚层与斑马鱼和爪蟾的部分神经结构相似，而物种间间质组成存在普遍的保守性；上述结果表明棘皮动物和脊索动物之间存在发育过程中的保守性。此外，从图中可以看出，斑马鱼和爪蟾的多种细胞类型存在保守性，如上皮祖细胞、后脑结构、视叶细胞、髓系免疫细胞、间质细胞等，表明鱼类和两栖类进一步的保守关系。

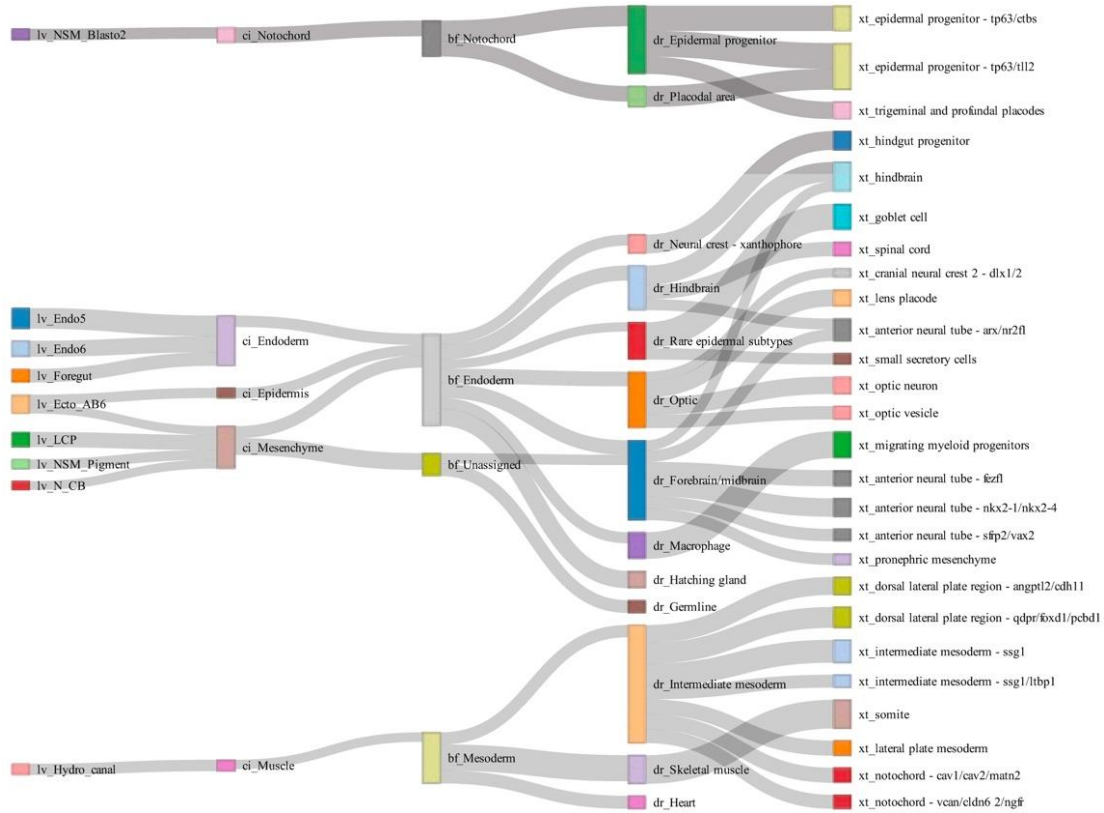


图 7.6. 绿海胆、玻璃海鞘、文昌鱼、斑马鱼和爪蟾胚胎之间的发育相关性。

7.3.2 在神经系统演化研究中的应用

神经系统作为生物体各项生命活动的控制中枢，对生物体的生存发挥着至关重要的作用。随着单细胞组学技术的发展和成熟，其在关键进化节点物种的神经系统相关的应用成果也陆续发表[9, 38-43]，对解析神经的起源和适应性进化提供了重要的数据和理论支撑。水生生物数据库中也收集整理了一系列物种的神经细胞数据集，本文利用其中典型的数据对其应用加以说明。

首先，本文选取玻璃海鞘（*Ciona intestinalis*）[9]、斑马鱼（*Danio rerio*）[42]、爪蟾（*Xenopus laevis*）[44]和红耳龟（*Trachemys scripta elegans*）[38]各物种幼体的神经元或者脑相关数据，利用 SAMap 对物种间细胞类型的保守性进行探究（图 7.7）。结果显示，玻璃海鞘的神经元细胞类型与斑马鱼脑多种神经元细胞类型、爪蟾脑 GABA 神经元细胞、红耳龟 aDVR（anterior dorsal ventricular ridge，前背侧室嵴）具有显著的保守性。基于物种间神经元细胞的相关性，我们进一步绘制了每两

个物种间 GABA 能神经元细胞保守基因的通路富集图（图 7.8），结果表明，各保守基因集富集到突触囊泡循环和 GABA 能突触等神经元细胞相关通路，说明物种间神经元相关细胞类型在分子表达水平也具有保守性。此外，玻璃海鞘幼体中的间质细胞与斑马鱼脑和爪蟾脑部分免疫细胞、爪蟾脑小胶质细胞显示出保守性，而玻璃海鞘原文中并未提及间质细胞免疫相关的内容，这一结果对于后续数据的深入分析和挖掘提供了方向和线索，也说明了基于数据库海量数据挖掘信息的必要性和重要性。

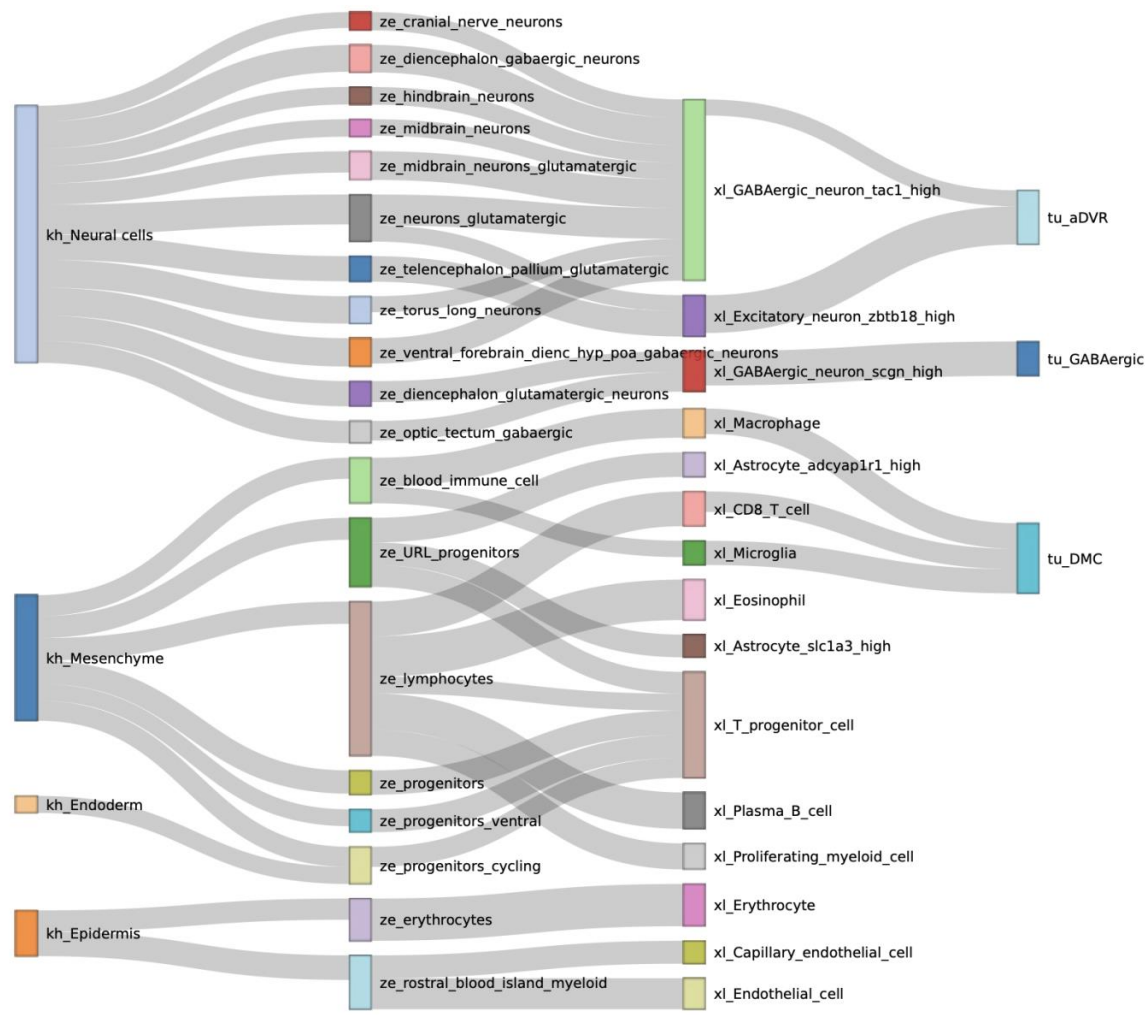


图 7.7. 玻璃海鞘、斑马鱼、爪蟾和红耳龟神经元细胞之间的发育相关性。

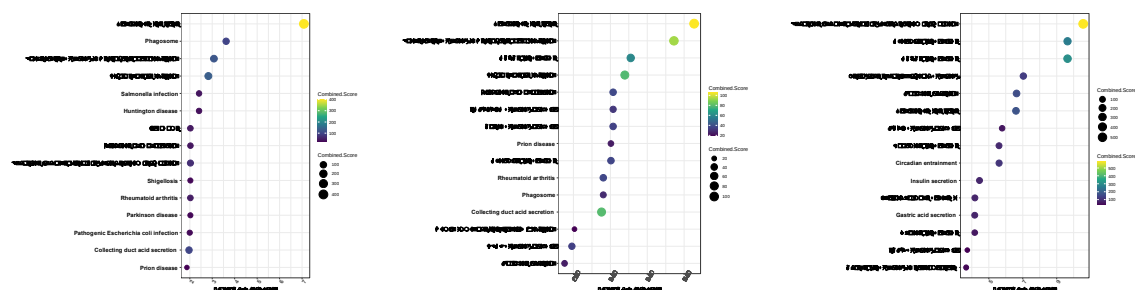


图 7.8 玻璃海鞘与斑马鱼（左）、斑马鱼与爪蟾（中）、爪蟾与红耳龟（右）之间神经元细胞保守基因集的富集 KEGG 通路。

除了对多个物种进行横向比较以外，我们研究并展示了玻璃海鞘幼体细胞数据转录调控元件的细胞类型特异性分布（图 7.9）。结果表明在神经元细胞中，多数转录因子具有较高评分，并且与神经元细胞的功能相对应。如 T-box 对神经元发育过程的影响[45]，bHLH 对神经元命运决定的作用[46]，zf-CCCH 对神经元形态的转录后调控等[47]。转录因子对于基因的选择性表达起到直接或者间接的调控作用，因此对其相关的分析有利于我们对细胞功能的深入理解。

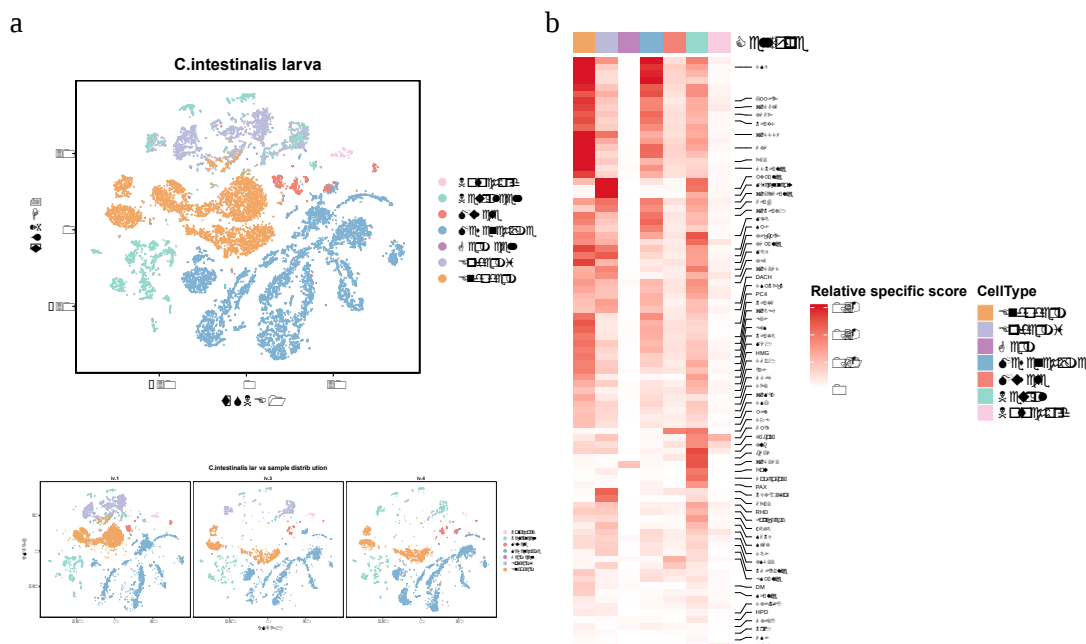


图 7.9. 玻璃海鞘幼虫的细胞图谱（a）和 TF 转录因子（b）的特异性评分。

7.4 讨论与展望

单细胞多组学技术对进化生物学领域的研究起到了强大的促进和推动作用，使得处于关键进化地位的非模式物种逐渐发展成为新模式物种，而其中水生生物占据了重要的环节，其丰富的物种门类和组成，对于揭示物种进化和起源的关键问题起到了至关重要的作用。多细胞生物体构成的复杂性意味着对其组学数据的整理和归纳更为必要，通过对不同器官、组织乃至生物系统数据的分级整理，有利于对单个物种进行细胞、分子和细胞通讯等多个角度的纵向分析，不同发育时期和实验组别的数据还可结合对其分化时序的推断进一步增加对其生物学现象的认识。而对位于不同进化节点物种在细胞或者器官水平的比较，有利于在宏观的时间维度上揭示细胞在进化过程中的保守性和异质性，进一步理解生物体和环境的相互作用和依赖关系。未来细胞组学数据的积累将会对进化发育生物学（Evolutionary developmental biology, evo-devo）等细化的交叉学科起到催化和推动作用，为生命科学的发展添砖加瓦。

单细胞测序和空间转录组学技术的发展使得研究者能够以单细胞分辨率研究细胞表型和细胞模式[48]，极大地促进了解析组织异质性、追踪胚胎细胞谱系进展和解码复杂器官发育的研究[49-51]。单细胞和空间组学的主要应用之一，是在不同组织和器官中创建了各种细胞类型的图谱[52]。单细胞和空间组学逐步揭示了各种水生生物的细胞景观，获得了不同细胞类型的大量细胞标记[9,11,53-55]。广泛的水生生物单细胞和空间组学数据可以提高我们对海洋生物发展和进化的认知。通过对不同发育阶段的细胞进行分析和注释，研究人员可以表征和追踪细胞命运决定过程和相应的调节因子[10,56,57]。通过比较不同物种的相似细胞类型，研究人员将深入了解特定细胞谱系和组织的演化历程[13]。

随着细胞组学技术的蓬勃发展和测序数据的产出，研究者们已经开始关注并陆续开发了一些跨物种细胞景观和细胞标记的数据库，这些数据库在细胞组学研究中已经开始显示它们的价值[53, 58]。我们开发的水生生物细胞图谱数据库重点侧重于收集水生物种的单细胞和空间组学数据以及细胞标记。考虑到海洋生物的生物多样性极高，以及海洋物种在早期生命和器官发育演化中发挥的关键作用，建立一个

特定的水生生物细胞图谱数据库对于全面准确地了解早期生命发育过程中的细胞类型和模式是非常必要的。该数据库将逐年持续更新。水生生物细胞图谱数据库和之前已发表的单细胞数据库将成为促进细胞组学研究的宝贵平台和资源。

本章作者：杜晓、刘群、张瑞华、李斌、秦雅婷、卢怡方、王睿、郭立东、韩凯、厉磊（青岛华大基因研究院）

本章参考文献

1. Altschuler, S.J. and L.F. Wu, *Cellular heterogeneity: do differences make a difference?* Cell, 2010. **141**(4): p. 559-63.
2. Potter, S.S., *Single-cell RNA sequencing for the study of development, physiology and disease.* Nat Rev Nephrol, 2018. **14**(8): p. 479-492.
3. Bingham, G.C., et al., *Spatial-omics: Novel approaches to probe cell heterogeneity and extracellular matrix biology.* Matrix Biology, 2020. **91-92**: p. 152-166.
4. Andersson, A., et al., *Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography.* Communications Biology, 2020. **3**(1): p. 565.
5. Dries, R., et al., *Advances in spatial transcriptomic data analysis.* Genome Res, 2021. **31**(10): p. 1706-1718.
6. Marioni, J.C. and D. Arendt, *How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology.* Annual Review of Cell and Developmental Biology, 2017. **33**(1): p. 537-553.
7. Zeppilli, D., et al., *Characteristics of meiofauna in extreme marine ecosystems: a review.* Marine Biodiversity, 2018. **48**(1): p. 35-71.
8. Jiang, M., X. Xu, and G. Guo, *Understanding embryonic development at single-cell resolution.* Cell Regen, 2021. **10**(1): p. 10.
9. Cao, C., et al., *Comprehensive single-cell transcriptome lineages of a proto-vertebrate.* Nature, 2019. **571**(7765): p. 349-354.
10. Farnsworth, D.R., L.M. Saunders, and A.C. Miller, *A single-cell transcriptome atlas for zebrafish development.* Developmental Biology, 2020. **459**(2): p. 100-108.
11. Wagner, D.E., et al., *Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo.* Science, 2018. **360**(6392): p. 981-987.
12. Ma, P., et al., *Joint profiling of gene expression and chromatin accessibility of amphioxus development at single cell resolution.* 2021.
13. Briggs, J.A., et al., *The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution.* Science, 2018. **360**(6392).
14. Massri, A.J., et al., *Developmental Single-cell transcriptomics in the Lytechinus variegatus Sea Urchin Embryo.* Cold Spring Harbor Laboratory, 2020.

15. Foster, S., et al., *Single cell RNA-seq in the sea urchin embryo show marked cell-type specificity in the Delta/Notch pathway*. Molecular Reproduction and Development, 2019. **86**.
16. Liu, C., et al., *Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis*. Cold Spring Harbor Laboratory, 2021.
17. Musser, J., et al., *Profiling cellular diversity in sponges informs animal cell type and nervous system evolution*. Cold Spring Harbor Laboratory, 2019.
18. Sharma, S., W. Wei, and A. Stolfi, *Single-cell transcriptome profiling of the Ciona larval brain*. Cold Spring Harbor Laboratory, 2018(2).
19. Bushra, et al., *Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain*. Nature Biotechnology, 2018.
20. Shafer, M.E.R., A.N. Sawh, and A.F. Schier, *Gene family evolution underlies cell type diversification in the hypothalamus of teleosts*. Nature Ecology & Evolution.
21. Tosches, M.A., et al., *Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles*. Science, 2018. **360**(6391): p. 881.
22. Willms, R.J., et al., *A Cell Atlas of Microbe-Responsive Processes in the Zebrafish Intestine*. Social Science Electronic Publishing.
23. Liu, X., et al., *Transcriptome Profiling of the Ovarian Cells at the Single-Cell Resolution in Adult Asian Seabass*. Frontiers in Cell and Developmental Biology, 2021. **9**.
24. Spanjaard, et al., *Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars*. Nature Biotechnology, 2018.
25. Wu, X., et al., *Single-Cell Atlas of Adult Testis in Protogynous Hermaphroditic Orange-Spotted Grouper, *Epinephelus coioides**. International journal of molecular sciences, 2021. **22**(22).
26. Wei, X., et al., *Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration*. Science, 2022. **377**(6610): p. eabp9444.
27. Moreno, P., et al., *User-friendly, scalable tools and workflows for single-cell RNA-seq analysis*. Nature Methods, 2021. **18**(4): p. 327-328.
28. Hu, C., et al., *CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data*. Nucleic Acids Research, 2022. **51**(D1): p. D870-D876.
29. Rozenblatt-Rosen, O., et al., *The Human Cell Atlas: from vision to reality*. Nature, 2017. **550**(7677): p. 451-453.
30. Han, X., et al., *Mapping the Mouse Cell Atlas by Microwell-Seq*. Cell, 2018. **172**(5): p. 1091-1107.e17.
31. Wang, J., et al., *Tracing cell-type evolution by cross-species comparison of cell atlases*. Cell Rep, 2021. **34**(9): p. 108803.
32. Geirsdottir, L., et al., *Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program*. Cell, 2019. **179**(7): p. 1609-1622.e16.
33. Massri, A.J., et al., *Developmental single-cell transcriptomics in the Lytechinus variegatus sea urchin embryo*. Development, 2021. **148**(19).
34. Qiu, C., et al., *Systematic reconstruction of cellular trajectories across mouse embryogenesis*. Nature Genetics, 2022. **54**(3): p. 328-341.

35. Ma, P., et al., *Joint profiling of gene expression and chromatin accessibility during amphioxus development at single-cell resolution*. Cell Reports, 2022. **39**(12): p. 110979.
36. Briggs, J.A., et al., *The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution*. Science, 2018. **360**(6392): p. eaar5780.
37. Tarashansky, A.J., et al., *Mapping single-cell atlases throughout Metazoa unravels cell type evolution*. eLife, 2021. **10**: p. e66747.
38. Tosches, M.A., et al., *Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles*. Science, 2018. **360**(6391): p. 881-888.
39. Fincher, C.T., et al., *Cell type transcriptome atlas for the planarian *Schmidtea mediterranea**. Science, 2018. **360**(6391): p. eaaq1736.
40. Styfahls, R., et al., *Cell type diversity in a developing octopus brain*. Nature Communications, 2022. **13**(1): p. 7392.
41. Sharma, S., W. Wang, and A. Stolfi, *Single-cell transcriptome profiling of the Ciona larval brain*. Dev Biol, 2019. **448**(2): p. 226-236.
42. Raj, B., et al., *Emergence of Neuronal Diversity during Vertebrate Brain Development*. Neuron, 2020. **108**(6): p. 1058-1074.e6.
43. Shafer, M.E.R., A.N. Sawh, and A.F. Schier, *Gene family evolution underlies cell-type diversification in the hypothalamus of teleosts*. Nature Ecology & Evolution, 2022. **6**(1): p. 63-76.
44. Liao, Y., et al., *Cell landscape of larval and adult Xenopus laevis at single-cell resolution*. Nature Communications, 2022. **13**(1): p. 4306.
45. Mihalas, A.B. and R.F. Hevner, *Control of Neuronal Development by T-Box Genes in the Brain*. Curr Top Dev Biol, 2017. **122**: p. 279-312.
46. Dennis, D.J., S. Han, and C. Schuurmans, *bHLH transcription factors in neural development, disease, and reprogramming*. Brain Res, 2019. **1705**: p. 48-65.
47. Murn, J., et al., *Recognition of distinct RNA motifs by the clustered CCCH zinc fingers of neuronal protein Unkempt*. Nat Struct Mol Biol, 2016. **23**(1): p. 16-23.
48. Yuan, G.C., et al., *Challenges and emerging directions in single-cell analysis*. Genome Biol, 2017. **18**(1): p. 84.
49. Han, X., et al., *Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing*. Genome Biol, 2018. **19**(1): p. 47.
50. Cochain, C., et al., *Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis*. Circ Res, 2018. **122**(12): p. 1661-1674.
51. Puram, S.V., et al., *Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer*. Cell, 2017. **171**(7): p. 1611-1624.e24.
52. Buettner, F., et al., *Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells*. Nat Biotechnol, 2015. **33**(2): p. 155-60.
53. Zhang, X., et al., *CellMarker: a manually curated resource of cell markers in human and mouse*. Nucleic Acids Res, 2019. **47**(D1): p. D721-d728.

54. Siebert, S., et al., *Stem cell differentiation trajectories in Hydra resolved at single-cell resolution*. Science, 2019. **365**(6451).
55. Seb -Pedr s, A., et al., *Early metazoan cell type diversity and the evolution of multicellular gene regulation*. Nat Ecol Evol, 2018. **2**(7): p. 1176-1188.
56. Marioni, J.C. and D. Arendt, *How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology*. Annu Rev Cell Dev Biol, 2017. **33**: p. 537-553.
57. Tambalo, M., R. Mitter, and D.G. Wilkinson, *A single cell transcriptome atlas of the developing zebrafish hindbrain*. Development, 2020. **147**(6).
58. Wang, R., et al., *Construction of a cross-species cell landscape at single-cell level*. Nucleic Acids Research, 2022.

附 录

附表 1.已公布的藻类基因组基本信息

Species	Phylum	Assembly Level	Total Length(bp)	Scaffold N50(bp)	Contig N50(bp)	GC(%)	Accession	Release Date
<i>Asterochloris erici</i>	Chlorophyta	Scaffold	55,586,817	313,547	215,402	58.5	GCA_019693375.1	2021/8/19
<i>Astrephomene gubernaculifera</i>	Chlorophyta	Contig	103,829,650	1,506,991	1,506,991	59.86	GCA_021605115.1	2021/11/19
<i>Auxenochlorella protothecoides</i>	Chlorophyta	Scaffold	22,924,637	285,543	35,091	63.54	GCA_000733215.1	2014/8/1
<i>Auxenochlorella protothecoides</i>	Chlorophyta	Scaffold	21,075,690	639,896	27,194	63.54	GCA_003709365.1	2018/11/1
<i>Auxenochlorella protothecoides</i>	Chlorophyta	Scaffold	32,730,026	16,332	5,798	63.54	GCA_002154875.1	2017/5/22
<i>Auxenochlorella pyrenoidosa</i>	Chlorophyta	Scaffold	56,992,954	1,392,758	11,732	66.61	GCA_001430745.1	2015/11/3
<i>Axilococcus clingmanii</i>	Chlorophyta	Scaffold	13,149,938	73,336	65,003	46.43	GCA_019693495.1	2021/8/19
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	11,239,930	14,789	12,932	60.71	GCA_029297925.1	2023/3/20
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	10,903,789	12,242	10,891	60.71	GCA_029297845.1	2023/3/20
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	9,429,038	8,371	7,642	60.71	GCA_029297815.1	2023/3/20
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	9,158,518	7,905	7,364	60.71	GCA_029297995.1	2023/3/20
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	9,530,899	7,940	7,109	60.71	GCA_029297885.1	2023/3/20
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	9,359,661	7,240	7,025	60.71	GCA_029298455.1	2023/3/20
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	2,684,685	4,273	3,896	60.71	GCA_029297835.1	2023/3/20
<i>Bathycoccaceae sp.</i>	Chlorophyta	Scaffold	4,479,146	5,239	3,446	60.71	GCA_029298105.1	2023/3/20
<i>Bathyococcus prasinos</i>	Chlorophyta	Chromosome	14,958,006	955,652	663,424	48.02	GCA_002220235.1	2012/11/22
<i>Bathyococcus prasinos</i>	Chlorophyta	Scaffold	8,892,781	5,665	4,915	48.02	GCA_029293465.1	2023/3/20
<i>Bathyococcus prasinos</i>	Chlorophyta	Scaffold	5,409,338	4,488	3,949	48.02	GCA_029290735.1	2023/3/20
<i>Bathyococcus prasinos</i>	Chlorophyta	Scaffold	2,504,736	3,912	3,416	48.02	GCA_029290765.1	2023/3/20
<i>Bathyococcus sp. NIOZ-UU96</i>	Chlorophyta	Scaffold	1,694,126	3,192	3,132	47.67	GCA_014466055.1	2020/9/3
<i>Bathyococcus sp. SAG1</i>	Chlorophyta	Contig	3,337,455	12,462	12,462	45.63	GCA_009618075.1	2019/11/7
<i>Bathyococcus sp. SAG2</i>	Chlorophyta	Contig	4,521,332	16,682	16,682	46.97	GCA_009618095.1	2019/11/7
<i>Bathyococcus sp. SAG3</i>	Chlorophyta	Contig	3,460,468	16,753	16,753	46.43	GCA_009618085.1	2019/11/7
<i>Bathyococcus sp. SAG4</i>	Chlorophyta	Contig	3,808,189	19,411	19,411	46.41	GCA_009618065.1	2019/11/7
<i>Bathyococcus sp. TOSAG39-1</i>	Chlorophyta	Scaffold	10,057,911	14,082	13,604	47.35	GCA_900128745.1	2016/12/2
<i>Botryococcus braunii</i>	Chlorophyta	Scaffold	184,381,702	372,998	163,917	50.83	GCA_002005505.1	2017/2/22
<i>Bryopsis sp. KO-2023</i>	Chlorophyta	Contig	90,654,258	6,725,592	6,725,592	46.07	GCA_030272585.1	2023/6/13
<i>Characiochloris sp. AAM3</i>	Chlorophyta	Scaffold	95,477,124	28,672	13,824	53.83	GCA_004335845.1	2019/3/8
<i>Chlamydomonas applanata</i>	Chlorophyta	Scaffold	78,504,178	105,699	25,013	59.39	GCA_001662365.1	2016/4/26
<i>Chlamydomonas asymmetrica</i>	Chlorophyta	Scaffold	141,916,335	114,158	22,718	58.87	GCA_001662385.1	2016/4/26
<i>Chlamydomonas eustigma</i>	Chlorophyta	Scaffold	66,605,202	465,125	59,700	44.81	GCA_002335675.1	2017/8/31
<i>Chlamydomonas incerta</i>	Chlorophyta	Contig	129,236,421	1,576,468	1,576,468	66.04	GCA_016834605.1	2021/2/8
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Chromosome	111,098,438	7,783,580	215,409	64.08	GCA_000002595.3	2018/1/26
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Complete	113,900,589	6,948,835	6,948,835	64.08	GCA_026108075.1	2022/11/9

<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Chromosome	111,112,239	6,886,220	6,779,811	64.08	GCA_013389655.1	2020/7/6
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Chromosome	112,471,279	6,862,525	5,839,277	64.08	GCA_018257845.1	2021/5/4
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Chromosome	109,022,210	8,096,202	4,253,305	64.08	GCA_027942765.1	2023/1/23
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Chromosome	110,269,002	6,791,016	3,808,610	64.08	GCA_027942735.1	2023/1/23
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Scaffold	114,515,278	7,007,943	6,313,329	64.08	GCA_949152325.1	2023/4/8
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Scaffold	112,603,837	7,061,149	3,072,450	64.08	GCA_949152315.1	2023/4/8
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	Scaffold	111,557,586	7,946,894	2,531,250	64.08	GCA_949141805.1	2023/4/8
<i>Chlamydomonas schloesseri</i>	Chlorophyta	Contig	130,201,002	1,206,536	1,206,536	64.38	GCA_016834595.1	2021/2/8
<i>Chlamydomonas</i> sp. 3112	Chlorophyta	Scaffold	120,063,769	14,239	6,379	62.76	GCA_004335865.1	2019/3/8
<i>Chlamydomonas</i> sp. 3222	Chlorophyta	Scaffold	120,063,769	14,239	6,379	62.76	GCA_004335795.1	2019/3/8
<i>Chlamydomonas</i> sp. AIC	Chlorophyta	Scaffold	52,664,990	32,757	17,683	56.88	GCA_004335895.1	2019/3/8
<i>Chlamydomonas</i> sp. ICE-L	Chlorophyta	Scaffold	541,862,442	19,230,455	221,978	49.23	GCA_013435795.1	2020/7/21
<i>Chlamydomonas</i> sp. UWO 241	Chlorophyta	Scaffold	211,641,456	375,902	279,208	60.64	GCA_016618255.1	2021/1/12
<i>Chlamydomonas</i> sp. WS3	Chlorophyta	Scaffold	114,119,837	24,298	22,412	67.01	GCA_004335755.1	2019/3/8
<i>Chlamydomonas</i> sp. haeroides	Chlorophyta	Scaffold	122,189,241	44,734	16,312	68.01	GCA_001662425.1	2016/4/26
<i>Chlorella desiccata</i> (nom nud)	Chlorophyta	Contig	21,550,525	1,641,916	1,641,916	44.98	GCA_019044685.2	2021/11/8
<i>Chlorella desiccata</i> (nom nud)	Chlorophyta	Contig	20,738,005	1,303,910	1,303,910	44.98	GCA_019202925.1	2021/7/13
<i>Chlorella ohadii</i>	Chlorophyta	Scaffold	57,091,209	328,977	31,681	63.82	GCA_025026875.1	2022/9/7
<i>Chlorella sorokiniana</i>	Chlorophyta	Complete	38,612,678	3,040,600	3,040,600	63.95	GCA_025917655.1	2022/10/31
<i>Chlorella sorokiniana</i>	Chlorophyta	Chromosome	58,362,567	4,091,730	3,818,101	63.95	GCA_003130725.1	2018/6/13
<i>Chlorella sorokiniana</i>	Chlorophyta	Contig	59,371,605	2,592,956	2,592,956	63.95	GCA_002245835.2	2018/3/16
<i>Chlorella sorokiniana</i>	Chlorophyta	Contig	61,202,522	2,415,094	2,415,094	63.95	GCA_002939045.1	2018/2/23
<i>Chlorella sorokiniana</i>	Chlorophyta	Contig	57,666,419	2,025,419	2,025,419	63.95	GCA_003116155.1	2018/5/11
<i>Chlorella sorokiniana</i>	Chlorophyta	Scaffold	53,781,828	4,138,376	131,078	63.95	GCA_006782985.1	2019/7/9
<i>Chlorella sorokiniana</i>	Chlorophyta	Scaffold	58,523,075	7,790,473	49,211	63.95	GCA_006782965.1	2019/7/12
<i>Chlorella sorokiniana</i>	Chlorophyta	Scaffold	57,528,089	15,040	7,878	63.95	GCA_026213835.1	2022/11/16
<i>Chlorella</i> sp.	Chlorophyta	Scaffold	44,185,683	24,761	24,761	63.91	GCA_029286985.1	2023/3/20
<i>Chlorella</i> sp. A99	Chlorophyta	Scaffold	40,934,037	1,727,419	22,979	70.46	GCA_003063905.1	2018/4/23
<i>Chlorella</i> sp. ArM0029B	Chlorophyta	Scaffold	92,961,256	805,067	17,364	63.34	GCA_002896455.3	2018/1/24
<i>Chlorella</i> sp. BAC 9706	Chlorophyta	Scaffold	32,957,766	521,038	35,389	61.92	GCA_013372505.1	2020/6/24
<i>Chlorella</i> sp. CH2018	Chlorophyta	Contig	56,831,849	17,989	17,989	65.21	GCA_009928355.1	2020/1/24
<i>Chlorella</i> sp. Dachan	Chlorophyta	Scaffold	60,207,685	2,575,675	35,243	65.19	GCA_006782975.1	2019/7/12
<i>Chlorella</i> sp. KRBP	Chlorophyta	Contig	126,281,769	10,159	10,159	61.34	GCA_004335735.1	2019/3/8
<i>Chlorella</i> sp. LC-ALGAE-MAG	Chlorophyta	Contig	18,969,561	46,087	46,087	71.46	GCA_025772305.1	2022/10/20
<i>Chlorella variabilis</i>	Chlorophyta	Scaffold	46,159,512	1,469,606	27,941	67.14	GCF_000147415.1	2010/9/16
<i>Chlorella vulgaris</i>	Chlorophyta	Scaffold	40,180,792	2,825,136	1,802,178	61.68	GCA_023343905.1	2022/5/16
<i>Chlorella vulgaris</i>	Chlorophyta	Scaffold	39,134,815	497,853	72,113	61.68	GCA_009720215.1	2019/11/28
<i>Chlorella vulgaris</i>	Chlorophyta	Scaffold	39,080,440	937,767	48,348	61.68	GCA_009720205.1	2019/11/28
<i>Chlorella vulgaris</i>	Chlorophyta	Scaffold	48,873,932	56,390	38,801	61.68	GCA_008119945.1	2019/8/20
<i>Chlorella vulgaris</i>	Chlorophyta	Scaffold	37,342,230	27,824	20,333	61.68	GCA_001021125.1	2015/6/5
<i>Chloroidium</i> sp. CF	Chlorophyta	Scaffold	54,312,065	2,512,641	45,266	52.17	GCA_004335625.1	2019/3/8
<i>Chloroidium</i> sp. JM	Chlorophyta	Scaffold	60,445,311	1,042,450	14,928	51.93	GCA_004335615.1	2019/3/8
<i>Chloromonas</i> sp. AAM2	Chlorophyta	Scaffold	102,365,718	32,535	16,278	53.1	GCA_004335635.1	2019/3/8
<i>Chlorophyta</i> sp.	Chlorophyta	Contig	29,100,557	7,000	7,000	53.23	GCA_007760615.1	2019/7/31

<i>Chloropicon primus</i>	Chlorophyta	Complete	17,584,162	1,118,318	1,118,318	57.52	GCA_023205875.1	2022/5/2
<i>Chloropicon primus</i>	Chlorophyta	Complete	17,400,691	1,089,399	1,089,399	57.52	GCA_007859695.1	2019/8/7
<i>Choricystis sp. odSpoLacu1</i>	Chlorophyta	Chromosome	14,467,306	996,781	673,657	69.3	GCA_958009055.1	2023/7/6
<i>Choricystis sp. odSpoLacu1</i>	Chlorophyta	Contig	14,609,959	673,657	673,657	69.3	GCA_958009065.1	2023/7/6
<i>Coccomyxa sp. LA000219</i>	Chlorophyta	Scaffold	48,546,496	2,254,067	694,571	54.7	GCA_000812005.1	2015/1/2
<i>Coccomyxa sp. Obi</i>	Chlorophyta	Chromosome	50,164,111	2,738,301	2,738,301	52.21	GCA_020887355.1	2021/10/5
<i>Coccomyxa sp. SUA001</i>	Chlorophyta	Contig	11,754,406	570	570	49.85	GCA_001244535.1	2015/8/5
<i>Coccomyxa subellipsoidea</i>	Chlorophyta	Scaffold	40,636,606	17,509	10,663	52.29	GCA_937857525.1	2023/1/12
<i>Coccomyxa subellipsoidea</i>	Chlorophyta	Scaffold	38,326,472	13,843	8,981	55.53	GCA_937857645.1	2023/1/12
<i>Coccomyxa subellipsoidea C-169</i>	Chlorophyta	Contig	48,826,616	1,959,569	1,959,569	52.94	GCF_000258705.1	2012/4/13
<i>Coelastrella sp. M60</i>	Chlorophyta	Contig	80,221,759	9,337	9,337	51.85	GCA_001630525.1	2016/4/25
<i>Coelastrella sp. MACC-549</i>	Chlorophyta	Scaffold	75,853,206	127,171	74,383	51.23	GCA_018290735.1	2021/5/13
<i>Coelastrella sp. UTEX B 3026</i>	Chlorophyta	Scaffold	151,301,762	10,679	7,135	51.97	GCA_002588565.1	2017/10/18
<i>Cymbomonas tetramitiformis</i>	Chlorophyta	Scaffold	281,270,454	10,932	5,748	52.52	GCA_001247695.1	2015/8/5
<i>Desmodesmus armatus</i>	Chlorophyta	Contig	116,304,009	341,806	341,806	56.64	GCA_007449985.2	2020/2/10
<i>Desmodesmus armatus</i>	Chlorophyta	Contig	110,532,431	217,430	217,430	56.64	GCA_021209675.1	2021/12/14
<i>Dunaliella primolecta</i>	Chlorophyta	Chromosome	210,950,326	12,721,864	2,890,256	48.87	GCA_914767535.2	2023/7/14
<i>Dunaliella primolecta</i>	Chlorophyta	Contig	20,300,542	4,968,441	4,968,441	48.87	GCA_914767515.2	2023/7/14
<i>Dunaliella salina</i>	Chlorophyta	Scaffold	343,704,392	353,034	7,231	49.08	GCA_002284615.2	2020/7/10
<i>Dunaliella sp. M2</i>	Chlorophyta	Scaffold	25,538,156	32,757	4,649	54.51	GCA_004335885.1	2019/3/8
<i>Dunaliella sp. RO</i>	Chlorophyta	Scaffold	121,879,027	8,610	4,107	48.47	GCA_004335775.1	2019/3/8
<i>Dunaliella sp. WIN1</i>	Chlorophyta	Scaffold	125,752,661	26,979	3,419	50.85	GCA_004335645.1	2019/3/8
<i>Dunaliella sp. YS1</i>	Chlorophyta	Scaffold	91,250,106	5,862	5,862	51.07	GCA_004335685.1	2019/3/8
<i>Edaphochlamys debaryana</i>	Chlorophyta	Contig	142,143,002	726,162	726,162	66.7	GCA_016858145.1	2021/2/9
<i>Edaphochlamys debaryana</i>	Chlorophyta	Scaffold	120,364,444	27,219	9,525	66.7	GCA_001662405.1	2016/4/26
<i>Ettlia oleoabundans</i>	Chlorophyta	Scaffold	48,907,151	21,691	16,270	58.19	GCA_019693395.1	2021/8/19
<i>Eudorina sp. 2006-703-Eu-15</i>	Chlorophyta	Scaffold	184,032,255	564,035	300,530	61.37	GCA_003117195.1	2018/4/6
<i>Eudorina sp. 2006-703-Eu-15</i>	Chlorophyta	Scaffold	168,620,790	377,357	341,314	61.37	GCA_003117095.1	2018/4/6
<i>Gonium pectorale</i>	Chlorophyta	Scaffold	148,806,073	1,267,136	16,298	64.53	GCA_001584585.1	2016/3/9
<i>Haematococcus lacustris</i>	Chlorophyta	Scaffold	309,350,987	9,426,245	144,907	58.37	GCA_030144725.1	2023/6/1
<i>Haematococcus lacustris</i>	Chlorophyta	Contig	365,773,732	11,925	11,925	58.37	GCA_003970955.1	2018/12/17
<i>Haematococcus lacustris</i>	Chlorophyta	Scaffold	171,794,631	38,941	5,283	58.37	GCA_011766145.1	2020/3/23
<i>Haematococcus sp. NG2</i>	Chlorophyta	Scaffold	77,829,956	591,061	47,126	58.23	GCA_004335575.1	2019/3/8
<i>Helicosporidium sp. ATCC 50920</i>	Chlorophyta	Contig	12,373,820	3,036	3,036	61.69	GCA_000690575.1	2014/5/13
<i>Mamiellaceae sp.</i>	Chlorophyta	Scaffold	3,152,822	8,500	7,376	47.5	GCA_029297915.1	2023/3/20
<i>Mamiellophyceae sp. 2017MT</i>	Chlorophyta	Scaffold	10,144,505	5,937	5,514	55.28	GCA_004115355.1	2019/1/28
<i>Messastrum gracile</i>	Chlorophyta	Scaffold	60,662,915	2,922	2,844	70	GCA_008037345.1	2019/8/20
<i>Micractinium conductrix</i>	Chlorophyta	Contig	60,773,875	1,210,495	1,210,495	67.2	GCA_002245815.2	2018/3/21
<i>Micractinium sp. NFX-FRZ</i>	Chlorophyta	Scaffold	68,098,937	115,803	101,544	65.17	GCA_029339195.1	2023/3/22
<i>Microglana sp. YARC</i>	Chlorophyta	Chromosome	572,689,865	94,760,622	141,564	48.63	GCA_022414725.1	2022/3/1
<i>Micromonas commoda</i>	Chlorophyta	Complete	20,989,326	1,394,110	1,394,110	63.82	GCA_000090985.2	2009/4/10
<i>Micromonas commoda</i>	Chlorophyta	Contig	18,874,706	43,253	43,253	63.82	GCA_937871665.1	2023/1/12
<i>Micromonas commoda</i>	Chlorophyta	Contig	18,568,673	27,986	27,986	63.82	GCA_937873045.1	2023/1/12
<i>Micromonas commoda</i>	Chlorophyta	Contig	17,521,099	27,831	27,831	63.82	GCA_937871525.1	2023/1/12
<i>Micromonas commoda</i>	Chlorophyta	Contig	17,248,016	24,650	24,650	63.82	GCA_937858595.1	2023/1/12

<i>Micromonas commoda</i>	Chlorophyta	Contig	16,695,881	16,526	16,526	63.82	GCA_937870615.1	2023/1/12
<i>Micromonas commoda</i>	Chlorophyta	Contig	12,885,859	14,962	14,962	63.82	GCA_937872605.1	2023/1/12
<i>Micromonas commoda</i>	Chlorophyta	Contig	9,753,388	8,671	8,671	63.82	GCA_937873085.1	2023/1/12
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	12,411,268	8,846	7,111	63.82	GCA_029293355.1	2023/3/20
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	13,697,947	12,873	6,997	63.82	GCA_029293475.1	2023/3/20
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	13,554,229	13,067	6,942	63.82	GCA_029293385.1	2023/3/20
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	12,083,793	8,485	6,722	63.82	GCA_029293545.1	2023/3/20
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	12,342,822	8,654	6,450	63.82	GCA_029293435.1	2023/3/20
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	11,972,191	8,272	6,253	63.82	GCA_029293455.1	2023/3/20
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	10,788,352	7,855	6,030	63.82	GCA_029293395.1	2023/3/20
<i>Micromonas commoda</i>	Chlorophyta	Scaffold	9,949,361	6,636	5,800	63.82	GCA_029290755.1	2023/3/20
<i>Micromonas pusilla</i>	Chlorophyta	Contig	16,973,892	12,507	12,507	67.66	GCA_937863545.1	2023/1/12
<i>Micromonas pusilla</i> CCMP1545	Chlorophyta	Scaffold	21,958,260	1,183,541	84,001	65.9	GCF_000151265.2	2009/4/9
<i>Micromonas sp. ASP10-01a</i>	Chlorophyta	Scaffold	19,582,411	22,484	12,133	59.38	GCA_001430725.1	2015/4/28
<i>Micromonas sp. SAG5</i>	Chlorophyta	Contig	957,347	2,516	2,516	39.8	GCA_009618055.1	2019/11/7
<i>Micromonas sp. SAG6</i>	Chlorophyta	Contig	614,822	1,837	1,837	39.31	GCA_009618155.1	2019/11/7
<i>Micromonas sp. SAG7</i>	Chlorophyta	Contig	609,943	7,399	7,399	41.55	GCA_009618165.1	2019/11/7
<i>Micromonas sp. SAG8</i>	Chlorophyta	Contig	558,822	6,768	6,768	38.84	GCA_009618195.1	2019/11/7
<i>Monoraphidium minutum</i>	Chlorophyta	Scaffold	68,170,630	259,450	259,450	72.06	GCA_025201885.1	2022/9/9
<i>Monoraphidium neglectum</i>	Chlorophyta	Scaffold	69,482,627	15,604	9,120	64.75	GCA_000611645.1	2015/2/26
<i>Monoraphidium sp. 549</i>	Chlorophyta	Contig	74,658,904	105,989	105,989	51.24	GCA_002814315.1	2017/12/8
<i>Mychonastes homosphaera</i>	Chlorophyta	Contig	24,204,206	2,001,712	2,001,712	72.43	GCA_009193075.1	2019/10/21
<i>Nannochloris sp. RS</i>	Chlorophyta	Scaffold	13,562,322	1,176,250	10,691	46.39	GCA_004335565.1	2019/3/8
<i>Nannochloris sp. X1</i>	Chlorophyta	Scaffold	13,742,038	52,627	5,621	46.03	GCA_004335555.1	2019/3/8
<i>Ostreobium quekettii</i>	Chlorophyta	Scaffold	151,902,089	71,032	69,688	52.41	GCA_905146915.1	2021/9/14
<i>Ostreococcus lucimarinus</i> CCE9901	Chlorophyta	Complete	13,204,888	708,927	708,927	60.44	GCA_000092065.1	2007/4/10
<i>Ostreococcus mediterraneus</i>	Chlorophyta	Scaffold	13,861,019	800,924	235,059	56.02	GCA_012295225.1	2020/4/15
<i>Ostreococcus sp. lucimarinus</i>	Chlorophyta	Contig	12,485,144	47,709	47,709	60.31	GCA_937863055.1	2023/1/12
<i>Ostreococcus sp. lucimarinus</i>	Chlorophyta	Contig	11,530,018	28,262	28,262	60.31	GCA_937863655.1	2023/1/12
<i>Ostreococcus sp. SAG10</i>	Chlorophyta	Contig	3,514,295	32,866	32,866	57.03	GCA_009618215.1	2019/11/7
<i>Ostreococcus sp. SAG11</i>	Chlorophyta	Contig	1,583,702	16,554	16,554	52	GCA_009618255.1	2019/11/7
<i>Ostreococcus sp. SAG12</i>	Chlorophyta	Contig	2,758,302	36,467	36,467	56.66	GCA_009618245.1	2019/11/7
<i>Ostreococcus sp. SAG9</i>	Chlorophyta	Contig	884,776	6,240	6,240	51.96	GCA_009618205.1	2019/11/7
<i>Ostreococcus tauri</i>	Chlorophyta	Chromosome	12,916,858	770,246	48,024	59.38	GCA_000214015.2	2014/10/2
<i>Ostreococcus tauri</i>	Chlorophyta	Scaffold	13,800,296	620,342	620,342	59.38	GCA_905339685.1	2021/3/26
<i>Ostreococcus tauri</i>	Chlorophyta	Scaffold	14,762,682	626,729	572,890	59.38	GCA_002158475.1	2017/5/26
<i>Ostreococcus tauri</i>	Chlorophyta	Scaffold	13,845,333	623,690	552,991	59.38	GCA_905339615.1	2021/3/26
<i>Ostreococcus tauri</i>	Chlorophyta	Contig	21,542,177	337,319	337,319	59.38	GCA_905339635.1	2021/3/26
<i>Ostreococcus tauri</i>	Chlorophyta	Scaffold	20,728,069	222,572	222,572	59.38	GCA_905339545.1	2021/3/26
<i>Ostreococcus tauri</i>	Chlorophyta	Contig	17,241,412	178,604	178,604	59.38	GCA_905339625.1	2021/3/26
<i>Parachlorella kessleri</i>	Chlorophyta	Contig	59,187,803	33,885	33,885	58.33	GCA_001598975.1	2015/12/19
<i>Parachlorella kessleri</i>	Chlorophyta	Contig	64,921,412	463,012	463,012	58.33	GCA_015712045.1	2020/12/2
<i>Pedinophyceae sp. YPF-701</i>	Chlorophyta	Contig	27,899,919	1,232,795	1,232,795	69.81	GCA_916048055.1	2022/2/18
<i>Picochlorum costavernella</i>	Chlorophyta	Scaffold	14,533,975	691,353	691,049	45.93	GCA_011316045.1	2020/3/17
<i>Picochlorum sp. BH-2019</i>	Chlorophyta	Contig	14,391,424	1,122,049	1,122,049	46.15	GCA_009650465.1	2019/11/12

<i>Picochlorum</i> sp. BPE23	Chlorophyta	Contig	14,824,110	1,257,599	1,257,599	46.19	GCA_025209345.1	2022/9/15
<i>Picochlorum</i> sp. BPE23	Chlorophyta	Contig	13,436,362	1,109,381	1,109,381	46.19	GCA_025209375.1	2022/9/14
<i>Picochlorum</i> sp. celeri	Chlorophyta	Scaffold	27,430,220	1,163,840	1,125,495	46.23	GCA_010909725.1	2020/2/20
<i>Picochlorum</i> sp. SENEW3	Chlorophyta	Scaffold	13,278,802	126,215	126,215	46.09	GCA_000876415.1	2014/7/28
<i>Picochlorum</i> sp. soloecismus	Chlorophyta	Scaffold	15,141,037	724,710	621,321	44.34	GCA_002818215.1	2017/12/8
<i>Picocystis</i> sp. L7	Chlorophyta	Contig	20,367,036	729,631	729,631	53.66	GCA_020975725.1	2021/11/30
<i>Picocystis</i> sp. ML	Chlorophyta	Contig	29,463,432	153,902	153,902	53.56	GCA_003665715.1	2018/10/16
<i>Pleodorina starrii</i>	Chlorophyta	Contig	136,124,605	3,124,857	3,124,857	62.82	GCA_030267565.1	2023/5/19
<i>Pleodorina starrii</i>	Chlorophyta	Contig	134,838,970	1,064,554	1,064,554	62.82	GCA_030267485.1	2023/5/19
<i>Pleodorina starrii</i>	Chlorophyta	Contig	134,979,981	833,246	833,246	62.82	GCA_030267705.1	2023/5/19
<i>Prasinocladus</i> sp. malaysianus	Chlorophyta	Scaffold	88,256,521	601	570	45.82	GCA_019702555.1	2021/8/20
<i>Prasinococcaceae</i> sp. CCMP 1430	Chlorophyta	Scaffold	28,571,423	305,482	21,851	66.96	GCA_013416795.1	2020/7/16
<i>Protoeuglena noctilucae</i>	Chlorophyta	Chromosome	26,184,876	1,046,115	275,497	66.18	GCA_958009085.1	2023/7/6
<i>Protoeuglena noctilucae</i>	Chlorophyta	Contig	20,113,892	338,936	338,936	66.18	GCA_958009145.1	2023/7/6
<i>Prototheca bovis</i>	Chlorophyta	Contig	24,744,895	7,940	7,940	73.53	GCA_003612995.1	2018/10/4
<i>Prototheca ciferrii</i>	Chlorophyta	Scaffold	26,448,891	7,016	5,354	67.62	GCA_003613005.1	2018/10/4
<i>Prototheca cutis</i>	Chlorophyta	Scaffold	20,029,045	1,409,608	58,511	61.23	GCA_002897115.2	2018/6/27
<i>Prototheca cutis</i>	Chlorophyta	Contig	19,388,884	5,802	5,802	61.23	GCA_016906445.1	2021/2/18
<i>Prototheca stagnorum</i>	Chlorophyta	Scaffold	16,896,228	1,107,247	34,340	73.31	GCA_002794665.1	2017/11/16
<i>Prototheca wickerhamii</i>	Chlorophyta	Scaffold	16,700,629	1,578,614	5,022	62.97	GCA_016906385.1	2021/2/18
<i>Prototheca wickerhamii</i>	Chlorophyta	Scaffold	27,590,722	31,071	7,951	62.97	GCA_003255715.1	2018/6/19
<i>Pycnococcus provasolii</i>	Chlorophyta	Chromosome	32,229,639	757,545	728,549	57.51	GCA_938743325.1	2022/5/3
<i>Pycnococcus provasolii</i>	Chlorophyta	Contig	22,700,889	534,558	534,558	57.51	GCA_015473125.1	2020/11/11
<i>Pycnococcus provasolii</i>	Chlorophyta	Scaffold	43,744,632	198,061	198,061	57.51	GCA_938743315.1	2022/5/2
<i>Pyramimonas parkeae</i>	Chlorophyta	Contig	249,414,800	3,128	3,128	46.86	GCA_018697155.1	2021/6/3
<i>Raphidocelis subcapitata</i>	Chlorophyta	Scaffold	51,162,697	341,804	91,765	71.62	GCA_003203535.1	2018/5/30
<i>Scenedesmus quadricauda</i>	Chlorophyta	Contig	65,353,879	8,094	8,094	63.24	GCA_002317545.1	2017/9/20
<i>Scenedesmus</i> sp. ARA	Chlorophyta	Scaffold	93,238,253	79,822	37,561	56.99	GCA_004335915.1	2019/3/8
<i>Scenedesmus</i> sp. KT-U	Chlorophyta	Contig	112,698,165	6,459	6,459	54.55	GCA_029853055.1	2023/4/21
<i>Scenedesmus</i> sp. NREL 46B-D3	Chlorophyta	Contig	151,896,789	204,857	204,857	57.26	GCA_014080715.1	2020/8/5
<i>Scenedesmus</i> sp. PABB004	Chlorophyta	Chromosome	39,756,104	1,281,093	1,281,093	78.38	GCA_014905635.1	2020/10/28
<i>Scenedesmus vacuolatus</i>	Chlorophyta	Contig	23,358,248	1,571	1,571	53.62	GCA_004764505.1	2019/4/9
<i>Tetraabaena socialis</i>	Chlorophyta	Scaffold	135,779,852	145,927	7,414	66.05	GCA_002891735.1	2018/1/18
<i>Tetradasmus acuminatus</i>	Chlorophyta	Scaffold	119,241,191	87,113	81,888	54.96	GCA_902809745.2	2020/4/27
<i>Tetradasmus obliquus</i>	Chlorophyta	Complete	104,665,984	6,124,559	6,124,559	56.82	GCA_030272155.1	2023/6/14
<i>Tetradasmus obliquus</i>	Chlorophyta	Complete	100,347,600	6,027,792	6,027,792	56.82	GCA_030272055.1	2023/6/14
<i>Tetradasmus obliquus</i>	Chlorophyta	Scaffold	107,715,903	186,615	151,600	56.82	GCA_900108755.1	2016/11/5
<i>Tetradasmus obliquus</i>	Chlorophyta	Contig	207,967,116	155,544	155,544	56.82	GCA_002149895.1	2017/5/19
<i>Tetradasmus obliquus</i>	Chlorophyta	Contig	172,339,266	94,410	94,410	56.82	GCA_014621415.1	2020/9/17
<i>Tetraselmis chuii</i>	Chlorophyta	Scaffold	362,171,450	155,666	153,541	53.03	GCA_026401115.1	2022/11/25
<i>Tetraselmis striata</i>	Chlorophyta	Contig	227,954,216	126,807	126,807	57.91	GCA_006384855.1	2019/6/21
<i>Tetratostichococcus</i> sp. P1	Chlorophyta	Contig	51,992,686	6,365	6,365	56.65	GCA_010646915.1	2020/2/14
<i>Trebouxia gelatinosa</i>	Chlorophyta	Scaffold	61,726,235	3,512,598	951,886	50.4	GCA_000818905.1	2015/1/16
<i>Trebouxia</i> sp. A1-2	Chlorophyta	Scaffold	52,896,788	848,255	151,515	49.99	GCA_008636185.1	2019/9/23
<i>Trebouxia</i> sp. A1-2	Chlorophyta	Scaffold	55,594,867	5,332	3,203	49.99	GCA_937857455.1	2023/1/12

<i>Trebouxia</i> sp. TZW2008	Chlorophyta	Scaffold	69,346,529	223,445	184,823	49.69	GCA_002118135.1	2017/3/31
<i>Trebouxiophyceae</i> sp. KSI-1	Chlorophyta	Scaffold	44,686,223	21,380	16,541	64.63	GCA_003568905.1	2018/8/29
<i>Trebouxiophyceae</i> sp. M-hakoo 311	Chlorophyta	Complete	15,638,232	1,246,908	1,246,908	73.07	GCA_026636295.1	2022/12/5
<i>Ulva compressa</i>	Chlorophyta	Scaffold	80,815,439	458,211	10,369	57.31	GCA_024500015.1	2022/8/1
<i>Ulva mutabilis</i>	Chlorophyta	Scaffold	98,484,689	600,008	527,412	57.16	GCA_900538255.1	2018/9/12
<i>Ulva prolifera</i>	Chlorophyta	Chromosome	88,938,772	11,195,790	2,057,339	55.95	GCA_023078555.1	2022/4/22
<i>Ulva prolifera</i>	Chlorophyta	Contig	103,843,212	4,106,524	4,106,524	55.95	GCA_023707965.1	2022/5/17
<i>Ulva prolifera</i>	Chlorophyta	Scaffold	87,889,290	675,063	391,640	55.95	GCA_004138255.1	2019/2/5
<i>uncultured Bathycoccus</i>	Chlorophyta	Contig	14,194,380	57,242	57,242	48.18	GCA_937871675.1	2023/1/12
<i>uncultured Bathycoccus</i>	Chlorophyta	Contig	5,183,084	44,020	44,020	48.18	GCA_000259855.1	2011/10/31
<i>uncultured Bathycoccus</i>	Chlorophyta	Scaffold	13,196,014	18,534	10,754	48.18	GCA_903798045.1	2020/8/12
<i>uncultured Bathycoccus</i>	Chlorophyta	Contig	12,952,815	5,388	5,388	48.18	GCA_937869125.1	2023/1/12
<i>uncultured Bathycoccus</i>	Chlorophyta	Contig	12,696,268	3,483	3,483	48.18	GCA_937863585.1	2023/1/12
<i>uncultured Chlorellales</i>	Chlorophyta	Contig	7,958,504	2,196	2,196	62.59	GCA_937867275.1	2023/1/12
<i>uncultured Trebouxiophyceae</i>	Chlorophyta	Contig	4,756,059	1,702	1,702	72.96	GCA_937872675.1	2023/1/12
<i>Volvox africanus</i>	Chlorophyta	Contig	141,009,466	3,954,136	3,954,136	52.91	GCA_030268105.1	2023/5/18
<i>Volvox africanus</i>	Chlorophyta	Contig	129,328,469	1,356,898	1,356,898	52.91	GCA_019650175.1	2021/3/2
<i>Volvox carteri f nagariensis</i>	Chlorophyta	Scaffold	137,684,403	1,491,501	43,981	55.97	GCA_000143455.1	2010/7/8
<i>Volvox reticuliferus</i>	Chlorophyta	Contig	133,065,142	1,907,605	1,907,605	53.93	GCA_019650235.1	2021/3/2
<i>Volvox reticuliferus</i>	Chlorophyta	Contig	133,961,728	1,866,906	1,866,906	53.93	GCA_019650255.1	2021/3/2
<i>Yamagishiella unicocca</i>	Chlorophyta	Contig	134,234,618	666,310	666,310	60.78	GCA_003116995.1	2018/4/6
<i>Yamagishiella unicocca</i>	Chlorophyta	Contig	140,837,241	543,037	543,037	60.78	GCA_003117035.1	2018/4/6
<i>Chara braunii</i>	Streptophyta	Scaffold	1,751,211,849	2,261,426	11,762	48.31	GCA_003427395.1	2018/7/13
<i>Chlorokybus atmophyticus</i>	Streptophyta	Scaffold	74,330,326	752,385	43,061	51.62	GCA_009103225.1	2019/10/15
<i>Closterium</i> sp. Naga37s-1	Streptophyta	Contig	975,230,015	2,838,063	2,838,063	57.09	GCA_947471745.1	2023/6/17
<i>Closterium</i> sp. NIES-64	Streptophyta	Scaffold	702,500,261	727,632	636,707	57.56	GCA_948144035.1	2023/6/17
<i>Closterium</i> sp. NIES-65	Streptophyta	Scaffold	876,339,800	1,233,613	1,206,981	57.3	GCA_948144175.1	2023/6/17
<i>Closterium</i> sp. NIES-67	Streptophyta	Contig	361,975,124	352,577	352,577	56.1	GCA_027943415.1	2022/12/27
<i>Closterium</i> sp. NIES-68	Streptophyta	Contig	338,860,182	276,113	276,113	55.81	GCA_027943385.1	2022/12/27
<i>Closterium</i> sp. Yama58-4	Streptophyta	Contig	437,979,428	952,287	952,287	56.74	GCA_947470075.1	2023/6/17
<i>Klebsormidium nitens</i>	Streptophyta	Scaffold	103,921,766	134,930	72,512	52.42	GCA_000708835.1	2014/6/3
<i>Mesostigma viride</i>	Streptophyta	Scaffold	441,698,809	2,558,729	319,906	54.32	GCA_009746045.1	2019/12/12
<i>Mesostigma viride</i>	Streptophyta	Scaffold	281,412,074	113,221	13,708	54.32	GCA_009103195.1	2019/10/15
<i>Mesostigma viride</i>	Streptophyta	Contig	145,898,619	3,263	3,263	54.32	GCA_004000685.1	2018/12/15
<i>Mesotaenium endlicherianum</i>	Streptophyta	Scaffold	173,751,473	448,375	19,321	51.87	GCA_009602735.1	2019/11/6
<i>Penium margaritaceum</i>	Streptophyta	Scaffold	3,661,744,145	116,159	10,892	51.1	GCA_023375785.1	2022/5/16
<i>Spirogloea muscicola</i>	Streptophyta	Scaffold	170,819,335	566,364	18,068	56.37	GCA_009602725.1	2019/11/6
<i>Asparagopsis taxiformis</i>	Rhodophyta	Contig	142,369,820	55,006	55,006	47.7	GCA_030407315.1	2023/7/5
<i>Asparagopsis taxiformis</i>	Rhodophyta	Scaffold	46,695,008	11,264	10,700	47.7	GCA_018397955.1	2021/5/18
<i>Asparagopsis taxiformis</i>	Rhodophyta	Scaffold	75,200,346	8,939	7,579	47.7	GCA_018397975.1	2021/5/18
<i>Chondria armata</i>	Rhodophyta	Scaffold	507,583,893	643,002	514,398	45.33	GCA_021653855.1	2022/1/27
<i>Chondrus crispus</i>	Rhodophyta	Scaffold	104,800,334	242,694	77,654	52.88	GCA_000350225.2	2013/5/22
<i>Cyanidiococcus yangmingshanensis</i>	Rhodophyta	Contig	12,072,675	653,084	653,084	54.61	GCA_013995675.1	2020/7/31
<i>Cyanidiococcus yangmingshanensis</i>	Rhodophyta	Contig	12,020,148	644,466	644,466	54.61	GCA_026122185.1	2022/11/10
<i>Cyanidioschyzon merolae</i>	Rhodophyta	Contig	12,115,884	616,115	616,115	55.02	GCA_006232315.1	2019/6/11

<i>Cyanidioschyzon merolae</i>	Rhodophyta	Contig	16,248,422	17,785	17,785	55.02	GCA_011170005.1	2020/2/6
<i>Cyanidioschyzon merolae</i>	Rhodophyta	Contig	16,233,476	17,500	17,500	55.02	GCA_010725195.1	2020/2/6
<i>Cyanidium caldarium</i>	Rhodophyta	Contig	8,787,356	488,039	488,039	40.32	GCA_026184775.1	2022/11/15
<i>Cyanidium caldarium</i>	Rhodophyta	Scaffold	11,413,374	13,838	13,324	40.32	GCA_019693505.1	2021/8/19
<i>Digenea simplex</i>	Rhodophyta	Contig	299,321,041	57,705	57,705	43.34	GCA_004798425.1	2019/4/16
<i>Galdieria phlegrea</i>	Rhodophyta	Contig	14,577,017	201,071	201,071	37.52	GCA_006232345.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Scaffold	13,712,004	172,322	116,786	37.7	GCF_000341285.1	2013/2/25
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	13,833,850	189,411	189,411	37.7	GCA_006232505.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	14,008,042	188,305	188,305	37.7	GCA_006232475.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	14,663,505	186,619	186,619	37.7	GCA_006232405.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	15,270,598	182,300	182,300	37.7	GCA_006232545.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	14,644,345	172,087	172,087	37.7	GCA_006232515.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	14,197,739	170,797	170,797	37.7	GCA_006232335.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	13,914,336	169,495	169,495	37.7	GCA_024341865.1	2022/6/15
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	13,744,627	165,034	165,034	37.7	GCA_006232395.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	14,073,046	158,217	158,217	37.7	GCA_006232365.1	2019/6/11
<i>Galdieria sulphuraria</i>	Rhodophyta	Contig	12,093,695	134,001	134,001	37.7	GCA_001704855.1	2016/8/14
<i>Galdieria sulphuraria</i>	Rhodophyta	Scaffold	14,584,420	6,906	3,832	37.7	GCA_019693475.1	2021/8/19
<i>Gracilaria domingensis</i>	Rhodophyta	Scaffold	77,739,776	189,322	132,761	50.23	GCA_022539475.1	2022/3/14
<i>Gracilariopsis chorda</i>	Rhodophyta	Contig	92,180,038	220,274	220,274	49.27	GCA_003194525.1	2018/6/6
<i>Gracilariopsis lemaneiformis</i>	Rhodophyta	Scaffold	88,693,116	34,594	29,144	49.94	GCA_003346895.1	2018/7/31
<i>Kappaphycus alvarezii</i>	Rhodophyta	Contig	336,052,185	849,038	849,038	45.32	GCA_002205965.3	2020/11/6
<i>Neoporphyra haitanensis</i>	Rhodophyta	Scaffold	53,254,617	5,758,810	1,068,871	71.26	GCA_008729055.1	2019/9/30
<i>Neoporphyra haitanensis</i>	Rhodophyta	Chromosome	49,666,066	7,796,351	650,156	71.26	GCA_021651115.1	2022/1/31
<i>Neopyropia yezoensis</i>	Rhodophyta	Chromosome	107,590,730	34,332,347	202,496	64.92	GCA_009829735.1	2020/1/6
<i>Porphyra umbilicalis</i>	Rhodophyta	Scaffold	87,698,808	202,021	171,716	65.7	GCA_002049455.2	2017/7/27
<i>Porphyridium purpureum</i>	Rhodophyta	Contig	22,192,338	1,853,716	1,853,716	55.91	GCA_008690995.1	2019/9/24
<i>Porphyridium purpureum</i>	Rhodophyta	Contig	19,451,899	20,534	20,534	55.91	GCA_000397085.1	2013/5/16
<i>Porphyridium purpureum</i>	Rhodophyta	Scaffold	19,672,903	18,818	11,707	55.91	GCA_019702435.1	2021/8/19
<i>Amoebophrya</i> sp. A120	Alveolata	Scaffold	115,520,608	9,243,444	339,444	51.23	GCA_905178155.1	2021/9/28
<i>Amoebophrya</i> sp. A25	Alveolata	Scaffold	116,004,390	1,024,096	223,917	47.79	GCA_905178165.1	2021/9/28
<i>Amoebophrya</i> sp. AT52	Alveolata	Contig	85,557,643	12,329	12,329	55.9	GCA_005223375.1	2019/5/7
<i>Amphidinium carterae</i>	Alveolata	Scaffold	478,166,117	558	558	46.45	GCA_019702695.1	2021/8/20
<i>Chromera velia</i>	Alveolata	Contig	187,454,854	2,868	2,868	49.94	GCA_000585135.1	2014/3/3
<i>Chromera velia</i> CCMP2878	Alveolata	Scaffold	193,672,810	57,903	31,704	49.1	GCA_018398765.1	2021/5/19
<i>Cladocopium goreau</i>	Alveolata	Scaffold	1,171,456,478	353,897	218,609	44.39	GCA_947184155.1	2022/10/28
<i>Effrenium voratum</i>	Alveolata	Scaffold	1,158,599,799	721,460	369,669	50.77	GCA_963377175.1	2023/8/24
<i>Effrenium voratum</i>	Alveolata	Scaffold	1,322,947,892	252,875	150,174	50.77	GCA_963377275.1	2023/8/24
<i>Effrenium voratum</i>	Alveolata	Scaffold	1,057,836,601	304,737	18,586	50.77	GCA_963377065.1	2023/8/23
<i>Polarella glacialis</i>	Alveolata	Scaffold	2,984,680,192	170,304	144,622	45.91	GCA_905237085.1	2021/2/20
<i>Polarella glacialis</i>	Alveolata	Scaffold	2,756,104,381	129,205	101,691	45.91	GCA_905237095.1	2021/2/20
<i>Prorocentrum minimum</i>	Alveolata	Contig	29,349,011	2,528	2,528	56.99	GCA_001652855.1	2016/5/26
<i>Symbiodinium kawagutii</i>	Alveolata	Scaffold	935,067,369	380,908	35,631	45.72	GCA_009767595.1	2019/12/18
<i>Symbiodinium microadriaticum</i>	Alveolata	Scaffold	808,226,712	573,512	33,339	50.51	GCA_001939145.1	2017/1/6
<i>Symbiodinium microadriaticum</i>	Alveolata	Scaffold	692,042,016	150,686	37,868	50.51	GCA_018327485.1	2021/3/19

<i>Symbiodinium microadriaticum</i>	Alveolata	Scaffold	775,008,844	49,975	11,136	50.51	GCA_905231925.1	2021/2/20
<i>Symbiodinium natans</i>	Alveolata	Scaffold	761,619,964	610,496	358,021	51.79	GCA_905221605.1	2021/2/20
<i>Symbiodinium necroappetens</i>	Alveolata	Scaffold	767,953,253	14,528	11,421	50.85	GCA_905231915.1	2021/2/20
<i>Symbiodinium pilosum</i>	Alveolata	Scaffold	1,089,424,773	62,444	17,505	48.21	GCA_905231905.1	2021/2/20
<i>Symbiodinium sp. CCMP2456</i>	Alveolata	Scaffold	694,902,460	58,075	11,149	50.36	GCA_905221635.1	2021/2/20
<i>Symbiodinium sp. CCMP2592</i>	Alveolata	Scaffold	1,103,301,044	651,264	356,695	51.01	GCA_905221615.1	2021/2/20
<i>Symbiodinium sp. clade A Y106</i>	Alveolata	Scaffold	766,659,703	133,469	52,831	50.43	GCA_003297005.1	2018/6/22
<i>Symbiodinium sp. clade C Y103</i>	Alveolata	Scaffold	703,698,624	249,182	38,666	44.73	GCA_003297045.1	2018/6/22
<i>Symbiodinium sp. KB8</i>	Alveolata	Scaffold	813,744,491	42,989	10,401	51.91	GCA_905221625.1	2021/2/20
<i>Vitrella brassicaformis CCMP3155</i>	Alveolata	Scaffold	72,700,666	147,631	33,811	58.09	GCA_001179505.1	2015/7/8
<i>Asterionella formosa</i>	Bacillariophyta	Scaffold	68,419,761	15,906	13,432	41.89	GCA_002256025.1	2017/8/16
<i>Asterionellopsis glacialis</i>	Bacillariophyta	Scaffold	66,879,591	21,950	11,523	39.95	GCA_014885115.2	2020/11/10
<i>Bacillariophyta sp.</i>	Bacillariophyta	Scaffold	37,632,609	22,224	20,820	44.93	GCA_010367165.1	2020/2/10
<i>Bacillariophyta sp.</i>	Bacillariophyta	Scaffold	23,569,900	12,848	12,833	44.93	GCA_020048575.1	2021/9/20
<i>Bacillariophyta sp.</i>	Bacillariophyta	Scaffold	19,794,973	10,902	10,715	44.93	GCA_020048535.1	2021/9/20
<i>Bacillariophyta sp.</i>	Bacillariophyta	Scaffold	41,386,207	8,319	8,055	44.93	GCA_013299085.1	2020/6/8
<i>Bacillariophyta sp.</i>	Bacillariophyta	Scaffold	21,036,362	6,478	5,963	44.93	GCA_010367175.1	2020/2/10
<i>Bacillariophyta sp.</i>	Bacillariophyta	Contig	96,830,142	2,569	2,569	44.93	GCA_023347125.1	2022/5/12
<i>Bacillariophyta sp.</i>	Bacillariophyta	Contig	19,038,625	2,260	2,260	44.93	GCA_023345945.1	2022/5/12
<i>Bacillariophyta sp.</i>	Bacillariophyta	Contig	6,029,054	2,006	2,006	44.93	GCA_023346715.1	2022/5/12
<i>Bacillariophyta sp.</i>	Bacillariophyta	Contig	15,632,304	1,917	1,917	44.93	GCA_023346925.1	2022/5/12
<i>Bacillariophyta sp.</i>	Bacillariophyta	Contig	14,327,404	1,638	1,638	44.93	GCA_023346865.1	2022/5/12
<i>Bacillariophyta sp.</i>	Bacillariophyta	Contig	26,376,286	1,610	1,610	44.93	GCA_023346745.1	2022/5/12
<i>Chaetoceros muellerii</i>	Bacillariophyta	Scaffold	37,735,205	21,121	16,859	41.48	GCA_019693545.1	2021/8/19
<i>Chaetoceros tenuissimus</i>	Bacillariophyta	Scaffold	41,001,288	2,427,249	2,121,213	38.83	GCA_021927905.1	2021/11/26
<i>Craspedostauros australis</i>	Bacillariophyta	Contig	74,631,628	1,724,159	1,724,159	53.14	GCA_026770025.1	2022/12/9
<i>Craspedostauros australis</i>	Bacillariophyta	Scaffold	65,888,406	30,927	28,326	53.14	GCA_030167865.1	2023/5/24
<i>Cyclotella cryptica</i>	Bacillariophyta	Scaffold	171,139,483	494,169	284,001	43.11	GCA_013187285.1	2020/5/29
<i>Cylindrotheca fusiformis</i>	Bacillariophyta	Scaffold	48,249,611	17,542	16,610	45.74	GCA_019693525.1	2021/8/19
<i>Epithemia pelagica</i>	Bacillariophyta	Chromosome	60,349,725	3,856,736	2,230,900	48.14	GCA_946965045.2	2023/4/7
<i>Epithemia pelagica</i>	Bacillariophyta	Contig	61,867,837	565,830	565,830	48.14	GCA_946965055.2	2023/4/8
<i>Fistulifera pelliculosa</i>	Bacillariophyta	Contig	35,793,069	176,105	176,105	49.16	GCA_026008555.1	2022/7/26
<i>Fistulifera pelliculosa</i>	Bacillariophyta	Scaffold	30,158,221	20,565	18,365	49.16	GCA_019693425.1	2021/8/19
<i>Fistulifera solaris</i>	Bacillariophyta	Chromosome	51,201,953	1,355,535	1,355,535	46.18	GCA_030295235.1	2022/8/1
<i>Fistulifera solaris</i>	Bacillariophyta	Scaffold	49,736,633	330,809	75,175	46.18	GCA_002217885.1	2017/6/26
<i>Fragilaria crotonensis</i>	Bacillariophyta	Contig	61,849,504	89,148	89,148	47.33	GCA_022925895.1	2022/4/12
<i>Fragilaria radians</i>	Bacillariophyta	Scaffold	98,377,760	100,875	12,314	39.31	GCA_900642245.1	2019/2/6
<i>Fragilariophycidae sp.</i>	Bacillariophyta	Contig	22,731,055	11,747	11,747	43.44	GCA_023346985.1	2022/5/12
<i>Fragilariophycidae sp.</i>	Bacillariophyta	Contig	15,189,913	10,553	10,553	43.44	GCA_023346625.1	2022/5/12
<i>Fragilariophycidae sp.</i>	Bacillariophyta	Contig	3,576,877	5,812	5,812	43.44	GCA_023346165.1	2022/5/12
<i>Fragilariophycidae sp.</i>	Bacillariophyta	Contig	23,958,011	5,287	5,287	43.44	GCA_023347145.1	2022/5/12
<i>Fragilariophycidae sp.</i>	Bacillariophyta	Contig	7,524,140	3,497	3,497	43.44	GCA_023346885.1	2022/5/12
<i>Fragilariophycidae sp.</i>	Bacillariophyta	Contig	11,033,108	2,333	2,333	43.44	GCA_023346465.1	2022/5/12
<i>Fragilariophycidae sp.</i>	Bacillariophyta	Contig	5,359,268	2,047	2,047	43.44	GCA_023346305.1	2022/5/12

<i>Fragilariophycidae</i> sp.	Bacillariophyta	Contig	10,936,309	1,579	1,579	43.44	GCA_023347105.1	2022/5/12
<i>Fragilariopsis cylindrus</i> CCMP1102	Bacillariophyta	Contig	68,974,517	190,760	190,760	38.51	GCA_900095095.1	2016/8/5
<i>Fragilariopsis cylindrus</i> CCMP1102	Bacillariophyta	Scaffold	80,540,407	1,295,603	78,226	38.51	GCA_001750085.1	2016/9/30
<i>Fragilariopsis kerguelensis</i>	Bacillariophyta	Complete	10,223	10,223	10,223	40.47	GCA_902825245.1	2020/6/28
<i>Halamphora</i> sp. AAB	Bacillariophyta	Scaffold	29,554,481	227,054	6,529	50.41	GCA_004335955.1	2019/3/8
<i>Halamphora</i> sp. MG8b	Bacillariophyta	Scaffold	53,206,049	1,555	1,430	46.6	GCA_004335815.1	2019/3/8
<i>Licmophora abbreviata</i>	Bacillariophyta	Contig	29,210,371	6,984	6,984	52.61	GCA_900291995.1	2018/4/24
<i>Mayamaea pseudoterrestris</i>	Bacillariophyta	Contig	30,627,229	1,067,527	1,067,527	49.52	GCA_027923505.1	2022/12/13
<i>Nitzschia inconspicua</i>	Bacillariophyta	Contig	99,706,970	3,618,388	3,618,388	45.37	GCA_019154785.2	2021/7/7
<i>Nitzschia palea</i>	Bacillariophyta	Scaffold	41,156,031	40,280	40,060	46.11	GCA_019593585.1	2021/8/13
<i>Nitzschia putrida</i>	Bacillariophyta	Contig	47,133,574	545,464	545,464	47.65	GCA_016586335.1	2020/12/9
<i>Nitzschia</i> sp. DOCU1	Bacillariophyta	Chromosome	51,966,319	3,685,028	1,543,651	48.9	GCA_951394045.1	2023/5/13
<i>Nitzschia</i> sp. DOCU1	Bacillariophyta	Scaffold	54,097,145	122,309	122,309	48.9	GCA_951393995.1	2023/5/13
<i>Nitzschia</i> sp. Nitz4	Bacillariophyta	Scaffold	27,229,399	91,515	62,281	48.02	GCA_013372465.1	2020/6/24
<i>Nitzschia</i> sp. pyKryTriq1	Bacillariophyta	Chromosome	40,484,363	2,856,296	1,332,194	48.16	GCA_956465555.1	2023/7/14
<i>Nitzschia</i> sp. pyKryTriq1	Bacillariophyta	Scaffold	39,990,705	243,925	243,925	48.16	GCA_956465565.1	2023/7/15
<i>Phaeodactylum tricornutum</i>	Bacillariophyta	Chromosome	27,820,512	1,122,386	1,122,386	48.83	GCA_914521175.1	2022/5/3
<i>Phaeodactylum tricornutum</i> CCAP 1055/1	Bacillariophyta	Chromosome	27,450,724	945,026	417,209	48.83	GCA_000150955.2	2008/12/12
<i>Psammoneis japonica</i>	Bacillariophyta	Scaffold	91,430,650	377,693	175,630	48.59	GCA_008632985.1	2019/9/23
<i>Pseudo-nitzschia multistriata</i>	Bacillariophyta	Scaffold	56,765,209	141,206	30,320	46.32	GCA_900660405.1	2019/2/12
<i>Seminavis robusta</i>	Bacillariophyta	Scaffold	125,572,603	50,704	48,501	48.37	GCA_903772945.1	2022/10/20
<i>Skeletonema costatum</i>	Bacillariophyta	Scaffold	51,094,165	97,960	92,667	45.13	GCA_018806925.1	2021/6/9
<i>Skeletonema marinoi</i>	Bacillariophyta	Chromosome	64,991,942	3,004,636	1,754,305	45.26	GCA_030544225.1	2023/8/8
<i>Skeletonema marinoi</i>	Bacillariophyta	Contig	54,623,658	1,237,151	1,237,151	45.26	GCA_030871285.1	2023/8/25
<i>Thalassiosira oceanica</i>	Bacillariophyta	Scaffold	83,512,925	3,960	3,572	53.28	GCA_019693575.1	2021/8/19
<i>Thalassiosira oceanica</i>	Bacillariophyta	Contig	92,043,847	3,628	3,628	53.28	GCA_000296195.2	2012/7/25
<i>Thalassiosira pseudonana</i> CCMP1335	Bacillariophyta	Chromosome	32,437,365	1,992,434	1,267,198	46.9	GCF_000149405.2	2009/1/16
<i>Thalassiosira</i> sp.	Bacillariophyta	Contig	134,500	15,270	15,270	29.44	GCA_023345645.1	2022/5/12
<i>Thalassiosira sundarbana</i>	Bacillariophyta	Contig	52,892,417	10,968	10,968	51.45	GCA_020086505.1	2021/9/21
uncultured Bacillariales	Bacillariophyta	Contig	19,069,982	14,509	14,509	45.54	GCA_937871795.1	2023/1/12
uncultured Thalassiosira	Bacillariophyta	Contig	30,739,597	5,996	5,996	51.59	GCA_947538825.1	2023/4/19
<i>Chroomonas mesostigmatica</i> CCMP1168	Cryptophyceae	Chromosome	702,852	232,699	232,699	25.94	GCA_000286095.1	2012/8/9
<i>Cryptomonas paramecium</i>	Cryptophyceae	Chromosome	487,066	160,189	160,189	26.19	GCF_000194455.1	2011/3/30
<i>Cryptophyta</i> sp. CCMP2293	Cryptophyceae	Scaffold	534,407,040	439,318	5,115	53.92	GCA_026770585.1	2022/12/9
<i>Goniomonas avonlea</i>	Cryptophyceae	Contig	91,516,952	3,834	3,834	55.21	GCA_003573635.1	2018/9/18
<i>Guillardia theta</i>	Cryptophyceae	Chromosome	551,264	180,915	180,915	53.1	GCF_000002975.1	2001/4/26
<i>Guillardia theta</i> CCMP2712	Cryptophyceae	Scaffold	87,145,349	545,808	40,445	53.1	GCF_000315625.1	2012/12/5
<i>Hemiselmis andersenii</i>	Cryptophyceae	Chromosome	571,872	184,755	184,755	25.22	GCF_000018645.1	2008/4/24
<i>Hemiselmis andersenii</i>	Cryptophyceae	Chromosome	571,872	184,755	184,755	25.22	GCA_000018645.1	2007/11/9
uncultured Cryptomonadales	Cryptophyceae	Contig	6,116,236	3,934	3,934	50.5	GCA_947538865.1	2023/4/19
<i>Euglena gracilis</i>	Euglenozoa	Contig	1,435,499,417	955	955	50.32	GCA_900893395.1	2019/4/12
<i>Euglena gracilis</i> var <i>bacillaris</i>	Euglenozoa	Contig	41,178,302	409	409	50.1	GCA_001638955.2	2016/5/6
<i>Rhabdomonas costata</i>	Euglenozoa	Scaffold	97,249,925	1,112	1,111	51.52	GCA_019059605.1	2021/6/29
<i>Cyanophora paradoxa</i>	Glaucomastophyceae	Contig	98,351,904	207,789	207,789	63.43	GCA_004431415.1	2019/2/5

<i>Chrysochromulina parva</i>	Haptophyta	Contig	65,752,923	16,059	16,059	63.58	GCA_002887195.1	2018/1/17
<i>Chrysochromulina tobinii</i>	Haptophyta	Contig	59,073,094	24,114	24,114	63.37	GCA_001275005.1	2015/8/26
<i>Diacronema lutheri</i>	Haptophyta	Contig	43,502,671	852,258	852,258	73.25	GCA_019448385.1	2021/8/4
<i>Diacronema lutheri</i>	Haptophyta	Scaffold	18,295,499	63,420	43,749	73.25	GCA_019693415.1	2021/8/19
<i>Emiliana huxleyi</i> CCMP1516	Haptophyta	Scaffold	167,675,649	404,808	29,715	65.67	GCA_000372725.1	2013/5/2
<i>Isochrysis galbana</i>	Haptophyta	Chromosome	92,733,715	6,994,209	117,000	58.44	GCA_018136815.1	2021/4/26
<i>Pavlova</i> sp. CCMP2436	Haptophyta	Scaffold	165,404,270	252,369	5,960	55.21	GCA_026770615.1	2022/12/9
uncultured <i>Chrysochromulina</i>	Haptophyta	Contig	59,972,424	5,785	5,785	63.45	GCA_947538845.1	2023/4/19
uncultured <i>Chrysochromulina</i>	Haptophyta	Contig	59,711,118	3,627	3,627	63.61	GCA_947538835.1	2023/4/19
<i>Aureococcus anophagefferens</i>	Ochrophyta	Scaffold	56,660,572	1,405,779	34,096	69.5	GCA_000186865.1	2011/2/15
<i>Aureococcus anophagefferens</i>	Ochrophyta	Contig	64,432,857	844,787	844,787	69.5	GCA_021029045.1	2022/1/11
<i>Aureococcus anophagefferens</i>	Ochrophyta	Contig	73,614,722	552,779	552,779	69.5	GCA_021029115.1	2022/1/11
<i>Aureococcus anophagefferens</i>	Ochrophyta	Contig	57,108,930	483,163	483,163	69.5	GCA_021029035.1	2022/1/11
<i>Aureococcus anophagefferens</i>	Ochrophyta	Scaffold	56,660,572	1,405,779	34,096	69.5	GCA_000186865.1	2011/2/15
<i>Aureococcus anophagefferens</i>	Ochrophyta	Scaffold	56,660,572	1,405,779	34,096	69.5	GCF_000186865.1	2011/2/15
<i>Aureococcus anophagefferens</i>	Ochrophyta	Contig	64,432,857	844,787	844,787	69.5	GCA_021029045.1	2022/1/11
<i>Aureococcus anophagefferens</i>	Ochrophyta	Contig	73,614,722	552,779	552,779	69.5	GCA_021029115.1	2022/1/11
<i>Aureococcus anophagefferens</i>	Ochrophyta	Contig	57,108,930	483,163	483,163	69.5	GCA_021029035.1	2022/1/11
<i>Chlorellidium tetrabotrys</i>	Ochrophyta	Scaffold	204,248,023	1,188	1,150	42.32	GCA_032273165.1	2023/10/2
<i>Chloridella minuta</i>	Ochrophyta	Scaffold	182,064,201	2,423	2,287	43.8	GCA_032273045.1	2023/10/2
<i>Chromulina nebulosa</i>	Ochrophyta	Scaffold	110,814,819	1,695	1,670	43.87	GCA_015148605.1	2020/10/30
<i>Chromulinospumella</i> sp. haerica	Ochrophyta	Scaffold	82,044,956	4,036	4,007	51.26	GCA_015144965.1	2020/10/30
<i>Chrysophaeum taylorii</i>	Ochrophyta	Contig	70,683,391	157,935	157,935	56.9	GCA_029834535.1	2023/4/19
<i>Cladosiphon okamuranus</i>	Ochrophyta	Scaffold	169,730,751	505,945	99,235	54.5	GCA_001742925.1	2016/8/18
<i>Dinobryon divergens</i>	Ochrophyta	Scaffold	113,367,926	1,836	1,794	45.06	GCA_015146095.1	2020/10/30
<i>Dinobryon</i> sp. LO226KS	Ochrophyta	Scaffold	87,456,568	2,529	2,359	51.61	GCA_015143705.1	2020/10/30
<i>Ectocarpus siliculosus</i>	Ochrophyta	Chromosome	195,810,619	497,380	32,613	53.59	GCA_000310025.1	2010/6/24
<i>Ectocarpus</i> sp. CCAP 1310/34	Ochrophyta	Scaffold	242,356,113	115,928	11,922	53.64	GCA_902602495.3	2020/3/23
<i>Ectocarpus</i> sp. CCAP 1310/34	Ochrophyta	Scaffold	242,239,399	510,467	8,040	53.64	GCA_902705575.1	2019/12/17
<i>Ectocarpus</i> sp. Ec32	Ochrophyta	Chromosome	195,781,519	6,813,345	34,774	53.59	GCA_004764655.1	2019/4/12
<i>Epipyxis</i> sp. PR26KG	Ochrophyta	Scaffold	57,533,909	11,846	11,258	33.39	GCA_015144975.1	2020/10/30
<i>Fucus vesiculosus</i>	Ochrophyta	Scaffold	1,508,024,975	38,557	522	47.89	GCA_014849475.1	2020/10/5
<i>Heterococcus</i> sp. DN1	Ochrophyta	Contig	60,742,704	4,231	4,231	48.19	GCA_000498555.1	2013/11/22
<i>Heterotrichella gracilis</i>	Ochrophyta	Scaffold	98,294,111	70,353	58,996	49.67	GCA_032358255.1	2023/10/4
<i>Hydrurus foetidus</i>	Ochrophyta	Contig	171,182,409	43,856	43,856	45.41	GCA_900617105.1	2019/4/1
<i>Macrocystis pyrifera</i>	Ochrophyta	Contig	409,091,383	2,573	2,573	49.98	GCA_012273945.1	2020/4/13
<i>Macrocystis pyrifera</i>	Ochrophyta	Scaffold	537,452,659	13,669,674	1,635,752	49.98	GCA_031763025.1	2023/9/20
<i>Mallomonas annulata</i>	Ochrophyta	Scaffold	108,460,286	7,396	7,062	40.07	GCA_015143345.1	2020/10/30
<i>Nannochloropsis gaditana</i>	Ochrophyta	Chromosome	30,708,667	1,141,545	1,081,029	55.32	GCA_002838785.1	2017/12/14
<i>Nannochloropsis gaditana</i>	Ochrophyta	Chromosome	27,432,140	1,065,987	39,853	55.32	GCA_000569095.1	2014/2/18
<i>Nannochloropsis gaditana</i>	Ochrophyta	Contig	25,618,953	12,737	12,737	55.32	GCA_001614215.1	2016/4/8
<i>Nannochloropsis gaditana</i>	Ochrophyta	Chromosome	30,708,667	1,141,545	1,081,029	55.32	GCA_002838785.1	2017/12/14
<i>Nannochloropsis gaditana</i>	Ochrophyta	Chromosome	27,432,140	1,065,987	39,853	55.32	GCA_000569095.1	2014/2/18
<i>Nannochloropsis gaditana</i>	Ochrophyta	Contig	25,618,953	12,737	12,737	55.32	GCA_001614215.1	2016/4/8

<i>Nannochloropsis gaditana</i> CCMP526	Ochrophyta	Scaffold	33,986,952	37,643	21,340	54.26	GCA_000240725.1	2012/1/4
<i>Nannochloropsis granulata</i> CCMP529	Ochrophyta	Scaffold	27,651,077	14,688	4,159	54.58	GCA_004335405.1	2019/3/8
<i>Nannochloropsis limnetica</i>	Ochrophyta	Contig	33,508,916	2,703	2,703	55.39	GCA_001614225.1	2016/4/8
<i>Nannochloropsis oceanica</i>	Ochrophyta	Chromosome	29,109,164	1,148,430	701,908	54.05	GCA_004519485.1	2019/3/28
<i>Nannochloropsis oceanica</i>	Ochrophyta	Chromosome	27,546,055	1,078,593	59,448	54.05	GCA_009014725.1	2019/10/12
<i>Nannochloropsis oceanica</i>	Ochrophyta	Chromosome	26,878,636	1,038,807	59,284	54.05	GCA_009014695.1	2019/10/12
<i>Nannochloropsis oceanica</i>	Ochrophyta	Contig	27,639,860	12,329	12,329	54.05	GCA_000226695.1	2011/9/29
<i>Nannochloropsis oceanica</i>	Ochrophyta	Chromosome	29,109,164	1,148,430	701,908	54.05	GCA_004519485.1	2019/3/28
<i>Nannochloropsis oceanica</i>	Ochrophyta	Chromosome	27,546,055	1,078,593	59,448	54.05	GCA_009014725.1	2019/10/12
<i>Nannochloropsis oceanica</i>	Ochrophyta	Chromosome	26,878,636	1,038,807	59,284	54.05	GCA_009014695.1	2019/10/12
<i>Nannochloropsis oceanica</i>	Ochrophyta	Contig	27,639,860	12,329	12,329	54.05	GCA_000226695.1	2011/9/29
<i>Nannochloropsis oceanica</i> CCMP531	Ochrophyta	Scaffold	29,568,935	26,762	2,765	53.84	GCA_004335385.1	2019/3/8
<i>Nannochloropsis oceanica</i> OZ-1	Ochrophyta	Contig	28,021,914	39,278	39,278	53.72	GCA_001614235.1	2016/4/8
<i>Nannochloropsis oceanica</i> strain IMET1	Ochrophyta	Scaffold	31,476,008	935,195	39,360	53.72	GCA_001870945.1	2016/11/9
<i>Nannochloropsis oculata</i> CCMP525	Ochrophyta	Scaffold	26,274,801	27,236	2,479	54.13	GCA_004335455.1	2019/3/8
<i>Nannochloropsis salina</i> CCMP1776	Ochrophyta	Scaffold	27,600,439	828,788	181,586	54.76	GCA_004565275.1	2019/5/15
<i>Nannochloropsis salina</i> CCMP1776	Ochrophyta	Contig	24,357,051	12,186	12,186	54.76	GCA_001614245.1	2016/4/8
<i>Nannochloropsis salina</i> CCMP537	Ochrophyta	Scaffold	23,055,047	24,449	15,756	54.25	GCA_004335465.1	2019/3/8
<i>Nannochloropsis</i> sp. QH25	Ochrophyta	Scaffold	29,301,793	1,040,719	939,345	54.14	GCA_024741955.1	2022/8/25
<i>Nereocystis luetkeana</i>	Ochrophyta	Scaffold	467,623,804	10,969,560	1,040,861	49.94	GCA_031213475.1	2023/9/7
<i>Ochromonadaceae</i> sp. CCMP2298	Ochrophyta	Scaffold	61,116,754	7,503	4,898	53.42	GCA_026770595.1	2022/12/9
<i>Ochromonas danica</i>	Ochrophyta	Scaffold	44,155,045	97,490	69,358	45.45	GCA_015146655.1	2020/10/30
<i>Ochromonas</i> sp.	Ochrophyta	Contig	29,237,651	11,486	11,486	42.55	GCA_023346525.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	15,775,809	7,760	7,760	42.55	GCA_023346265.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	22,979,148	5,966	5,966	42.55	GCA_023345975.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	8,269,859	5,761	5,761	42.55	GCA_023346365.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	5,840,804	5,296	5,296	42.55	GCA_023346705.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	25,801,272	2,737	2,737	42.55	GCA_023346245.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	22,432,152	1,964	1,964	42.55	GCA_023346895.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	25,824,802	1,947	1,947	42.55	GCA_023346965.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	19,080,090	1,791	1,791	42.55	GCA_023346535.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	11,803,606	1,610	1,610	42.55	GCA_023346425.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	29,237,651	11,486	11,486	42.55	GCA_023346525.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	15,775,809	7,760	7,760	42.55	GCA_023346265.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	22,979,148	5,966	5,966	42.55	GCA_023345975.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	8,269,859	5,761	5,761	42.55	GCA_023346365.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	5,840,804	5,296	5,296	42.55	GCA_023346705.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	25,801,272	2,737	2,737	42.55	GCA_023346245.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	22,432,152	1,964	1,964	42.55	GCA_023346895.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	25,824,802	1,947	1,947	42.55	GCA_023346965.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	19,080,090	1,791	1,791	42.55	GCA_023346535.1	2022/5/12
<i>Ochromonas</i> sp.	Ochrophyta	Contig	11,803,606	1,610	1,610	42.55	GCA_023346425.1	2022/5/12

<i>Ophiocytium majus</i>	Ochrophyta	Scaffold	293,300,872	7,483	7,214	52.52	GCA_032358315.1	2023/10/4
<i>Parmales sp. scaly parma</i>	Ochrophyta	Scaffold	30,979,897	51,725	18,742	51.26	GCA_030267985.1	2023/5/25
<i>Pedospumella encystans</i>	Ochrophyta	Scaffold	61,203,055	3,884	3,618	51.43	GCA_015143685.1	2020/10/30
<i>Pedospumella encystans</i>	Ochrophyta	Scaffold	55,330,094	1,033	1,026	51.43	GCA_015148585.1	2020/10/30
<i>Pelagococcus subviridis</i>	Ochrophyta	Scaffold	21,596,650	15,065	9,415	65.81	GCA_019702425.1	2021/8/19
<i>Pelagomonas calceolata</i>	Ochrophyta	Contig	32,445,274	5,385,517	5,385,517	63.56	GCA_918797485.1	2021/11/19
<i>Pelagophyceae sp. CCMP2097</i>	Ochrophyta	Scaffold	85,824,453	186,141	14,243	62	GCA_026770575.1	2022/12/9
<i>Pinguicoccus pyrenoidosus</i>	Ochrophyta	Scaffold	32,008,640	27,803	19,902	59.67	GCA_019677205.1	2021/8/18
<i>Poterioochromonas sp. DS</i>	Ochrophyta	Scaffold	57,519,221	22,724	12,958	40.36	GCA_015145745.1	2020/10/30
<i>Poteriospumella lacustris</i>	Ochrophyta	Scaffold	54,674,105	52,526	9,979	53.13	GCA_008000375.1	2019/8/15
<i>Poteriospumella lacustris</i>	Ochrophyta	Scaffold	49,390,929	29,656	8,061	53.13	GCA_008000325.1	2019/8/15
<i>Poteriospumella lacustris</i>	Ochrophyta	Scaffold	52,749,073	24,654	6,673	53.13	GCA_008000405.1	2019/8/15
<i>Pterygophora californica</i>	Ochrophyta	Scaffold	93,119,657	692	681	46.64	GCA_032273145.1	2023/10/2
<i>Saccharina japonica</i>	Ochrophyta	Scaffold	535,382,851	342,726	90,005	49.35	GCA_008828725.1	2019/10/4
<i>Saccharina japonica</i>	Ochrophyta	Scaffold	543,425,876	252,007	44,572	49.35	GCA_000978595.1	2015/4/22
<i>Spumella sp. AR4D6</i>	Ochrophyta	Scaffold	69,798,425	2,507	2,446	51.72	GCA_015146115.1	2020/10/30
<i>Spumella vulgaris</i>	Ochrophyta	Scaffold	89,618,401	2,679	2,657	47.94	GCA_015148565.1	2020/10/30
<i>Synura sp. LO234KE</i>	Ochrophyta	Scaffold	116,222,657	2,025	1,923	46.86	GCA_015143735.1	2020/10/30
<i>Tetraparma gracilis</i>	Ochrophyta	Scaffold	43,592,913	10,699	3,715	65.85	GCA_030267505.1	2023/5/25
<i>Tribonema minus</i>	Ochrophyta	Contig	158,345,074	768,631	768,631	56.96	GCA_017506865.1	2021/3/22
<i>Triparma columacea</i>	Ochrophyta	Scaffold	42,982,915	63,509	27,173	51.12	GCA_030267965.1	2023/5/25
<i>Triparma laevis f inornata</i>	Ochrophyta	Scaffold	42,573,550	83,218	14,736	51.26	GCA_012489335.1	2020/4/11
<i>Triparma laevis f longispina</i>	Ochrophyta	Scaffold	41,411,691	77,879	26,940	51.29	GCA_030267805.1	2023/5/25
<i>Triparma retinervis</i>	Ochrophyta	Scaffold	36,536,487	8,214	3,044	53.31	GCA_030267625.1	2023/5/25
<i>Triparma strigata</i>	Ochrophyta	Scaffold	35,156,025	73,438	52,227	52.21	GCA_030267945.1	2023/5/25
<i>Triparma verrucosa</i>	Ochrophyta	Scaffold	35,535,122	73,983	53,064	52.13	GCA_030267905.1	2023/5/25
<i>Undaria pinnatifida</i>	Ochrophyta	Scaffold	511,280,173	16,510,065	1,670,562	50.14	GCA_012845835.1	2020/4/27
<i>Undaria pinnatifida</i>	Ochrophyta	Chromosome	634,119,850	13,610,167	406,296	50.14	GCA_020975765.1	2021/11/30
<i>Vaucheria bursata</i>	Ochrophyta	Scaffold	191,659,160	16,466	15,654	39.44	GCA_032358265.1	2023/10/4

附表 2. 已公布的无脊椎动物基因组基本信息

Species	Phylum	Assembly Level	Total Length(bp)	Scaffold N50(bp)	Contig N50(bp)	GC(%)	Accession	Release Date
<i>Pomphorhynchus laevis</i>	Acanthocephala	Contig	253,271,445	126,570	126,570	33.07	GCA_012934845.2	2022/3/18
<i>Dimorphilus gyrotilatus</i>	Annelida	Scaffold	77,897,245	2,241,898	979,161	31.36	GCA_904063045.1	2020/8/25
<i>Hirudinaria manillensis</i>	Annelida	Contig	151,673,145	-	2,277,518	35.50	GCA_015345955.1	2020/11/10
<i>Piscicola geometra</i>	Annelida	Chromosome	171,080,115	9,301,023	1,427,829	37.00	GCA_943735955.1	2022/6/24
<i>Branchellion lobata</i>	Annelida	Chromosome	174,127,743	10,374,467	502,393	36.50	GCA_947562095.1	2022/12/13
<i>Whitmania pigra</i>	Annelida	Contig	178,785,238	2,001,490	2,001,490	35.46	GCA_021613335.1	2022/1/26
<i>Hirudo medicinalis</i>	Annelida	Scaffold	187,546,627	97,775	24,951	30.00	GCA_011800805.1	2020/3/31
<i>Helobdella robusta</i>	Annelida	Scaffold	235,376,169	3,060,193	52,195	34.20	GCF_000326865.1	2012/12/19
<i>Capitella teleta</i>	Annelida	Scaffold	333,283,208	188,402	21,930	41.80	GCA_000328365.1	2013/1/25
<i>Owenia fusiformis</i>	Annelida	Scaffold	499,544,331	38,876,441	2,574,000	34.71	GCA_903813345.2	2022/3/14
<i>Paralvinella palmiformis</i>	Annelida	Contig	601,252,053	-	557,207	36.00	GCA_025642065.1	2022/10/11
<i>Ridgeia piscesae</i>	Annelida	Contig	658,248,171	-	365,412	40.50	GCA_025642075.1	2022/10/11
<i>Lamellibrachia satsuma</i>	Annelida	Contig	664,980,951	267,064	266,880	40.03	GCA_022478865.1	2022/3/4
<i>Alitta virens</i>	Annelida	Chromosome	671,169,925	42,251,612	7,783,370	37.62	GCA_932294295.1	2022/3/11
<i>Lamellibrachia luymesii</i>	Annelida	Scaffold	687,706,213	372,990	24,086	35.10	GCA_009193005.1	2019/10/21
<i>Amphiduros pacificus</i>	Annelida	Chromosome	694,322,451	54,531,576	7,010,533	37.50	GCA_949316495.1	2023/3/10
<i>Streblospio benedicti</i>	Annelida	Chromosome	701,453,095	56,106,436	47,837	37.93	GCA_019095985.1	2021/7/6
<i>Terebella lapidaria</i>	Annelida	Chromosome	765,232,019	43,950,887	6,684,000	41.00	GCA_949152475.1	2023/4/8
<i>Protula sp._h_YS-2021</i>	Annelida	Chromosome	863,481,455	102,984,657	1,014,507	39.00	GCA_949752745.1	2023/4/8
<i>Lepidonotus clava</i>	Annelida	Chromosome	1,010,237,070	55,150,419	7,105,433	40.50	GCA_936440205.1	2022/5/3
<i>Amyntas corticis</i>	Annelida	Scaffold	1,017,224,661	539	505	40.27	GCA_900184025.1	2017/7/18
<i>Hydroides elegans</i>	Annelida	Scaffold	1,026,046,400	17,329	5,980	34.60	GCA_001703475.1	2016/8/11
<i>Paraescarpia echinospica</i>	Annelida	Chromosome	1,101,720,295	69,127,315	237,029	40.88	GCA_020002185.1	2021/9/21
<i>Aporrectodea caliginosa</i>	Annelida	Scaffold	1,111,828,997	1,177,782	650,792	40.18	GCA_020284085.1	2021/10/6
<i>Sthenelais limicola</i>	Annelida	Chromosome	1,131,133,856	136,979,486	10,232,024	38.50	GCA_942159475.1	2022/6/10
<i>Alentia gelatinosa</i>	Annelida	Contig	1,187,348,741	-	833,961	42.00	GCA_950023025.1	2023/4/29
<i>Sipunculus nudus</i>	Annelida	Chromosome	1,426,776,566	80,869,746	28,461,674	36.50	GCA_026874595.1	2022/12/14
<i>Platynereis dumerilii</i>	Annelida	Scaffold	1,469,601,079	54,798,709	4,288,053	37.50	GCA_026936325.1	2022/12/16
<i>Eisenia fetida</i>	Annelida	Scaffold	1,471,976,452	9,314	897	40.28	GCA_003999395.1	2019/1/4
<i>Acholoe squamosa</i>	Annelida	Chromosome	1,481,470,071	84,155,211	5,523,906	41.50	GCA_949317995.1	2023/4/4
<i>Harmothoe impar</i>	Annelida	Chromosome	1,512,304,043	83,397,641	1,337,000	42.50	GCA_947462335.1	2022/12/6
<i>Trichoceridae sp._BV-2014</i>	Arthropoda	Scaffold	41,711,596	1,520	1,395	29.00	GCA_001014425.1	2015/5/28
<i>Diptera sp._S2_005_002R2</i>	Arthropoda	Scaffold	43,768,535	6,837	6,817	42.00	GCA_004026295.1	2019/1/16
<i>Chironomidae sp.</i>	Arthropoda	Contig	50,753,847	-	14,176	35.50	GCA_023345965.1	2022/5/12
<i>Lepidurus arcticus</i>	Arthropoda	Scaffold	73,105,941	118,958	82,939	41.60	GCA_003724045.1	2018/11/13
<i>Lepidurus couesii</i>	Arthropoda	Scaffold	76,388,658	40,400	38,661	42.30	GCA_022832235.1	2022/4/6

<i>Lepidurus_apus_apus</i>	Arthropoda	Scaffold	78,857,107	136,761	124,430	41.50	GCA_022832285.1	2022/4/6
<i>Triops_longicaudatus</i>	Arthropoda	Scaffold	83,658,226	380,856	267,456	40.00	GCA_022885665.1	2022/4/11
<i>Oithona_nana</i>	Arthropoda	Scaffold	85,010,107	400,614	38,620	38.90	GCA_900157175.1	2017/2/21
<i>Clunio_marinus</i>	Arthropoda	Scaffold	85,491,412	1,871,155	154,800	32.50	GCA_900005825.1	2016/11/28
<i>Lepidurus_apus_lubbocki</i>	Arthropoda	Scaffold	87,970,075	42,769	15,890	42.00	GCA_003723985.1	2018/11/13
<i>Paracyclopina_nana</i>	Arthropoda	Contig	91,992,532	11,574,028	11,574,028	42.20	GCA_019096065.1	2021/7/2
<i>Hexapoda_sp.</i>	Arthropoda	Scaffold	99,041,005	17,347	16,929	36.00	GCA_002916965.1	2018/2/5
<i>Themira_minor</i>	Arthropoda	Scaffold	99,885,510	3,571	2,825	41.50	GCA_001014575.1	2015/5/28
<i>Platynothrus_peltifer</i>	Arthropoda	Scaffold	100,532,934	1,570	1,236	28.90	GCA_000988905.1	2015/5/6
<i>Leptestheria_dahalacensis</i>	Arthropoda	Scaffold	104,016,550	38,854	37,279	40.90	GCA_022114935.1	2022/2/9
<i>Daphnia_dubia</i>	Arthropoda	Scaffold	107,502,057	18,465	440	33.10	GCA_013387435.1	2020/7/2
<i>Lepidurus_packardii</i>	Arthropoda	Scaffold	108,645,424	12,481,803	1,298,445	40.50	GCA_023053545.1	2022/4/19
<i>Daphnia_carinata</i>	Arthropoda	Chromosome	109,458,891	3,869,624	3,869,624	40.80	GCA_022539665.1	2022/3/10
<i>Triops_cancriformis</i>	Arthropoda	Scaffold	114,976,066	56,108	48,879	36.20	GCA_020615345.1	2021/11/4
<i>Myllaena_sp._GC000078</i>	Arthropoda	Scaffold	119,221,412	11,375	9,424	43.50	GCA_030264135.1	2023/6/13
<i>Eulimnadia_texana</i>	Arthropoda	Scaffold	120,535,642	18,070,303	10,428,323	39.30	GCA_002872375.1	2018/1/10
<i>Daphnia_sinensis</i>	Arthropoda	Chromosome	131,590,582	8,418,570	2,802,846	40.89	GCA_013167095.1	2020/5/26
<i>Ceriodaphnia_dubia</i>	Arthropoda	Contig	132,649,347	2,288,570	2,288,570	42.40	GCA_022539655.1	2022/3/10
<i>Daphnia_pulex</i>	Arthropoda	Chromosome	133,196,385	12,288,052	8,301,990	40.14	GCF_021134715.1	2021/12/9
<i>Daphnia_galeata</i>	Arthropoda	Scaffold	133,304,630	756,671	483,996	38.70	GCA_918697745.1	2021/12/1
<i>Pandalus_platyceros</i>	Arthropoda	Scaffold	146,260,629	1,512	1,138	36.90	GCA_005815305.1	2019/5/29
<i>Anopheles_farauti_No._4</i>	Arthropoda	Contig	146,379,160	-	16,229	44.50	GCA_000956215.1	2015/3/19
<i>Aleochara_sp._GC000083</i>	Arthropoda	Scaffold	152,715,001	29,507	27,552	37.50	GCA_030264555.1	2023/6/13
<i>Chydorus_sphaericus</i>	Arthropoda	Chromosome	154,710,702	13,753,613	1,098,146	37.00	GCA_030141595.1	2023/6/5
<i>Amitus_sp._ZL-2020</i>	Arthropoda	Scaffold	160,412,742	1,590	1,585	35.50	GCA_022817785.1	2022/4/5
<i>Daphnia_magna</i>	Arthropoda	Chromosome	161,467,056	12,546,303	1,507,563	40.20	GCF_020631705.1	2021/10/27
<i>Strigamia_maritima</i>	Arthropoda	Scaffold	176,210,797	139,451	24,745	35.80	GCA_000239455.1	2011/12/22
<i>Anopheles_farauti</i>	Arthropoda	Scaffold	183,103,254	12,895,223	161,110	44.85	GCA_000473445.2	2014/1/17
<i>Daphnia_pulex</i>	Arthropoda	Chromosome	185,057,732	18,885,472	3,523,768	41.21	GCF_021234035.1	2022/1/11
<i>Tigriopus_californicus</i>	Arthropoda	Chromosome	191,142,546	15,806,032	44,438	42.32	GCA_007210705.1	2019/7/29
<i>Tigriopus_japonicus</i>	Arthropoda	Chromosome	196,599,007	10,654,335	70,641	39.93	GCA_010645155.1	2020/2/21
<i>Adinopsis_sp._GC000173</i>	Arthropoda	Scaffold	205,153,063	16,059	16,056	31.00	GCA_030264715.1	2023/6/13
<i>Cletocampus_dominicanus</i>	Arthropoda	Scaffold	209,620,252	1,121	1,108	43.00	GCA_028565815.1	2023/2/9
<i>Anopheles_melas</i>	Arthropoda	Scaffold	224,162,116	18,103	11,312	44.50	GCA_000473525.2	2014/1/17
<i>Anoncus_sp._PSUC_FEM_10030000</i>	Arthropoda	Scaffold	226,362,449	808	808	38.50	GCA_018902145.1	2021/6/16
<i>Coelopa_frigida</i>	Arthropoda	Chromosome	230,965,604	37,402,801	313,772	40.72	GCA_017309665.1	2021/3/9
<i>Dusona_sp._PSUC_FEM_10030013</i>	Arthropoda	Scaffold	241,082,464	3,759	3,758	36.50	GCA_018902295.1	2021/6/16
<i>Apocyclops_royi</i>	Arthropoda	Scaffold	262,264,339	3,257	2,108	34.10	GCA_900607525.1	2019/2/27
<i>Coelopa_pilipes</i>	Arthropoda	Chromosome	262,952,336	47,252,759	1,280,259	42.50	GCA_947389925.1	2022/12/4
<i>Sacculina_carcini</i>	Arthropoda	Chromosome	264,490,643	9,565,286	3,521,841	41.95	GCA_916048095.2	2022/2/9
<i>Anopheles_gambiae_str._P EST</i>	Arthropoda	Chromosome	265,027,044	12,309,988	85,548	42.40	GCF_000005575.2	2006/10/16
<i>Tisbe_holothuriae</i>	Arthropoda	Scaffold	276,259,850	5,742	5,083	40.20	GCA_900659605.1	2019/7/29
<i>Myscelus_nobilis</i>	Arthropoda	Contig	291,092,801	-	1,789	36.00	GCA_018245955.1	2021/5/3

<i>Anopheles_merus</i>	Arthropoda	Chromosome	294,379,759	50,316,797	2,729,089	45.01	GCF_017562075.2	2021/3/24
<i>Notodromas_monacha</i>	Arthropoda	Scaffold	376,665,953	42,325	32,600	38.20	GCA_905338405.1	2021/5/18
<i>Darwinula_stevensoni</i>	Arthropoda	Scaffold	382,098,207	56,422	38,545	42.50	GCA_905338385.1	2021/5/19
<i>Branchinecta_lindahli</i>	Arthropoda	Scaffold	384,837,704	29,112,420	3,495,440	41.50	GCA_023053555.1	2022/4/19
<i>Eurytemora_affinis</i>	Arthropoda	Scaffold	389,032,277	252,275	67,724	32.60	GCF_000591075.1	2017/12/13
<i>Metis_holothuriae</i>	Arthropoda	Scaffold	422,854,744	2,372	2,109	44.00	GCA_028566655.1	2023/2/9
<i>Capitulum_mitella</i>	Arthropoda	Chromosome	477,053,020	31,794,967	1,800,705	50.00	GCA_030062745.1	2023/5/22
<i>Caligus_rogercresseyi</i>	Arthropoda	Chromosome	478,201,101	27,802,916	38,017	31.50	GCA_013387185.1	2019/11/13
<i>Semibalanus_balanoides</i>	Arthropoda	Scaffold	486,009,102	56,726	45,551	49.05	GCA_014673585.1	2020/9/17
<i>Notopais_sp._c_TS-2021</i>	Arthropoda	Contig	535,477,874	-	1,015	29.00	GCA_028455975.1	2023/2/3
<i>Hyalella_azteca</i>	Arthropoda	Scaffold	550,885,727	215,427	114,415	38.50	GCA_000764305.3	2019/9/9
<i>Branchinecta_lynchi</i>	Arthropoda	Scaffold	575,750,106	23,437,814	902,028	41.00	GCA_023053575.1	2022/4/19
<i>Balanus_glandula</i>	Arthropoda	Scaffold	582,475,369	3,341	3,257	47.00	GCA_030265075.1	2023/6/13
<i>Pagurus_longicarpus</i>	Arthropoda	Scaffold	609,410,934	1,721	1,297	38.50	GCA_028571265.1	2023/2/9
<i>Notopais_sp._p_TS-2021</i>	Arthropoda	Contig	609,453,061	-	1,695	29.00	GCA_028455955.1	2023/2/3
<i>Anisolabis_maritima</i>	Arthropoda	Scaffold	641,675,028	1,383,245	48,444	31.10	GCA_010014785.1	2020/1/30
<i>Lepeophtheirus_salmonis</i>	Arthropoda	Chromosome	647,176,227	48,457,437	4,499,712	30.85	GCF_016086655.3	2020/12/16
<i>Longipedia_americana</i>	Arthropoda	Scaffold	657,676,259	2,481	2,476	32.50	GCA_028571025.1	2023/2/9
<i>Pantala_flavescens</i>	Arthropoda	Chromosome	663,273,809	56,028,107	16,157,025	34.46	GCA_020796165.1	2021/11/8
<i>Metacarcinus_magister</i>	Arthropoda	Scaffold	730,208,932	16,129,408	330,138	40.00	GCA_029783475.1	2023/4/17
<i>Pollicipes_pollicipes</i>	Arthropoda	Scaffold	770,089,732	47,009,503	109,725	52.30	GCF_011947565.2	2020/10/27
<i>Amphibalanus_amphitrite</i>	Arthropoda	Contig	807,619,925	536,785	536,785	49.70	GCF_019059575.1	2021/6/29
<i>Orchestia_grillus</i>	Arthropoda	Scaffold	812,068,892	16,847	16,630	39.90	GCA_014899125.1	2020/10/20
<i>Pagurus_granosimanus</i>	Arthropoda	Scaffold	841,335,030	1,401	896	38.00	GCA_030265335.1	2023/6/13
<i>Artemia_franciscana</i>	Arthropoda	Scaffold	850,560,342	112,377	61,131	34.70	GCA_019857095.1	2021/9/2
<i>Trinorchestia_longiramus</i>	Arthropoda	Scaffold	886,359,443	120,570	29,650	33.50	GCA_006783055.1	2019/7/9
<i>Acartia_tonsa</i>	Arthropoda	Scaffold	989,163,677	3,610	3,244	31.60	GCA_900241095.1	2018/1/29
<i>Callinectes_sapidus</i>	Arthropoda	Chromosome	997,995,872	18,846,596	9,312	39.47	GCA_020233015.1	2021/10/4
<i>Portunus_trituberculatus</i>	Arthropoda	Chromosome	1,005,062,047	21,793,880	4,121,416	39.80	GCF_017591435.1	2021/3/30
<i>Platorchestia_sp._MABIK</i>	Arthropoda	Scaffold	1,177,921,742	86,542	21,305	36.70	GCA_014220935.1	2020/8/17
<i>Artemia_sp._Kazakhstan_ARC1039</i>	Arthropoda	Scaffold	1,246,959,422	27,462	11,347	35.50	GCA_029168905.1	2023/3/13
<i>Eriocheir_sinensis</i>	Arthropoda	Chromosome	1,272,135,116	17,608,299	3,161,423	41.96	GCA_013436485.1	2020/7/21
<i>Penaeus_chinensis</i>	Arthropoda	Chromosome	1,466,079,641	36,870,704	470,155	37.30	GCF_019202785.1	2021/7/13
<i>Hetaerina_americana</i>	Arthropoda	Scaffold	1,630,025,798	86,194,728	4,979,805	39.00	GCA_022747635.1	2022/3/28
<i>Penaeus_vannamei</i>	Arthropoda	Scaffold	1,663,581,301	605,555	86,864	37.10	GCF_003789085.1	2018/11/16
<i>Carcinoscorpius_rotundicauda</i>	Arthropoda	Chromosome	1,668,682,064	102,344,000	7,608,042	33.60	GCA_011833715.1	2020/4/1
<i>Artemia_sinica</i>	Arthropoda	Chromosome	1,701,023,953	67,194,031	1,689,053	35.00	GCA_027921565.1	2023/1/20
<i>Penaeus_japonicus</i>	Arthropoda	Scaffold	1,705,010,925	234,949	132,830	36.60	GCF_017312705.1	2021/2/12
<i>Platycnemis_pennipes</i>	Arthropoda	Chromosome	1,793,356,121	144,837,943	18,171,807	39.27	GCA_933228895.1	2022/3/23
<i>Limulus_polyphemus</i>	Arthropoda	Scaffold	1,828,271,751	254,089	11,441	34.50	GCF_000517525.1	2014/1/3

<i>Panulirus_ornatus</i>	Arthropoda	Scaffold	1,925,793,637	8,139	5,433	39.10	GCA_018397875.1	2021/5/18
<i>Penaeus_indicus</i>	Arthropoda	Scaffold	1,935,640,391	34,405,730	463,417	32.00	GCA_018983055.1	2021/6/21
<i>Caridina_multidentata</i>	Arthropoda	Scaffold	1,948,953,281	819	819	33.30	GCA_002091895.1	2017/4/13
<i>Oppeia_nova</i>	Arthropoda	Scaffold	1,967,245,510	6,753	6,155	30.40	GCA_905397405.1	2021/9/27
<i>Macrobrachium_nipponense</i>	Arthropoda	Chromosome	1,985,027,095	83,001,933	267,348	36.85	GCA_015104395.1	2020/10/29
<i>Chionoecetes_opilio</i>	Arthropoda	Scaffold	2,002,919,378	208,145	149,622	41.80	GCA_016584305.1	2021/1/8
<i>Tachypleus_tridentatus</i>	Arthropoda	Chromosome	2,167,470,406	169,002,194	1,644,156	33.10	GCA_004210375.1	2019/4/3
<i>Homarus_americanus</i>	Arthropoda	Scaffold	2,292,076,018	759,644	133,311	43.00	GCF_018991925.1	2021/6/23
<i>Penaeus_monodon</i>	Arthropoda	Chromosome	2,394,347,767	44,862,054	45,084	38.20	GCF_015228065.1	2020/11/5
<i>Procambarus_clarkii</i>	Arthropoda	Chromosome	2,735,361,232	17,011,522	217,730	44.06	GCF_020424385.1	2021/10/12
<i>Parhyale_hawaiiensis</i>	Arthropoda	Scaffold	2,752,560,740	20,228,728	10,438	32.60	GCA_001587735.2	2018/6/22
<i>Gammarus_rossi</i>	Arthropoda	Scaffold	3,239,433,295	4,812	4,736	42.80	GCA_016164225.1	2020/12/21
<i>Cherax_destructor</i>	Arthropoda	Scaffold	3,336,744,225	87,184	80,900	41.40	GCA_009830355.1	2020/1/6
<i>Procambarus_virginialis</i>	Arthropoda	Scaffold	3,700,845,369	144,428	12,232	28.90	GCA_020271785.1	2021/10/4
<i>Paralithodes_camtschaticus</i>	Arthropoda	Scaffold	3,809,538,835	7,036	5,760	39.50	GCA_018397895.1	2021/5/18
<i>Paralithodes_platypus</i>	Arthropoda	Chromosome	4,805,175,703	51,153,954	147,786	41.57	GCA_013283005.1	2020/6/11
<i>Cherax_quadricarinatus</i>	Arthropoda	Chromosome	5,247,605,071	45,061,517	146,379	42.00	GCF_026875155.1	2022/12/14
<i>Bathynomus_jamesi</i>	Arthropoda	Contig	5,894,292,451	-	586,500	37.00	GCA_023014485.1	2022/4/15
<i>Palaemon_carinicauda</i>	Arthropoda	Scaffold	6,699,723,695	962	696	36.80	GCA_004011675.1	2019/1/11
<i>Euphausia_superba</i>	Arthropoda	Chromosome	#####	#####	178,990	29.36	CNP0001930	2023/3/16
<i>Lingula_anatina</i>	Brachiopoda	Scaffold	406,308,128	460,090	56,058	36.84	GCF_001039355.2	2015/6/24
<i>Cellarinelloides_crassus</i>	Bryozoa	Scaffold	74,423,554	882	875	50.00	GCA_027564095.1	2023/1/3
<i>Bugula_neritina</i>	Bryozoa	Contig	214,708,255	94,086	94,086	34.72	GCA_010799875.2	2020/8/6
<i>Bugulina_stolonifera</i>	Bryozoa	Chromosome	234,968,257	20,271,823	12,561,000	36.50	GCA_935421135.1	2022/5/3
<i>Membranipora_membranacea</i>	Bryozoa	Chromosome	339,374,674	29,963,023	7,007,262	38.01	GCA_914767715.1	2021/9/16
<i>Cristatella_mucedo</i>	Bryozoa	Scaffold	573,890,644	4,115,582	3,552,225	46.92	GCA_009760855.2	2020/10/23
<i>Cryptosula_pallasiana</i>	Bryozoa	Chromosome	605,593,191	46,264,604	1,788,101	37.00	GCA_945261195.1	2022/7/13
<i>Himantozoum_antarcticum</i>	Bryozoa	Scaffold	683,059,872	2,980	2,956	38.00	GCA_027564295.1	2023/1/3
<i>Oikopleura dioica</i>	Chordata	Chromosome	64,272,340	16,158,756	4,700,139	41.00	GCA_907165135.1	2021/5/17
<i>Ciona_intestinalis</i>	Chordata	Chromosome	115,212,710	3,102,162	37,096	35.50	GCF_000224145.3	2013/9/17
<i>Halocynthia_rosea</i>	Chordata	Scaffold	119,682,510	200,671	71,606	36.00	GCA_013436055.1	2020/7/21
<i>Halocynthia_aurantium</i>	Chordata	Scaffold	128,100,739	30,875	19,964	36.50	GCA_013436065.1	2020/7/21
<i>Ascidia_mentula</i>	Chordata	Chromosome	196,955,708	21,171,019	11,890,000	39.00	GCA_947561715.1	2022/12/11
<i>Clavelina_lepadiiformis</i>	Chordata	Chromosome	210,074,726	25,070,715	8,263,119	35.50	GCA_947623445.1	2022/12/19
<i>Phallusia_fumigata</i>	Chordata	Scaffold	231,877,689	12,731	11,099	38.00	GCA_008931825.1	2019/10/9
<i>Phallusia_mammillata</i>	Chordata	Scaffold	233,920,252	95,113	46,189	37.50	GCA_003260075.1	2018/6/22
<i>Oikopleura_longicauda</i>	Chordata	Contig	308,740,099	-	2,440	37.00	GCA_004367895.1	2019/3/25

<i>Salpa_thompsoni</i>	Chordata	Scaffold	318,747,957	934	636	35.50	GCA_001749815.1	2016/9/29
<i>Styela_clava</i>	Chordata	Scaffold	340,475,108	20,774,986	735,968	35.00	GCF_013122585.1	2021/2/2
<i>Oikopleura_albicans</i>	Chordata	Scaffold	365,958,051	213,763	10,311	35.00	GCA_004367875.1	2019/3/25
<i>Branchiostoma_floridae_x_Branchiostoma_japonicum</i>	Chordata	Chromosome	381,985,860	19,529,791	14,214,614	41.50	GCA_013266295.2	2020/7/23
<i>Bathochordaeus_stygius</i>	Chordata	Scaffold	396,466,979	1,347	1,346	36.50	GCA_004367955.1	2019/3/25
<i>Branchiostoma_belcheri</i>	Chordata	Scaffold	426,108,443	2,325,619	47,280	41.00	GCF_001625305.1	2016/4/21
<i>Asymmetron_lucayanum</i>	Chordata	Scaffold	460,590,149	3,154	1,984	41.00	GCA_001663935.1	2016/6/16
<i>Branchiostoma_lanceolatum</i>	Chordata	Chromosome	474,791,770	23,752,511	1,191,949	41.50	GCA_927797965.1	2022/2/2
<i>Branchiostoma_floridae_x_Branchiostoma_belcheri</i>	Chordata	Chromosome	490,439,570	25,499,959	5,730,782	41.00	GCA_019207045.1	2021/7/13
<i>Branchiostoma_floridae</i>	Chordata	Chromosome	513,445,848	25,441,410	50,413	41.00	GCF_000003815.2	2020/4/29
<i>Botryllus_schlosseri</i>	Chordata	Scaffold	579,633,380	7,220	6,924	40.50	GCA_000444245.1	2013/8/7
<i>Ciona_savignyi</i>	Chordata	Scaffold	587,352,817	601,536	22,563	36.50	GCA_000149265.1	2004/2/25
<i>Aplidium_turbinatum</i>	Chordata	Chromosome	605,489,456	34,009,763	14,915,238	37.50	GCA_918807975.1	2021/11/19
<i>Oikopleura_vanhoeffeni</i>	Chordata	Scaffold	643,662,591	259,996	20,125	32.00	GCA_004367855.1	2019/3/25
<i>Mesochordaeus_erythrocephalus</i>	Chordata	Scaffold	874,018,965	1,905	1,905	36.00	GCA_004367975.1	2019/3/25
<i>Kudoa_iwatai</i>	Cnidaria	Scaffold	31,177,209	40,195	39,532	23.60	GCA_001407235.1	2015/10/22
<i>Myxobolus_squamalis</i>	Cnidaria	Contig	43,671,844	1,286	1,286	27.30	GCA_010108815.1	2020/2/3
<i>Henneguya_salminicola</i>	Cnidaria	Scaffold	61,443,780	7,570	7,570	28.96	GCA_009887335.1	2020/1/15
<i>Enteromyxum_leei</i>	Cnidaria	Scaffold	67,975,573	996	996	33.50	GCA_001455295.1	2015/12/2
<i>Ceratonova_shasta</i>	Cnidaria	Scaffold	69,794,587	36,279	9,293	23.50	GCA_918559645.1	2021/11/20
<i>Renilla_reniformis</i>	Cnidaria	Scaffold	131,554,742	1,843	1,741	36.60	GCA_900177555.1	2017/4/20
<i>Thelohanellus_kitaei</i>	Cnidaria	Scaffold	150,348,159	149,756	13,033	31.00	GCA_000827895.1	2015/1/23
<i>Sphaeromyxa_zaharoni</i>	Cnidaria	Scaffold	173,585,031	4,474	4,474	28.00	GCA_001455285.1	2015/12/2
<i>Tamoya_ohboya</i>	Cnidaria	Scaffold	183,949,938	686	686	38.00	GCA_028566775.1	2023/2/9
<i>Sanderia_malayensis</i>	Cnidaria	Scaffold	184,371,355	4,676,287	411,334	37.32	GCA_013076295.1	2020/5/13
<i>Calvadosia_cruxmelitensis</i>	Cnidaria	Scaffold	209,392,379	16,443	11,652	39.95	GCA_900245855.1	2018/1/27
<i>Haliclystus_inabai</i>	Cnidaria	Scaffold	212,279,008	2,579	2,304	35.00	GCA_025505365.1	2022/9/30
<i>Nemopilema_nomurai</i>	Cnidaria	Scaffold	213,621,014	2,711,397	849,297	38.22	GCA_003864495.1	2018/12/5
<i>Epizoanthus_planus</i>	Cnidaria	Scaffold	215,922,282	775	775	40.00	GCA_025388665.1	2022/9/23
<i>Turritopsis_rubra</i>	Cnidaria	Scaffold	216,274,671	71,856	40,797	34.00	GCA_025167575.1	2022/9/13
<i>Xenia_sp._Carnegie-2017</i>	Cnidaria	Scaffold	222,699,550	14,832,246	1,122,448	35.90	GCF_021976095.1	2022/2/4
<i>Pocillopora_damicornis</i>	Cnidaria	Scaffold	234,333,463	326,133	25,941	37.82	GCA_003704095.1	2018/10/31
<i>Actinia_tenebrosa</i>	Cnidaria	Scaffold	238,179,736	188,789	11,161	37.16	GCA_009602425.1	2019/11/6
<i>Virgularia_gustaviana</i>	Cnidaria	Scaffold	243,681,559	3,138	2,873	38.00	GCA_025770025.1	2022/10/20
<i>Myxobolus_honghuensis</i>	Cnidaria	Scaffold	254,310,768	3,351	444	27.60	GCA_022478855.1	2022/3/1
<i>Exaiptasia_diaphana</i>	Cnidaria	Scaffold	256,132,296	442,145	14,401	38.80	GCF_001417965.1	2015/10/28
<i>Rhopilema_esculentum</i>	Cnidaria	Scaffold	256,689,583	12,930,448	193,069	36.30	GCA_013076305.1	2020/5/13
<i>Haliclystus_oxoradiatus</i>	Cnidaria	Chromosome	261,880,027	28,991,210	1,466,457	33.80	GCA_916610825.1	2021/10/1
<i>Chrysaora_fuscescens</i>	Cnidaria	Scaffold	266,405,745	1,938	1,737	37.72	GCA_009936425.1	2020/1/27
<i>Nematostella_vectensis</i>	Cnidaria	Chromosome	269,418,438	17,865,913	2,927,189	40.50	GCF_932526225.1	2022/3/23
<i>Hydra_viridissima</i>	Cnidaria	Scaffold	284,246,930	1,079,704	19,520	28.95	GCA_014706445.1	2020/9/23
<i>Dendronephthya_gigantea</i>	Cnidaria	Contig	286,131,786	1,445,523	1,445,523	36.84	GCA_004324835.1	2019/3/4

<i>Anthopleura_sola</i>	Cnidaria	Scaffold	288,950,296	10,852,815	2,720,395	37.50	GCA_023349425.1	2022/5/12
<i>Stichodactyla_mertensii</i>	Cnidaria	Scaffold	292,949,330	443,501	21,519	37.78	GCA_011800005.1	2020/3/31
<i>Stichodactyla_helianthus</i>	Cnidaria	Scaffold	296,173,587	813,926	27,789	37.90	GCA_015163945.1	2020/11/1
<i>Churabana_kuroshioae</i>	Cnidaria	Scaffold	303,676,354	550	548	38.50	GCA_027563615.1	2023/1/3
<i>Acropora_palmata</i>	Cnidaria	Scaffold	304,252,072	6,820,020	230,835	38.50	GCA_025960835.1	2022/11/2
<i>Diadumene_lineata</i>	Cnidaria	Chromosome	312,988,696	17,670,263	2,745,340	35.20	GCA_918843875.1	2021/11/18
<i>Chrysaora_quinquecirrha</i>	Cnidaria	Scaffold	337,419,359	3,825,607	227,000	35.65	GCA_014526335.1	2020/9/11
<i>Pteroeides_caledonicum</i>	Cnidaria	Scaffold	346,410,713	3,178	2,946	36.50	GCA_026546875.1	2022/11/30
<i>Pocillopora_meandrina</i>	Cnidaria	Contig	347,233,126	-	4,753,879	37.50	GCA_942486045.1	2023/5/15
<i>Cassiopea_xamachana</i>	Cnidaria	Scaffold	366,839,413	17,884,241	17,635	36.50	GCA_900291935.2	2022/10/28
<i>Paleozoanthus_reticulatus</i>	Cnidaria	Scaffold	367,864,585	325	324	42.50	GCA_026546615.1	2022/11/30
<i>Blastomussa_wellsi</i>	Cnidaria	Chromosome	367,991,705	27,335,774	418,556	39.50	GCA_947652115.1	2022/12/21
<i>Hydrozoanthus_antumbrosus</i>	Cnidaria	Scaffold	372,657,702	12,926	11,018	37.50	GCA_026260235.1	2022/11/21
<i>Astreopora_myriophthalma</i>	Cnidaria	Scaffold	373,482,476	1,634,538	31,248	38.40	GCA_014634185.1	2020/9/4
<i>Hydrozoanthus_tunicans</i>	Cnidaria	Scaffold	373,613,569	12,644	10,921	37.50	GCA_026413305.1	2022/11/27
<i>Epizoanthus_rinbou</i>	Cnidaria	Scaffold	374,392,259	3,330	3,312	38.00	GCA_025505405.1	2022/9/30
<i>Aurelia_aurita</i>	Cnidaria	Scaffold	376,932,402	1,042,981	33,962	37.00	GCA_004194415.1	2019/2/14
<i>Heteractis_magnifica</i>	Cnidaria	Scaffold	376,947,355	1,098,995	14,063	38.17	GCA_011763375.1	2020/3/27
<i>Pocillopora_verrucosa</i>	Cnidaria	Scaffold	380,505,547	333,696	75,683	38.00	GCA_014529365.1	2020/9/11
<i>Acropora_microphthalma</i>	Cnidaria	Scaffold	383,861,656	1,061,150	36,638	35.30	GCA_014634165.1	2020/9/5
<i>Metridium_senile</i>	Cnidaria	Chromosome	390,884,731	20,827,098	1,819,000	35.00	GCA_949775045.1	2023/4/8
<i>Acropora_selago</i>	Cnidaria	Scaffold	392,938,768	657,172	34,309	36.50	GCA_014634525.1	2020/9/5
<i>Acropora_acuminata</i>	Cnidaria	Scaffold	394,735,330	1,005,452	51,873	36.90	GCA_014633975.1	2020/9/5
<i>Epizoanthus_scotinus</i>	Cnidaria	Scaffold	396,351,196	3,070	2,897	38.00	GCA_024757865.1	2022/8/28
<i>Stylophora_pistillata</i>	Cnidaria	Scaffold	397,587,675	460,160	20,684	38.50	GCF_002571385.2	2017/10/17
<i>Heteractis_crispa</i>	Cnidaria	Scaffold	399,044,133	1,013,117	11,629	39.10	GCA_015164035.1	2020/11/1
<i>Anemonia_viridis</i>	Cnidaria	Scaffold	400,604,393	2,087	1,315	38.30	GCA_900234385.1	2018/8/1
<i>Acropora_gemmifera</i>	Cnidaria	Scaffold	400,993,167	1,141,272	25,126	35.20	GCA_014634125.1	2020/9/5
<i>Acropora_echinata</i>	Cnidaria	Scaffold	401,478,809	1,917,675	24,392	33.20	GCA_014634105.1	2020/9/4
<i>Acropora_tenuis</i>	Cnidaria	Scaffold	403,138,309	1,165,953	27,215	36.10	GCA_014633955.1	2020/9/5
<i>Cassiopea_andromeda</i>	Cnidaria	Scaffold	406,637,004	307,826	16,678	33.70	GCA_018155075.1	2021/4/28
<i>Actinia_equina</i>	Cnidaria	Contig	409,058,306	492,607	492,607	37.51	GCA_011057435.1	2020/3/2
<i>Acropora_nasuta</i>	Cnidaria	Scaffold	416,422,065	1,051,962	40,130	36.20	GCA_014634205.1	2020/9/5
<i>Galaxea_fascicularis</i>	Cnidaria	Chromosome	416,848,215	33,853,504	1,749,000	39.50	GCA_948470475.1	2023/2/3
<i>Acropora_intermedia</i>	Cnidaria	Scaffold	416,875,905	577,312	40,253	36.90	GCA_014634585.1	2020/9/5
<i>Acropora_muricata</i>	Cnidaria	Scaffold	420,731,913	574,627	38,011	36.40	GCA_014634545.1	2020/9/5
<i>Clytia_hemisphaerica</i>	Cnidaria	Scaffold	420,978,673	1,182,947	300,839	35.45	GCA_902728285.1	2020/2/4
<i>Acropora_cytherea</i>	Cnidaria	Scaffold	426,326,548	1,091,490	50,557	36.60	GCA_014634045.1	2020/9/5
<i>Entacmaea_quadricolor</i>	Cnidaria	Scaffold	428,330,991	2,481	2,318	36.00	GCA_024752375.1	2022/8/27
<i>Acropora_awi</i>	Cnidaria	Scaffold	428,844,380	1,089,856	22,742	33.80	GCA_014634005.1	2020/9/5
<i>Heliopora_coerulea</i>	Cnidaria	Contig	429,909,776	-	1,427,998	37.00	GCA_030142905.1	2023/5/31
<i>Chrysaora_achlyos</i>	Cnidaria	Scaffold	432,406,514	506,867	14,028	42.10	GCA_015164055.1	2020/11/1
<i>Epizoanthus_ramosus</i>	Cnidaria	Scaffold	432,671,475	4,068	4,019	39.00	GCA_026413505.1	2022/11/27

<i>Phymanthus_crucifer</i>	Cnidaria	Scaffold	437,438,562	570,669	13,886	43.32	GCA_009858155.1	2020/1/13
<i>Acropora_yongei</i>	Cnidaria	Scaffold	438,047,505	3,033,871	45,501	36.40	GCA_014634225.1	2020/9/5
<i>Acropora_florida</i>	Cnidaria	Scaffold	442,824,714	751,376	34,202	35.40	GCA_014634605.1	2020/9/5
<i>Acropora_hyacinthus</i>	Cnidaria	Chromosome	446,425,040	26,527,962	2,162,224	37.50	GCA_020536085.1	2021/10/20
<i>Acropora_digitifera</i>	Cnidaria	Scaffold	447,497,157	483,559	10,915	40.50	GCA_000222465.2	2011/8/1
<i>Chrysaora_chesapeakei</i>	Cnidaria	Scaffold	457,121,517	432,188	14,759	40.28	GCA_011763395.1	2020/3/27
<i>Zoanthus_gigantus</i>	Cnidaria	Scaffold	457,932,737	4,086	4,002	38.50	GCA_025331005.2	2023/6/1
<i>Isaurus_tuberculatus</i>	Cnidaria	Scaffold	460,552,712	3,157	2,916	38.00	GCA_025447995.2	2023/6/1
<i>Hydrozoanthus_gracilis</i>	Cnidaria	Scaffold	461,816,543	2,641	2,350	37.00	GCA_025388745.1	2022/9/23
<i>Chironex_yamaguchii</i>	Cnidaria	Scaffold	467,325,690	12,149	10,044	36.50	GCA_024741275.1	2022/8/24
<i>Porites_rus</i>	Cnidaria	Scaffold	470,011,093	137,187	5,319	38.86	GCA_900290455.1	2018/5/28
<i>Turritopsis_dohrnii</i>	Cnidaria	Contig	474,898,964	-	683,747	34.50	GCA_027922465.1	2022/12/26
<i>Acropora_millepora</i>	Cnidaria	Chromosome	475,381,253	19,840,543	1,091,365	37.10	GCF_013753865.1	2020/7/29
<i>Hydractinia_symbiolongicarpus</i>	Cnidaria	Chromosome	480,922,349	33,314,500	25,832,406	35.50	GCF_029227915.1	2023/6/5
<i>Telmatactis_stephensoni</i>	Cnidaria	Contig	484,670,911	-	602,521	37.00	GCA_029948255.1	2023/5/4
<i>Orbicella_faveolata</i>	Cnidaria	Scaffold	485,548,939	1,162,446	12,468	41.90	GCA_002042975.1	2017/3/20
<i>Bergia_puertoricense</i>	Cnidaria	Scaffold	485,766,386	3,882	3,263	37.50	GCA_026167885.1	2022/11/14
<i>Zoanthus_sanctae-helenae</i>	Cnidaria	Scaffold	486,054,230	5,343	5,281	36.00	GCA_028564885.2	2023/6/7
<i>Antipathozoanthus_remengesai</i>	Cnidaria	Scaffold	492,361,077	3,493	3,453	39.00	GCA_025434805.1	2022/9/27
<i>Antipathozoanthus_obscurus</i>	Cnidaria	Scaffold	501,183,593	816	816	53.00	GCA_024741335.1	2022/8/24
<i>Montipora_sp._Colony_1_RG-2022</i>	Cnidaria	Contig	514,221,667	-	47,391	39.00	GCA_023091755.1	2022/4/22
<i>Ricordea_florida</i>	Cnidaria	Chromosome	537,844,786	33,962,563	2,968,506	38.00	GCA_949710005.1	2023/3/28
<i>Zoanthus_sansibaricus</i>	Cnidaria	Scaffold	537,977,332	2,499	2,495	39.50	GCA_026122405.2	2023/6/7
<i>Umimayanthus_nakama</i>	Cnidaria	Scaffold	539,628,319	5,405	4,585	38.00	GCA_027574985.1	2023/1/6
<i>Zoanthus_solanderi</i>	Cnidaria	Scaffold	543,858,579	3,285	3,259	39.00	GCA_024336925.2	2023/6/7
<i>Parazoanthus_darwini</i>	Cnidaria	Scaffold	549,208,561	2,031	2,028	39.50	GCA_025403605.1	2022/9/26
<i>Palythoa_grandis</i>	Cnidaria	Scaffold	554,359,501	2,366	2,260	38.50	GCA_025448215.2	2023/6/1
<i>Desmophyllum_pertusum</i>	Cnidaria	Scaffold	556,858,542	1,614,486	438,598	39.50	GCA_029204205.1	2023/3/8
<i>Palythoa_caribaeorum</i>	Cnidaria	Scaffold	564,042,941	2,992	2,950	39.50	GCA_026546465.2	2023/5/25
<i>Hydractinia_echinata</i>	Cnidaria	Scaffold	565,047,976	904,215	554,117	35.50	GCA_030067915.1	2023/5/24
<i>Epizoanthus_illoricatus</i>	Cnidaria	Scaffold	566,019,393	2,832	2,667	39.50	GCA_026546655.1	2022/11/30
<i>Porites_australiensis</i>	Cnidaria	Scaffold	576,510,620	554,882	17,662	35.30	GCA_022179025.1	2022/1/14
<i>Trachythela_sp._YZ-2020</i>	Cnidaria	Contig	578,189,354	3,563,727	3,563,727	35.90	GCA_016169945.1	2020/12/21
<i>Phenganax_marumi</i>	Cnidaria	Scaffold	583,317,065	1,311	1,303	46.50	GCA_025400095.1	2022/9/25
<i>Parazoanthus_atlanticus</i>	Cnidaria	Scaffold	592,849,034	2,987	2,971	39.50	GCA_025448195.2	2023/6/7
<i>Parazoanthus_swiftii</i>	Cnidaria	Scaffold	593,601,463	3,655	3,150	38.50	GCA_025403585.1	2022/9/26
<i>Porites_evermanni</i>	Cnidaria	Scaffold	603,805,388	171,385	33,681	39.00	GCA_942486025.1	2023/5/15
<i>Paramuricea_clavata</i>	Cnidaria	Scaffold	606,969,498	23,918	19,152	36.50	GCA_902702795.2	2022/11/8
<i>Cladonema_radiatum</i>	Cnidaria	Scaffold	618,598,448	2,803	2,566	33.00	GCA_025330745.1	2022/9/20
<i>Dendrophyllia_cribrrosa</i>	Cnidaria	Scaffold	627,238,274	48,620,881	2,020,450	38.50	GCA_024195265.1	2022/7/8
<i>Montipora_efflorescens</i>	Cnidaria	Scaffold	643,322,625	1,132,343	24,965	36.00	GCA_014634505.1	2020/9/5
<i>Porites_lobata</i>	Cnidaria	Contig	646,152,978	-	2,154,615	39.00	GCA_942486035.1	2023/5/15
<i>Montipora_cactus</i>	Cnidaria	Scaffold	652,738,455	898,511	26,103	36.40	GCA_014634245.1	2020/9/5

<i>Catalaphyllia_jardinei</i>	Cnidaria	Scaffold	652,767,473	28,947,999	20,477	36.80	GCA_022496165.1	2022/3/7
<i>Zoanthus_pulchellus</i>	Cnidaria	Scaffold	659,744,211	2,334	2,176	36.50	GCA_027575285.2	2023/6/7
<i>Umimayanthus_kanabou</i>	Cnidaria	Scaffold	663,987,347	2,419	2,291	38.00	GCA_025403525.1	2022/9/26
<i>Palythoa_heliodiscus</i>	Cnidaria	Scaffold	668,467,692	1,734	1,674	37.50	GCA_025403565.2	2023/6/1
<i>Aurelia_coerulea</i>	Cnidaria	Scaffold	669,416,919	637,308	13,659	40.96	GCA_011634815.1	2020/3/23
<i>Craspedacusta_sowerbii</i>	Cnidaria	Scaffold	673,186,662	6,208	4,859	40.93	GCA_003687565.1	2018/10/24
<i>Phenganax_stokvisi</i>	Cnidaria	Scaffold	689,040,755	1,419	1,374	45.00	GCA_025400075.2	2023/6/1
<i>Montipora_capitata</i>	Cnidaria	Chromosome	689,311,190	47,608,193	2,047,503	39.50	GCA_949126865.1	2023/4/8
<i>Palythoa_grandiflora</i>	Cnidaria	Scaffold	739,717,117	1,966	1,960	41.50	GCA_026546935.2	2023/6/1
<i>Sphenopus_marsupialis</i>	Cnidaria	Scaffold	748,645,361	897	889	37.00	GCA_027575005.1	2023/1/6
<i>Hydrozoanthus_sils</i>	Cnidaria	Scaffold	751,578,864	1,772	1,671	39.00	GCA_027575215.2	2023/6/1
<i>Epizoanthus_stellaris</i>	Cnidaria	Scaffold	768,013,631	1,939	1,918	42.50	GCA_025403545.1	2022/9/26
<i>Hydra_vulgaris</i>	Cnidaria	Chromosome	819,386,744	55,045,248	756,747	27.05	GCF_022113875.1	2022/2/14
<i>Umimayanthus_parasiticus</i>	Cnidaria	Scaffold	859,641,056	1,171	1,156	38.50	GCA_027575125.1	2023/1/6
<i>Cladopsammia_gracilis</i>	Cnidaria	Contig	878,658,226	1,308,184	1,308,184	38.80	GCA_022457155.1	2022/3/3
<i>Phenganax_subtilis</i>	Cnidaria	Scaffold	880,046,452	1,082	1,039	46.00	GCA_025434845.1	2022/9/27
<i>Palythoa_mutuki</i>	Cnidaria	Scaffold	916,983,906	828	828	46.00	GCA_027575235.2	2023/6/7
<i>Umimayanthus_chanpuru</i>	Cnidaria	Scaffold	918,735,318	1,426	1,405	38.00	GCA_025448255.1	2022/9/27
<i>Morbakka_virulenta</i>	Cnidaria	Scaffold	951,575,644	2,173,999	11,447	35.59	GCA_003991215.1	2019/1/3
<i>Pachyseris_speciosa</i>	Cnidaria	Contig	984,430,632	766,649	766,649	39.60	GCA_016490735.1	2021/1/5
<i>Bythotiarra_depressa</i>	Cnidaria	Scaffold	1,037,087,141	2,252	2,070	33.00	GCA_027564595.1	2023/1/3
<i>Alatina_alata</i>	Cnidaria	Scaffold	1,075,357,049	7,756	7,514	38.03	GCA_008930755.1	2019/10/9
<i>Platygyra_sinensis</i>	Cnidaria	Contig	1,096,359,891	32,425	32,425	39.20	GCA_019787425.1	2021/8/27
<i>Carybdea_marsupialis_au</i> <i>ct._non_(Linnaeus,_1758)</i>	Cnidaria	Scaffold	1,218,614,357	2,268	2,213	38.00	GCA_010016065.2	2023/6/1
<i>Scolionema_suvaense</i>	Cnidaria	Scaffold	1,259,201,252	2,530	2,468	39.00	GCA_027575255.1	2023/1/6
<i>Alatinidae_sp._Z8VKAUB</i> <i>7J3</i>	Cnidaria	Scaffold	1,323,732,878	38,669,433	12,229	39.00	GCA_010016025.2	2023/6/1
<i>Beroe_ovata</i>	Ctenophora	Scaffold	91,597,139	233,527	18,289	41.50	GCA_946803715.1	2022/9/15
<i>Hormiphora_californensis</i>	Ctenophora	Chromosome	110,691,255	8,537,259	580,398	43.13	GCA_020137815.1	2021/10/5
<i>Mnemiopsis_leidy</i>	Ctenophora	Scaffold	155,865,547	187,314	11,914	39.10	GCA_000226015.1	2011/9/19
<i>Pleurobrachia_bachei</i>	Ctenophora	Scaffold	156,121,975	20,628	6,132	37.60	GCA_000695325.1	2014/5/21
<i>Beroe_forskalii</i>	Ctenophora	Scaffold	218,981,974	382,677	208,539	40.20	GCA_011033025.1	2020/3/5
<i>Bolinopsis_microptera</i>	Ctenophora	Chromosome	265,435,384	20,825,282	3,068,845	38.50	GCA_026151205.1	2022/11/14
<i>Dicyema_japonicum</i>	Dicyemida	Scaffold	67,542,437	1,000,208	128,166	36.70	GCA_011109175.1	2020/3/10
<i>Parastichopus_regalis</i>	Echinodermata	Scaffold	171,204,695	314	313	37.00	GCA_027574725.1	2023/1/6
<i>Enypniastes_eximia</i>	Echinodermata	Scaffold	171,706,426	924	923	37.00	GCA_025399915.1	2022/9/25
<i>Sclerodactyla_briareus</i>	Echinodermata	Contig	338,973,358	-	284,909	42.50	GCA_026419785.1	2022/11/27
<i>Acanthaster_planci</i>	Echinodermata	Scaffold	383,860,178	1,521,119	49,680	41.50	GCF_001949145.1	2016/12/20
<i>Pisaster_ochraceus</i>	Echinodermata	Chromosome	401,943,971	20,188,303	7,642	40.36	GCA_010994315.1	2020/2/24
<i>Asterias_rubens</i>	Echinodermata	Chromosome	417,601,740	20,558,067	1,383,654	38.77	GCF_902459465.1	2020/2/12
<i>Orphnurgus_asper</i>	Echinodermata	Scaffold	460,520,177	4,119	3,030	39.50	GCA_027574785.1	2023/1/6
<i>Luidia_sarsii</i>	Echinodermata	Chromosome	471,186,708	23,238,236	755,984	44.00	GCA_949987565.1	2023/5/3
<i>Apostichopus_leukothele</i>	Echinodermata	Scaffold	480,516,499	1,493,354	29,050	36.90	GCA_010014835.1	2020/1/30

<i>Marthasterias glacialis</i>	Echinodermata	Chromosome	520,943,681	25,218,880	1,872,146	39.45	GCA_911728455.2	2022/1/30
<i>Pisaster brevispinus</i>	Echinodermata	Scaffold	549,786,689	20,488,448	3,702,536	39.50	GCA_023634235.1	2022/6/1
<i>Plazaster borealis</i>	Echinodermata	Chromosome	561,050,340	24,975,817	3,100,000	38.90	GCA_021014325.1	2021/11/30
<i>Hemicentrotus pulcherrimus</i>	Echinodermata	Scaffold	568,911,661	142,559	9,641	30.70	GCA_003118195.1	2018/3/21
<i>Paracaudina chilensis</i>	Echinodermata	Scaffold	575,466,867	3,408	3,334	37.50	GCA_027580055.1	2023/1/6
<i>Anneissia japonica</i>	Echinodermata	Scaffold	589,628,371	623,489	19,051	35.00	GCF_011630105.1	2020/3/23
<i>Nesometra sesokonis</i>	Echinodermata	Scaffold	597,070,271	7,143	6,777	34.50	GCA_025631205.1	2022/10/9
<i>Patiria miniata</i>	Echinodermata	Scaffold	608,344,308	23,093,800	2,102,889	40.55	GCF_015706575.1	2020/12/1
<i>Patiria pectinifera</i>	Echinodermata	Chromosome	615,361,208	25,782,902	875,000	40.50	GCA_029964075.1	2023/5/12
<i>Stichopus horrens</i>	Echinodermata	Scaffold	689,130,727	3,896	2,830	37.50	GCA_009801055.1	2019/12/27
<i>Stichopus chloronotus</i>	Echinodermata	Chromosome	719,130,506	32,706,042	4,484,386	37.92	GCA_021234535.1	2021/12/20
<i>Molpadia intermedia</i>	Echinodermata	Scaffold	734,575,371	2,298	2,240	37.00	GCA_027574765.1	2023/1/6
<i>Echinometra lucunter</i>	Echinodermata	Chromosome	760,415,207	30,013,327	128,000	37.00	GCA_029962745.1	2023/5/12
<i>Apostichopus californicus</i>	Echinodermata	Scaffold	786,358,564	9,679	7,238	39.10	GCA_020650975.1	2021/10/28
<i>Apostichopus japonicus</i>	Echinodermata	Scaffold	804,620,767	487,241	307,423	36.80	GCA_002754855.1	2017/11/6
<i>Lytechinus variegatus</i>	Echinodermata	Chromosome	869,582,435	45,571,165	5,852,084	36.30	GCF_018143015.1	2021/4/29
<i>Apostichopus parvimensis</i>	Echinodermata	Scaffold	873,088,235	89,133	9,587	38.50	GCA_000934455.1	2015/2/27
<i>Heliocidaris tuberculata</i>	Echinodermata	Chromosome	897,798,483	44,002,265	1,137,987	36.50	GCA_025618425.1	2022/10/12
<i>Actinopyga echinites</i>	Echinodermata	Scaffold	899,119,518	1,907	1,894	39.10	GCA_010015985.1	2020/1/30
<i>Strongylocentrotus purpuratus</i>	Echinodermata	Scaffold	921,855,793	37,282,239	2,052,140	37.40	GCF_000002235.5	2005/3/11
<i>Paracentrotus lividus</i>	Echinodermata	Scaffold	927,475,755	41,462,573	37,479	36.00	GCA_940671915.1	2022/5/7
<i>Patiriella regularis</i>	Echinodermata	Scaffold	949,333,185	557	218	37.00	GCA_900067625.1	2017/2/10
<i>Scotoplanes globosa</i>	Echinodermata	Scaffold	975,746,094	2,930	2,906	37.50	GCA_027575025.1	2023/1/6
<i>Lytechinus pictus</i>	Echinodermata	Chromosome	999,463,329	46,031,485	219,603	36.00	GCF_015342785.2	2023/3/8
<i>Zoroaster sp._YZ-2022</i>	Echinodermata	Chromosome	1,001,964,507	40,360,388	376,173	38.00	GCA_029582265.1	2023/4/6
<i>Strongylocentrotus pallidus</i>	Echinodermata	Scaffold	1,044,176,929	23,511,293	2,522	36.50	GCA_030265145.1	2023/6/13
<i>Heliocidaris erythrogramma</i>	Echinodermata	Chromosome	1,064,235,828	52,154,090	955,812	37.50	GCA_025617745.1	2022/10/12
<i>Chiridota heheva</i>	Echinodermata	Scaffold	1,106,931,595	1,221,604	1,217,086	36.20	GCA_020152595.1	2021/9/27
<i>Holothuria glaberrima</i>	Echinodermata	Scaffold	1,127,811,786	1,221,172	19,067	38.60	GCA_009936505.1	2020/1/27
<i>Holothuria scabra</i>	Echinodermata	Scaffold	1,181,446,114	1,604,227	1,604,227	38.00	GCA_026123075.1	2022/11/10
<i>Pentamera pediparva</i>	Echinodermata	Scaffold	1,232,190,579	1,984	1,981	42.50	GCA_025400115.1	2022/9/25
<i>Australostichopus mollis</i>	Echinodermata	Scaffold	1,252,116,408	626	256	33.30	GCA_900067635.1	2020/7/10
<i>Paelopatides confundens</i>	Echinodermata	Scaffold	1,378,868,978	504,687	8,832	37.40	GCA_011317855.1	2020/3/16
<i>Holothuria leucospilota</i>	Echinodermata	Chromosome	1,391,422,550	56,107,593	56,107,593	39.00	GCA_029531755.1	2023/4/6
<i>Psolus chitonoides</i>	Echinodermata	Scaffold	1,597,988,068	2,522	2,507	40.50	GCA_027564615.1	2023/1/3
<i>Eucidaris tribuloides</i>	Echinodermata	Scaffold	2,187,261,492	39,192	6,630	37.20	GCA_001188425.1	2015/4/14
<i>Ophioderma brevispina</i>	Echinodermata	Scaffold	2,684,526,435	48,503	48,378	38.00	GCA_023653785.1	2022/6/6
<i>Ophiothrix spiculata</i>	Echinodermata	Scaffold	2,764,315,159	72,780	6,474	35.10	GCA_000969725.1	2015/2/3
<i>Saccoglossus kowalevskii</i>	Hemichordata	Scaffold	775,823,358	245,820	10,074	38.10	GCA_000003605.1	2009/9/9

<i>Ptychodera flava</i>	Hemichordata	Scaffold	1,228,694,352	196,299	13,428	38.00	GCA_001465055.1	2015/12/1
<i>Muusoctopus eicomar</i>	Mollusca	Scaffold	344,064,830	732	679	36.50	GCA_027122395.1	2022/12/20
<i>Lottia gigantea</i>	Mollusca	Scaffold	359,505,668	1,870,055	96,027	36.00	GCF_000327385.1	2013/1/25
<i>Chrysomallon squamiferum</i>	Mollusca	Chromosome	404,610,835	30,197,426	1,883,489	29.80	GCA_012295275.1	2020/4/15
<i>Lottia scabra</i>	Mollusca	Scaffold	414,045,076	17,311,013	5,334,417	34.00	GCA_029955415.1	2023/5/8
<i>Pomacea maculata</i>	Mollusca	Scaffold	432,264,763	375,864	75,997	40.30	GCA_004794325.1	2019/4/15
<i>Goniobranchus leopardus</i>	Mollusca	Scaffold	432,546,281	1,177	1,176	40.00	GCA_028566475.1	2023/2/9
<i>Pomacea canaliculata</i>	Mollusca	Chromosome	440,143,917	31,531,291	1,072,857	40.80	GCF_003073045.1	2018/5/11
<i>Lanistes nyassanus</i>	Mollusca	Scaffold	507,389,202	317,839	25,785	42.30	GCA_004794575.1	2019/4/15
<i>Wirenia argentea</i>	Mollusca	Scaffold	528,769,856	2,992	2,969	39.50	GCA_025802215.1	2022/10/23
<i>Marisa cornuarietis</i>	Mollusca	Contig	535,287,142	4,359,112	4,359,112	40.90	GCA_004794655.1	2019/4/15
<i>Epimenia babai</i>	Mollusca	Scaffold	537,167,073	5,504	5,209	43.50	GCA_011762755.2	2022/11/30
<i>Lutraria rhynchaena</i>	Mollusca	Contig	543,903,147	2,143,760	2,143,760	33.60	GCA_008271625.1	2019/9/4
<i>Elysia chlorotica</i>	Mollusca	Scaffold	557,480,303	441,954	30,474	36.50	GCA_003991915.1	2019/1/4
<i>Dirona pellucida</i>	Mollusca	Scaffold	562,136,326	2,554	2,307	38.00	GCA_030265205.1	2023/6/13
<i>Pinna nobilis</i>	Mollusca	Scaffold	586,463,971	7,576	5,822	42.00	GCA_016161895.1	2020/12/22
<i>Acanthopleura granulata</i>	Mollusca	Scaffold	606,536,932	23,921,462	1,098,986	35.70	GCA_016165875.1	2020/12/21
<i>Crassostrea hongkongensis</i>	Mollusca	Chromosome	607,923,123	55,855,599	2,576,225	33.50	GCA_015776775.1	2020/12/8
<i>Crassostrea angulata</i>	Mollusca	Chromosome	624,273,838	60,476,782	10,870,631	33.50	GCF_025612915.1	2022/10/7
<i>Ostrea denselamellosa</i>	Mollusca	Scaffold	635,891,909	13,952,798	13,952,798	35.00	GCA_024699665.1	2022/8/15
<i>Crassostrea gigas</i>	Mollusca	Chromosome	647,887,097	58,462,999	1,564,469	33.40	GCF_902806645.1	2020/3/11
<i>Watasenia scintillans</i>	Mollusca	Contig	649,175,396	1,283	1,283	35.20	GCA_015471945.1	2020/11/9
<i>Japetella diaphana</i>	Mollusca	Scaffold	661,078,023	1,134	985	37.50	GCA_029822975.1	2023/4/18
<i>Crassostrea ariakensis</i>	Mollusca	Chromosome	663,145,247	66,336,742	5,932,265	33.50	GCA_020458035.1	2021/10/15
<i>Mopalia vespertina</i>	Mollusca	Scaffold	676,010,105	7,077	6,583	36.50	GCA_030265115.1	2023/6/13
<i>Goniobranchus coi</i>	Mollusca	Scaffold	677,083,788	1,643	1,634	37.00	GCA_025762795.1	2022/10/19
<i>Patella depressa</i>	Mollusca	Chromosome	683,711,263	77,532,951	13,889,351	35.50	GCA_948474765.1	2023/2/3
<i>Crassostrea virginica</i>	Mollusca	Chromosome	684,741,128	75,944,018	1,971,208	34.80	GCF_002022765.2	2017/9/1
<i>Patella vulgata</i>	Mollusca	Chromosome	694,468,775	87,764,759	18,376,887	36.10	GCA_932274485.1	2022/3/10
<i>Physella acuta</i>	Mollusca	Contig	705,887,305	-	491,577	37.00	GCA_028476545.1	2023/2/6
<i>Patella pellucida</i>	Mollusca	Chromosome	712,057,054	87,163,620	3,173,890	36.30	GCA_917208275.1	2021/10/13
<i>Batillaria attramentaria</i>	Mollusca	Contig	717,569,005	1,290,776	1,290,776	44.80	GCA_018292915.1	2021/5/6
<i>Argopecten purpuratus</i>	Mollusca	Scaffold	724,780,576	1,022,003	80,106	35.40	n/a	2018/4/1
<i>Chromodoris orientalis</i>	Mollusca	Scaffold	727,712,454	1,798	1,791	37.50	GCA_028571245.1	2023/2/9
<i>Nautilus pompilius</i>	Mollusca	Contig	729,019,367	1,094,646	1,094,646	36.60	GCA_018389105.1	2021/5/17
<i>Perna viridis</i>	Mollusca	Scaffold	731,874,576	4,106,954	70,096	32.10	GCA_018327765.1	2021/3/19
<i>Enteroctopus dofleini</i>	Mollusca	Scaffold	743,946,058	1,386	1,120	37.00	GCA_027564315.1	2023/1/3
<i>Cerastoderma edule</i>	Mollusca	Chromosome	755,460,783	39,684,391	1,015,184	35.50	GCA_947846245.1	2023/1/5
<i>Magallana hongkongensis</i>	Mollusca	Scaffold	757,928,205	72,332,161	17,202	32.20	GCA_016163765.1	2020/12/21
<i>Biomphalaria pfeifferi</i>	Mollusca	Contig	771,819,456	-	3,253,497	35.50	GCA_030265305.1	2023/6/13
<i>Saccostrea glomerata</i>	Mollusca	Scaffold	788,100,799	804,232	39,540	33.30	GCA_003671525.1	2018/10/22
<i>Elysia marginata</i>	Mollusca	Scaffold	790,322,091	225,654	6,205	32.70	GCA_019649035.1	2021/5/9
<i>Chalmys farreri</i>	Mollusca	Scaffold	816,498,441	555,171	20,229	35.50	n/a	2017/11/23

<i>Alviniconcha_marisindica</i>	Mollusca	Contig	829,610,088	727,552	727,552	45.50	GCA_018857735.1	2021/6/11
<i>Lymnaea_stagnalis</i>	Mollusca	Contig	833,229,131	5,751	5,751	37.40	GCA_900036025.1	2016/3/29
<i>Atrina_japonica</i>	Mollusca	Contig	835,332,209	-	911,473	33.50	GCA_030159075.1	2023/2/20
<i>Argopecten_irradians_irradians</i>	Mollusca	Scaffold	835,595,382	1,533,165	78,654	35.20	GCA_004382745.1	2019/3/26
<i>Biomphalaria_glabrata</i>	Mollusca	Chromosome	850,620,091	48,536,009	10,774,472	36.00	GCF_947242115.1	2022/10/29
<i>Mopalia_swanii</i>	Mollusca	Scaffold	860,836,749	2,400	2,201	36.50	GCA_030265315.1	2023/6/13
<i>Anisus_vortex</i>	Mollusca	Chromosome	869,531,111	45,417,137	3,531,458	35.00	GCA_949126835.1	2023/4/8
<i>Argopecten_irradians_concentricus</i>	Mollusca	Scaffold	874,784,041	1,246,717	63,725	35.20	GCA_004382765.1	2019/3/26
<i>Scapharca_broughtonii</i>	Mollusca	Chromosome	884,566,040	44,995,656	1,797,717	33.70	n/a	2019/7/9
<i>Ostrea_edulis</i>	Mollusca	Chromosome	894,786,893	94,306,699	2,427,391	35.00	GCF_947568905.1	2022/12/16
<i>Cyclina_sinensis</i>	Mollusca	Scaffold	903,119,975	46,470,132	2,587,078	34.20	GCA_012932295.1	2020/4/29
<i>Goniobranchus_geometricus</i>	Mollusca	Scaffold	907,216,199	2,471	2,464	38.00	GCA_028565895.1	2023/2/9
<i>Tritonia_festiva</i>	Mollusca	Scaffold	908,035,819	1,452	1,418	39.00	GCA_030265255.1	2023/6/13
<i>Radix_auricularia</i>	Mollusca	Scaffold	909,750,324	578,730	24,354	33.70	GCA_002072015.1	2017/6/7
<i>Tegillarca_granosa</i>	Mollusca	Chromosome	915,485,251	45,661,362	463,414	31.00	GCA_029721355.1	2023/4/18
<i>Pecten_maximus</i>	Mollusca	Chromosome	918,306,378	44,824,366	1,258,799	36.80	GCF_902652985.1	2019/12/4
<i>Aplysia_californica</i>	Mollusca	Scaffold	927,310,431	917,541	9,586	42.00	GCF_000002075.1	2013/5/15
<i>Plakobranchus_ocellatus</i>	Mollusca	Scaffold	927,888,823	1,453,842	5,479	32.80	GCA_019648995.1	2021/5/9
<i>Goniobranchus_fidelis</i>	Mollusca	Scaffold	929,280,104	2,040	2,036	37.50	GCA_028565935.1	2023/2/9
<i>Pinctada_fucata</i>	Mollusca	Chromosome	930,916,280	66,396,765	2,496,427	35.00	GCA_028142955.1	2022/12/19
<i>Spisula_solida</i>	Mollusca	Chromosome	932,088,914	52,695,421	15,021,671	35.50	GCA_947247005.1	2022/10/30
<i>Phorcus_lineatus</i>	Mollusca	Chromosome	957,834,846	49,790,511	4,940,451	35.10	GCA_921293015.1	2021/12/5
<i>Melanoides_tuberculata</i>	Mollusca	Scaffold	970,667,096	1,061	1,059	41.50	GCA_028565955.2	2023/6/1
<i>Mimachlamys_varia</i>	Mollusca	Chromosome	975,359,695	50,749,371	19,361,000	36.50	GCA_947623455.1	2022/12/19
<i>Mactra_quadrangularis</i>	Mollusca	Chromosome	979,316,952	49,304,579	7,487,386	33.00	GCA_025267735.1	2022/9/25
<i>Sinotia_purificata</i>	Mollusca	Chromosome	984,328,742	141,970,838	37,213,934	34.50	GCA_028829895.1	2023/2/24
<i>Mizuhopecten_yessoensis</i>	Mollusca	Scaffold	987,588,634	803,631	65,014	37.60	GCF_002113885.1	2017/6/12
<i>Pinctada_imbricata</i>	Mollusca	Chromosome	990,984,031	59,032,463	21,518	36.80	GCA_002216045.1	2017/7/20
<i>Biomphalaria_straminea</i>	Mollusca	Scaffold	1,004,745,081	25,272,813	9,530	34.20	GCA_021533235.1	2022/1/20
<i>Amphioctopus_burryi</i>	Mollusca	Scaffold	1,031,232,840	3,795,778	869	37.50	GCA_026873925.1	2022/12/13
<i>Mytilisepta_virgata</i>	Mollusca	Chromosome	1,034,479,799	63,395,483	16,191,354	33.00	GCA_028015205.1	2023/1/31
<i>Tridacna_crocea</i>	Mollusca	Chromosome	1,048,956,009	60,802,674	32,149,973	36.00	GCA_943736015.1	2022/6/24
<i>Muusoctopus_longibrachius</i>	Mollusca	Scaffold	1,067,355,155	2,244	1,743	36.50	GCA_027122495.1	2022/12/20
<i>Hippopus_hippopus</i>	Mollusca	Chromosome	1,068,488,762	63,014,760	4,905,753	36.00	GCA_946811185.1	2022/9/15
<i>Ampullaceana_balthica</i>	Mollusca	Scaffold	1,105,887,359	816,280	435,188	36.00	GCA_944989445.1	2022/7/13
<i>Babylonia_areolata</i>	Mollusca	Scaffold	1,108,395,005	6,321	5,933	43.00	GCA_011634625.1	2020/3/23
<i>Anadara_kagoshimensis</i>	Mollusca	Chromosome	1,115,236,308	60,635,260	1,935,399	33.60	GCA_021292105.1	2021/12/28
<i>Ruditapes_philippinarum</i>	Mollusca	Chromosome	1,123,164,463	345,005	29,238	32.30	GCA_009026015.1	2019/10/15
<i>Ostrea_lurida</i>	Mollusca	Scaffold	1,140,787,867	12,947	7,815	34.50	GCA_903981925.1	2020/7/21

<i>Gigantopelta_aegis</i>	Mollusca	Chromosome	1,149,605,64 2	81,591,406	461,769	37.30	GCF_016097555.1	2020/12/17
<i>Fragum_fragum</i>	Mollusca	Chromosome	1,153,070,08 9	63,927,008	2,298,020	35.00	GCA_946902895.1	2022/9/22
<i>Saxidomus_purpurata</i>	Mollusca	Scaffold	1,161,000,14 7	52,225,674	511,514	34.50	GCA_022818135.1	2022/4/5
<i>Dracogyra_subfuscus</i>	Mollusca	Scaffold	1,161,245,40 4	5,907	4,083	36.50	GCA_016106625.1	2020/12/17
<i>Tridacna_gigas</i>	Mollusca	Chromosome	1,175,943,10 4	68,447,427	9,438,000	36.50	GCA_945859785.2	2022/9/22
<i>Haliotis_cracherodii</i>	Mollusca	Scaffold	1,182,234,24 6	60,096,789	17,462,865	40.60	GCA_022045235.1	2022/2/8
<i>Goniobranchus_kuniei</i>	Mollusca	Scaffold	1,190,951,30 4	2,743	2,720	37.50	GCA_025770095.1	2022/10/20
<i>Mya_arenaria</i>	Mollusca	Chromosome	1,216,068,29 1	71,556,252	3,381,420	35.00	GCF_026914265.1	2022/12/15
<i>Sinonovacula_constricta</i>	Mollusca	Chromosome	1,220,848,27 2	65,929,677	976,936	35.40	GCA_007844125.1	2019/8/8
<i>Bulinus_truncatus</i>	Mollusca	Scaffold	1,221,776,97 9	4,956,851	234,305	36.30	GCA_021962125.1	2022/2/4
<i>Fragum_whitleyi</i>	Mollusca	Chromosome	1,224,640,63 6	69,779,563	2,802,427	35.50	GCA_948146395.1	2023/1/18
<i>Dreissena_rostriformis</i>	Mollusca	Scaffold	1,241,703,71 2	131,390	45,905	33.10	GCA_007657795.1	2019/7/30
<i>Steromphala_cineraria</i>	Mollusca	Chromosome	1,270,484,90 5	70,748,747	6,200,493	35.40	GCA_916613615.1	2021/10/1
<i>Solen_grandis</i>	Mollusca	Chromosome	1,324,491,67 2	67,678,117	50,000	34.80	GCA_021229015.1	2021/12/19
<i>Botula_fusca</i>	Mollusca	Scaffold	1,333,706,66 5	2,171	2,129	34.50	GCA_028566455.1	2023/2/9
<i>Haliotis_rufescens</i>	Mollusca	Scaffold	1,334,454,21 4	45,695,856	8,868,657	40.50	GCF_023055435.1	2022/4/19
<i>Limnoperna_fortunei</i>	Mollusca	Chromosome	1,335,559,95 9	97,051,362	1,498,882	33.50	GCA_944474755.1	2022/7/13
<i>Haliotis_rubra</i>	Mollusca	Scaffold	1,378,265,26 4	1,227,833	1,177,711	40.50	GCA_003918875.1	2018/12/7
<i>Lithophaga_antillarum</i>	Mollusca	Scaffold	1,419,759,20 5	2,680	2,637	33.50	GCA_028566495.1	2023/2/9
<i>Gibbula_magus</i>	Mollusca	Chromosome	1,470,388,82 2	80,454,948	3,416,053	36.00	GCA_936450465.1	2022/5/3
<i>Panopea_generosa</i>	Mollusca	Chromosome	1,474,161,28 9	73,788,920	1,571,249	34.00	GCA_029582155.1	2023/4/6
<i>Lithophaga_nigra</i>	Mollusca	Scaffold	1,492,027,03 2	2,310	2,277	34.00	GCA_028566555.1	2023/2/9
<i>Muusoctopus_leioderma</i>	Mollusca	Scaffold	1,531,747,14 4	3,814	3,109	36.50	GCA_027124095.1	2022/12/20
<i>Archivesica_marissinica</i>	Mollusca	Chromosome	1,544,610,69 5	74,312,544	79,144	39.20	GCA_014843695.1	2020/10/7
<i>Congerina_kusceri</i>	Mollusca	Chromosome	1,550,098,61 8	72,441,503	6,387,040	35.50	GCA_027627225.1	2023/1/13
<i>Mytilus_coruscus</i>	Mollusca	Chromosome	1,566,529,93 8	99,542,347	1,481,111	32.30	GCA_017311375.1	2021/3/10
<i>Venustaconcha_ellipsiformis</i>	Mollusca	Scaffold	1,590,009,04 5	6,657	2,813	33.40	GCA_003401595.1	2018/8/16
<i>Gari_tellinella</i>	Mollusca	Chromosome	1,597,614,75 3	85,279,272	19,181,371	36.50	GCA_922989275.1	2021/12/19
<i>Octopus_insularis</i>	Mollusca	Contig	1,617,927,58 2	-	1,978	36.00	GCA_029874955.1	2023/4/26
<i>Mytilus_edulis</i>	Mollusca	Chromosome	1,651,313,23 6	116,503,18 0	490,737	32.30	GCA_019925275.1	2021/9/14
<i>Mytilus_californianus</i>	Mollusca	Scaffold	1,651,950,17 1	117,871,51 2	16,323,199	32.60	GCA_021869535.1	2022/2/3
<i>Gigantidas_platifrons</i>	Mollusca	Scaffold	1,658,191,95 3	343,341	12,602	30.20	GCA_002080005.1	2017/4/5
<i>Mytilus_galloprovincialis</i>	Mollusca	Chromosome	1,658,655,99 8	71,952,646	26,323	32.00	GCA_025277285.1	2022/9/23

<i>Haliotis laevis</i>	Mollusca	Scaffold	1,762,655,385	81,233	3,353	26.80	GCA_008038995.1	2019/8/20
<i>Potamilus streckersoni</i>	Mollusca	Scaffold	1,776,755,686	2,051,244	2,032,685	33.80	GCA_016746295.1	2021/1/21
<i>Dreissena polymorpha</i>	Mollusca	Chromosome	1,797,942,340	117,515,028	1,111,654	35.10	GCA_020536995.1	2021/10/26
<i>Mercenaria mercenaria</i>	Mollusca	Chromosome	1,858,199,728	82,914,371	82,432,825	34.50	GCF_021730395.1	2022/2/1
<i>Haliotis discus hannai</i>	Mollusca	Scaffold	1,865,475,499	200,099	14,194	40.50	n/a	2017/3/7
<i>Conchocele bisecta</i>	Mollusca	Chromosome	1,904,932,765	106,883,709	515,412	39.00	GCA_029237695.1	2023/3/20
<i>Conus consors</i>	Mollusca	Scaffold	2,049,318,684	1,128	749	39.00	GCA_004193615.1	2019/2/12
<i>Octopus mimus</i>	Mollusca	Scaffold	2,057,065,956	80,823,707	1,141	36.50	GCA_027568975.1	2023/1/3
<i>Phymorhynchus buccinoides</i>	Mollusca	Contig	2,114,589,556	336,037	336,037	42.80	GCA_017654935.1	2021/3/31
<i>Conus tribblei</i>	Mollusca	Scaffold	2,160,493,082	2,681	854	42.50	GCA_001262575.1	2015/8/6
<i>Rapana venosa</i>	Mollusca	Chromosome	2,300,182,741	64,632,560	433,295	42.00	GCA_028751875.1	2023/2/24
<i>Octopus bimaculoides</i>	Mollusca	Scaffold	2,338,188,782	475,182	5,532	37.80	GCF_001194135.1	2015/11/9
<i>Octopus maya</i>	Mollusca	Scaffold	2,341,750,473	89,813,523	2,361	36.50	GCA_027122535.1	2022/12/20
<i>Megalonaia nervosa</i>	Mollusca	Scaffold	2,365,201,850	50,649	50,186	35.80	GCA_016617855.1	2021/1/12
<i>Unio pictorum</i>	Mollusca	Contig	2,434,325,818	-	10,612,599	34.50	GCA_030141615.1	2023/5/31
<i>Margaritifera margaritifera</i>	Mollusca	Contig	2,453,525,054	-	3,425,502	35.00	GCA_029931535.1	2023/5/3
<i>Octopus americanus</i>	Mollusca	Scaffold	2,534,512,599	100,565,213	1,553	37.00	GCA_027568995.1	2023/1/3
<i>Modiolus philippinarum</i>	Mollusca	Scaffold	2,629,556,424	100,161	18,389	32.30	GCA_002080025.1	2017/4/5
<i>Architeuthis dux</i>	Mollusca	Scaffold	2,693,618,964	4,852,590	5,397	30.00	GCA_006491835.1	2019/6/26
<i>Octopus sinensis</i>	Mollusca	Chromosome	2,719,136,158	105,892,736	490,217	36.40	GCF_006345805.1	2019/6/17
<i>Octopus rubescens</i>	Mollusca	Scaffold	2,757,134,468	23,148,921	2,937	37.00	GCA_027564515.1	2023/1/3
<i>Limacina bulimoides</i>	Mollusca	Scaffold	2,901,930,333	893	884	34.10	GCA_009866985.1	2020/1/14
<i>Conus betulinus</i>	Mollusca	Scaffold	3,429,626,206	232,489	182,638	43.80	GCA_016801955.1	2021/2/2
<i>Conus ventricosus</i>	Mollusca	Chromosome	3,592,060,885	93,519,712	114,325	43.90	GCA_018398815.1	2021/5/25
<i>Haplochlarea maculosa</i>	Mollusca	Scaffold	4,009,598,772	931,835	1,130	22.60	GCA_015501135.1	2020/11/16
<i>Doryteuthis pealeii</i>	Mollusca	Chromosome	4,598,001,602	107,400,528	230,938	33.50	GCA_023376005.1	2022/5/25
<i>Sepia pharaonis</i>	Mollusca	Contig	4,785,531,890	1,926,397	1,926,397	33.20	GCA_903632075.3	2021/2/2
<i>Euprymna scolopes</i>	Mollusca	Scaffold	5,280,013,996	3,549,550	3,558	21.80	GCA_004765925.1	2019/4/9
<i>Bunonema sp._RGD898</i>	Nematoda	Scaffold	34,132,167	14,240	8,202	54.99	GCA_014805405.1	2020/9/30
<i>Meloidogyne graminicola</i>	Nematoda	Scaffold	41,549,413	294,907	294,907	23.21	GCA_014773135.1	2020/9/28
<i>Aphelenchoides bicaudatus</i>	Nematoda	Contig	46,428,382	-	11,905,581	38.50	GCA_024699845.1	2022/8/15
<i>Hoplolaimus columbus</i>	Nematoda	Contig	46,941,449	7,546	7,546	41.66	GCA_016433075.1	2020/12/29
<i>Aphelenchoides besseyi</i>	Nematoda	Scaffold	47,428,222	17,779,264	1,080,231	42.00	GCA_024699855.1	2022/8/15
<i>Halicephalobus sp._NKZ32</i>	Nematoda	Scaffold	47,438,423	60,651	38,223	36.98	GCA_009761265.1	2019/12/17

<i>Panagrolaimus_sp._PS1579</i>	Nematoda	Scaffold	48,802,921	2,935	2,091	30.55	GCA_901779485.1	2019/7/31
<i>Trichinella_sp._T9</i>	Nematoda	Scaffold	49,075,249	212,690	113,887	33.00	GCA_001447505.1	2015/11/24
<i>Trichinella_pseudospiralis</i>	Nematoda	Scaffold	49,152,658	287,133	141,979	32.50	GCA_001447645.1	2015/11/24
<i>Trichinella_sp._T8</i>	Nematoda	Scaffold	49,331,761	239,129	120,751	33.50	GCA_001447745.1	2015/11/24
<i>Trichinella_patagoniensis</i>	Nematoda	Scaffold	49,773,209	154,103	121,031	33.50	GCA_001447655.1	2015/11/24
<i>Rhabditida_sp._S2_005_001R2</i>	Nematoda	Scaffold	50,500,136	8,213	8,120	38.50	GCA_004026265.1	2019/1/16
<i>Trichinella_sp._T6</i>	Nematoda	Scaffold	50,853,738	158,103	106,131	33.00	GCA_001447435.1	2015/11/24
<i>Strongyloides_papillosus</i>	Nematoda	Contig	58,232,448	-	1,445,102	25.50	GCA_005656395.1	2019/5/22
<i>Caenorhabditis_sp._36_P RJB53466</i>	Nematoda	Scaffold	59,029,622	9,393,127	2,590,546	45.00	GCA_946814055.1	2022/9/15
<i>Trichinella_sp._17WV049- YT159</i>	Nematoda	Scaffold	59,190,369	7,343	7,221	34.00	GCA_011764315.1	2020/3/27
<i>Panagrellus_sp._LJ2285</i>	Nematoda	Contig	62,270,399	-	10,706,094	44.00	GCA_024447235.1	2022/7/26
<i>Trichinella_spiralis</i>	Nematoda	Scaffold	63,525,422	6,373,445	76,707	33.50	GCF_000181795.1	2011/1/11
<i>Panagrolaimus_sp._JU765</i>	Nematoda	Scaffold	64,262,255	10,861	1,669	35.89	GCA_901765185.1	2019/7/31
<i>Panagrellus_redivivus</i>	Nematoda	Scaffold	65,093,147	262,414	19,707	44.25	GCA_000341325.1	2013/2/22
<i>Caenorhabditis_sp._32_LS -2015</i>	Nematoda	Scaffold	65,098,870	136,669	26,457	46.41	GCA_900536325.3	2019/4/11
<i>Panagrellus_sp._LJ2284</i>	Nematoda	Contig	74,454,495	-	11,844,156	44.00	GCA_024447225.1	2022/7/26
<i>Caenorhabditis_panamensis</i>	Nematoda	Scaffold	79,045,575	487,150	89,877	42.00	GCA_900883565.2	2019/4/11
<i>Caenorhabditis_tropicalis</i>	Nematoda	Chromosome	81,560,419	13,926,232	5,142,224	37.50	GCA_016735795.1	2021/1/25
<i>Panagrolaimus_sp._PS1159</i>	Nematoda	Scaffold	84,957,092	9,924	2,144	28.19	GCA_901765195.1	2019/7/31
<i>Caenorhabditis_becei</i>	Nematoda	Scaffold	87,867,412	767,505	71,727	42.00	GCA_900536315.3	2019/4/11
<i>Dirofilaria_immitis</i>	Nematoda	Contig	89,153,617	284,390	284,390	29.73	GCA_009829315.1	2020/1/3
<i>Panagrolaimus_sp._ES5</i>	Nematoda	Scaffold	90,240,976	6,312	2,687	28.86	GCA_901766855.1	2019/7/31
<i>Caenorhabditis_sp._39_LS -2015</i>	Nematoda	Scaffold	91,416,478	15,123	8,426	41.91	GCA_900536345.3	2019/4/11
<i>Caenorhabditis_sp._21_LS -2015</i>	Nematoda	Scaffold	93,733,926	44,405	12,682	39.29	GCA_900536235.3	2019/4/11
<i>Caenorhabditis_sp._38_M B-2015</i>	Nematoda	Scaffold	100,371,645	139,430	75,067	34.81	GCA_900536415.3	2019/4/11
<i>Caenorhabditis_sp._26_LS -2015</i>	Nematoda	Scaffold	101,086,038	91,328	44,851	38.31	GCA_900536285.3	2019/4/11
<i>Caenorhabditis_sp._40_LS -2015</i>	Nematoda	Scaffold	101,228,815	224,519	75,561	41.51	GCA_900536305.3	2019/4/11
<i>Caenorhabditis_sp._31_LS -2015</i>	Nematoda	Scaffold	104,046,868	177,024	39,592	36.01	GCA_900536295.3	2019/4/11
<i>Caenorhabditis_inopinata</i>	Nematoda	Chromosome	122,994,328	20,594,552	9,051,690	38.00	GCA_003052745.1	2017/6/22
<i>Ditylenchus_destructor</i>	Nematoda	Contig	139,411,310	-	781,965	37.00	GCA_022814885.1	2022/4/4
<i>Anisakis_simplex</i>	Nematoda	Scaffold	144,322,410	6,833	6,798	37.30	GCA_900576815.1	2019/3/4
<i>Mesodorylaimus_sp._YZB24</i>	Nematoda	Scaffold	147,972,742	409,169	409,169	45.50	GCA_030248295.1	2023/6/9
<i>Theristus_sp._LFF4_11</i>	Nematoda	Contig	169,171,958	-	57,348	43.50	GCA_030248135.1	2023/6/9
<i>Epsilonema_sp._ZAB3_2</i>	Nematoda	Contig	171,870,121	-	36,920	45.50	GCA_030247985.1	2023/6/9
<i>Paralinhomoeus_sp._GSC O2_6</i>	Nematoda	Scaffold	177,079,101	79,981	79,981	52.00	GCA_030248085.1	2023/6/9
<i>Heterodera_schachtii</i>	Nematoda	Scaffold	179,246,932	1,273,070	286,103	33.23	GCA_019095935.1	2021/7/2
<i>Panagrolaimus_sp._LJ2400</i>	Nematoda	Contig	182,164,411	-	7,797,374	30.00	GCA_024447215.1	2022/7/26
<i>Plectus_sambesii</i>	Nematoda	Scaffold	186,672,612	23,450	15,583	42.48	GCA_002796945.1	2017/11/28

<i>Panagrolaimus_sp._LJ241</i> 4	Nematoda	Contig	217,946,463	-	3,283,130	30.00	GCA_024447195.1	2022/7/26
<i>Ditylenchus_dipsaci</i>	Nematoda	Scaffold	227,234,012	287,390	246,852	37.50	GCA_004194705.1	2019/2/14
<i>Ptycholaimellus_sp._GST1</i> _10	Nematoda	Contig	238,323,431	-	33,268	48.00	GCA_030248095.1	2023/6/9
<i>Linhomoeus_sp._GSCO2_</i> 2	Nematoda	Contig	240,517,537	-	28,911	49.00	GCA_030248115.1	2023/6/9
<i>Acrobeloides_nanus</i>	Nematoda	Scaffold	248,054,631	19,614	8,635	35.47	GCA_900406225.1	2018/6/9
<i>Aphelenchus_avenae</i>	Nematoda	Scaffold	264,011,204	141,150	10,739	49.72	GCA_020875895.1	2021/11/15
<i>Panagrolaimus_sp._k_MP-</i> 2021	Nematoda	Contig	265,835,997	-	3,823,143	29.50	GCA_028622995.1	2023/2/16
<i>Enoplolaimus_lenunculus</i>	Nematoda	Scaffold	300,868,979	164,672	164,672	38.00	GCA_030248215.1	2023/6/9
<i>Romanomermis_culicivora</i> x	Nematoda	Scaffold	322,765,759	17,632	1,417	36.27	GCA_001039655.1	2014/1/8
<i>Sabatieria_punctata</i>	Nematoda	Scaffold	390,719,928	91,013	90,858	43.00	GCA_030248155.1	2023/6/9
<i>Trissonchulus_sp._WLG1_</i> 4	Nematoda	Scaffold	446,957,608	36,265	36,222	38.50	GCA_030248245.1	2023/6/9
<i>Microlaimidae_sp._YZB2_</i> 3	Nematoda	Contig	575,776,021	-	24,089	32.50	GCA_030248175.1	2023/6/9
<i>Trileptium_ribeirensis</i>	Nematoda	Contig	745,954,240	-	41,950	38.50	GCA_030248225.1	2023/6/9
<i>Trissonchulus_latispiculu</i> m	Nematoda	Scaffold	792,433,438	40,542	40,526	40.00	GCA_030248195.1	2023/6/9
<i>Lineus_longissimus</i>	Nemertea	Chromosome	391,180,816	20,961,375	9,910,080	41.41	GCA_910592395.2	2022/2/13
<i>Notospermus_geniculatus</i>	Nemertea	Scaffold	858,599,399	239,235	22,604	42.90	GCA_002633025.1	2017/10/23
<i>Intoshia_linei</i>	Orthonectida	Scaffold	41,603,068	26,268	25,991	25.90	GCA_001642005.1	2016/5/10
<i>Phoronis_ovalis</i>	Phoronida	Scaffold	326,042,390	5,749	4,788	34.00	GCA_028565635.1	2023/2/9
<i>Phoronis_pallida</i>	Phoronida	Scaffold	425,364,260	2,362	2,318	36.00	GCA_028565715.1	2023/2/9
<i>Phoronis_australis</i>	Phoronida	Scaffold	498,443,662	655,058	68,151	38.50	GCA_002633005.1	2017/10/23
<i>Trichoplax_sp._H2</i>	Placozoa	Scaffold	94,878,730	376,320	91,942	32.75	GCA_003344405.1	2018/7/27
<i>Trichoplax_adhaerens</i>	Placozoa	Scaffold	105,631,681	5,978,658	204,191	34.50	GCA_000150275.1	2008/6/17
<i>Gyrodactylus_salaris</i>	Platyhelminthes	Scaffold	67,380,660	18,390	14,673	33.90	GCA_000715275.1	2014/6/27
<i>Gyrodactylus_bullatarudis</i>	Platyhelminthes	Scaffold	84,346,884	308,184	124,054	30.90	GCA_012064415.1	2020/4/6
<i>Taenia_crassiceps</i>	Platyhelminthes	Scaffold	107,053,072	10,553,375	3,809,406	43.50	GCA_023375655.1	2022/5/16
<i>Cardicola_forsteri</i>	Platyhelminthes	Contig	216,531,548	-	430,422	28.50	GCA_025895645.1	2022/10/28
<i>Macrostomum_hystrix</i>	Platyhelminthes	Contig	220,453,370	-	13,505,968	45.00	GCA_950097015.1	2023/6/1
<i>Macrostomum_cliftonense</i>	Platyhelminthes	Contig	226,746,154	-	1,017,626	44.50	GCA_950096795.1	2023/6/1
<i>Schistosoma_mansoni</i>	Platyhelminthes	Chromosome	364,518,427	32,115,376	76,721	35.00	GCF_000237925.1	2011/12/13
<i>Schistosoma_bovis</i>	Platyhelminthes	Chromosome	391,788,201	47,558,671	10,931,731	34.50	GCA_944470425.2	2022/9/20
<i>Schistosoma_haematobium</i>	Platyhelminthes	Chromosome	400,271,889	48,328,128	20,231,032	35.00	GCF_000699445.3	2022/2/10
<i>Schistosoma_japonicum</i>	Platyhelminthes	Chromosome	406,576,294	49,539,924	2,378,822	33.50	GCA_025215515.1	2022/9/16
<i>Dibothriocephalus_latus</i>	Platyhelminthes	Scaffold	531,434,409	6,744	2,113	43.19	GCA_900617775.1	2018/12/5
<i>Clonorchis_sinensis</i>	Platyhelminthes	Chromosome	558,124,894	168,711,085	760,604	44.07	GCA_003604175.2	2021/4/12
<i>Atriophallophorus_winterbourni</i>	Platyhelminthes	Scaffold	601,728,533	40,108	33,380	40.73	GCA_013407085.1	2020/7/13
<i>Protopolystoma_xenopodis</i>	Platyhelminthes	Scaffold	617,344,661	2,899	2,191	37.72	GCA_900617795.1	2019/2/12
<i>Opisthorchis_viverrini</i>	Platyhelminthes	Scaffold	620,452,578	1,349,843	36,994	43.50	GCF_000715545.1	2014/7/9
<i>Schistocephalus_solidus</i>	Platyhelminthes	Scaffold	625,207,408	121,651	120,142	43.66	GCA_017591395.1	2021/3/26
<i>Opisthorchis_felineus</i>	Platyhelminthes	Scaffold	679,245,959	621,022	54,248	39.40	GCA_004794785.1	2019/4/15
<i>Paragonimus_kellicotti</i>	Platyhelminthes	Scaffold	696,536,264	56,030	7,921	42.50	GCA_014220965.1	2020/8/17
<i>Fasciolopsis_buski</i>	Platyhelminthes	Scaffold	748,157,630	180,478	51,097	42.50	GCA_008360955.1	2019/9/9

<i>Macrostomum lignano</i>	Platyhelminthes	Scaffold	764,410,778	246,204	215,276	45.80	GCA_002269645.1	2017/8/24
<i>Schmidtea mediterranea</i>	Platyhelminthes	Chromosome	773,974,862	262,873,270	1,113,910	29.63	GCA_022537955.1	2022/3/10
<i>Paragonimus heterotremus</i>	Platyhelminthes	Scaffold	841,211,332	92,457	23,957	42.00	GCA_013368495.1	2020/6/23
<i>Dugesia japonica</i>	Platyhelminthes	Scaffold	854,187,087	1,740	1,068	26.84	GCA_001938525.1	2017/1/6
<i>Paragonimus westermani</i>	Platyhelminthes	Scaffold	922,829,509	135,171	42,170	43.00	GCA_008508345.1	2019/9/18
<i>Eudiplozoon nipponicum</i>	Platyhelminthes	Contig	939,802,929	-	87,067	34.50	GCA_029291075.1	2023/3/20
<i>Trichobilharzia regenti</i>	Platyhelminthes	Chromosome	954,242,115	125,840,967	3,882,242	38.00	GCA_944472135.2	2022/9/15
<i>Fasciola gigantica</i>	Platyhelminthes	Scaffold	1,347,749,759	4,892,498	4,892,498	44.00	GCA_018104335.1	2021/4/20
<i>Girardia tigrina</i>	Platyhelminthes	Scaffold	1,428,149,015	12,094	4,936	29.43	GCA_001938485.1	2017/1/6
<i>Prostheceraeus crozeri</i>	Platyhelminthes	Scaffold	2,065,465,794	292,050	287,821	37.50	GCA_907163375.1	2022/7/16
<i>Oopsacas minuta</i>	Porifera	Scaffold	61,460,524	676,369	303,044	35.50	GCA_024704765.1	2022/8/16
<i>Oscarella lobularis</i>	Porifera	Chromosome	65,248,392	3,475,296	506,373	46.00	GCA_947507565.1	2022/12/6
<i>Halichondria panicea</i>	Porifera	Contig	73,970,439	-	2,555	42.00	GCA_020423275.1	2021/10/12
<i>Hexactinellida sp. DTS-2022</i>	Porifera	Scaffold	112,398,167	5,914,077	5,377,299	36.00	GCA_028753945.1	2023/2/21
<i>Chondrosia reniformis</i>	Porifera	Chromosome	117,372,766	8,459,200	4,706,000	37.00	GCA_947172415.1	2022/10/27
<i>Amphimedon queenslandica</i>	Porifera	Scaffold	166,699,561	120,365	11,817	37.50	GCF_000090795.1	2010/5/28
<i>Petrosia ficiformis</i>	Porifera	Chromosome	191,254,924	9,942,894	3,950,187	33.50	GCA_947044365.1	2022/10/7
<i>Spongilla lacustris</i>	Porifera	Chromosome	248,704,581	10,124,357	254,100	43.00	GCA_949361645.1	2023/3/11
<i>Agelas oroides</i>	Porifera	Chromosome	260,658,830	10,369,553	2,293,598	44.00	GCA_949130485.1	2023/4/7
<i>Ephydatia muelleri</i>	Porifera	Scaffold	322,619,961	9,883,643	211,556	43.19	GCA_013339895.1	2020/6/12
<i>Nicola tetela</i>	Porifera	Scaffold	459,937,682	4,652	3,708	43.50	GCA_028565695.1	2023/2/9
<i>Clathrina lutea</i>	Porifera	Scaffold	596,185,498	3,617	3,558	46.50	GCA_028571005.1	2023/2/9
<i>Cladorhizidae sp. MBARI D1340-BT2</i>	Porifera	Scaffold	857,639,048	44,572,050	1,326,106	41.50	GCA_028752905.1	2023/2/21
<i>Priapulus caudatus</i>	Priapula	Scaffold	511,738,253	209,727	10,617	45.70	GCA_000485595.2	2013/10/29
<i>Tubiluchus corallicola</i>	Priapulida	Scaffold	345,366,841	363,593	296,101	35.50	GCA_030141605.1	2023/5/31
<i>Proales similis</i>	Rotifera	contig	32,722,929	2,428,301	2,428,301	45.51	GCA_019059635.1	2021/6/29
<i>Seison nebaliae</i>	Rotifera	Contig	43,872,782	-	45,684	31.50	GCA_023231475.1	2022/5/3
<i>Brachionus angularis</i>	Rotifera	Contig	56,513,240	5,422,639	5,422,639	23.65	GCA_021432725.1	2022/1/11
<i>Brachionus rotundiformis</i>	Rotifera	Scaffold	57,707,736	2,691,490	169,622	28.19	GCA_016802295.1	2021/2/2
<i>Brachionus sp. 'Tiscar'</i>	Rotifera	Contig	70,983,965	24,091	24,091	26.18	GCA_006352455.1	2019/6/18
<i>Brachionus koreanus</i>	Rotifera	Scaffold	85,731,059	1,856,740	25,075	24.40	GCA_009177125.1	2019/10/18
<i>Brachionus plicatilis</i>	Rotifera	Scaffold	106,939,251	1,156,906	55,822	26.75	GCA_010279815.1	2020/2/6
<i>Brachionus manjavacas</i>	Rotifera	scaffold	114,172,357	6,365,377	6,365,377	27.99	GCA_019054805.1	2021/6/29
<i>Brachionus calyciflorus</i>	Rotifera	Scaffold	129,626,318	785,733	25,945	25.86	GCA_002922825.1	2018/2/7
<i>Brachionus rubens</i>	Rotifera	Contig	132,737,077	2,506,101	2,506,101	29.96	GCA_022207745.1	2022/2/11
<i>Adineta ricciae</i>	Rotifera	Scaffold	135,599,987	283,762	178,075	35.50	GCA_905250025.1	2021/3/19
<i>Adineta steineri</i>	Rotifera	Scaffold	171,140,834	200,132	32,736	29.00	GCA_905250115.1	2021/3/20
<i>Rotaria magnacalcarata</i>	Rotifera	Scaffold	180,458,959	53,250	24,068	31.90	GCA_900239745.1	2018/2/10
<i>Brachionus asplanchnoides</i>	Rotifera	Contig	230,116,674	-	3,065,462	30.00	GCA_025491095.1	2022/9/29
<i>Rotaria macrura</i>	Rotifera	Scaffold	234,715,442	73,242	55,551	32.80	GCA_900239685.1	2018/2/10
<i>Rotaria sp. Silwood2</i>	Rotifera	Scaffold	296,468,297	102,503	16,004	31.02	GCA_905329745.1	2021/3/20

<i>Rotaria_sp._Silwood1</i>	Rotifera	Scaffold	475,793,899	298,918	50,874	30.52	GCA_905250075.1	2021/3/21
<i>Macrotrachela_quadricornifera</i>	Rotifera	Scaffold	525,309,123	242,167	32,858	29.58	GCA_022393335.1	2022/2/25
<i>Ramazzottius_varieornatus</i>	Tardigrada	Scaffold	55,828,384	4,740,345	130,106	47.50	GCA_001949185.1	2016/11/9
<i>Hypsibius_exemplaris</i>	Tardigrada	Scaffold	104,154,999	342,180	76,809	45.00	GCA_002082055.1	2017/4/7
<i>Hofstenia_miamia</i>	Xenacoelomorpha	Scaffold	949,845,848	1,044,515	9,279	31.81	GCA_900660155.1	2019/6/19
<i>Symsagittifera_roscoffensis</i>	Xenacoelomorpha	Scaffold	1,103,011,097	1,039,899	239,323	36.50	GCA_025153585.1	2022/9/14

附表 3. 已公布的鱼类基因组信息

Species	Order	Assembly Level	Genome size (bp)	Scaffold N50 (bp)	Contig N50 (bp)	Accession	Release Date
<i>Callorhinchus milii</i>	Chimaeriformes	Scaffold	974,498,586	4,521,921	46,577	GCA_000165045.2	2013/12/11
<i>Chimaera phantasma</i>	Chimaeriformes	Scaffold	1,149,020,171	6,239,293	#N/A	Newly sequenced	#N/A
<i>Amblyraja radiata</i>	Rajiformes	Chromosome	2,558,801,738	62,052,929	1,462,921	GCA_010909765.1	2020/2/21
<i>Raja kenoei</i>	Rajiformes	Scaffold	3,185,554,497	1,299,035	#N/A	Newly sequenced	#N/A
<i>Torpedinidae sp.</i>	Torpediniformes	Scaffold	2,610,681,668	2,881	#N/A	Newly sequenced	#N/A
<i>Torpedo nobiliana</i>	Torpediniformes	Scaffold	2,078,615,216	3,804	#N/A	Newly sequenced	#N/A
<i>Pristis pectinata</i>	Pristiformes	Chromosome	2,267,857,790	101,699,621	17,009,522	GCA_009764475.1	2019/12/19
<i>Rhina ancylostoma</i>	Pristiformes	Scaffold	2,401,038,731	12,966,293	#N/A	Newly sequenced	#N/A
<i>Gymnura bimaculata</i>	Myliobatiformes	Scaffold	3,423,079,185	9,666,153	#N/A	Newly sequenced	#N/A
<i>Myliobatis tobijei</i>	Myliobatiformes	Scaffold	3,152,792,569	6,274,978	#N/A	Newly sequenced	#N/A
<i>Dasyatidae</i>	Myliobatiformes	Scaffold	2,020,857,199	3,438	#N/A	Newly sequenced	#N/A
<i>Plesiobatis daviesi</i>	Myliobatiformes	Scaffold	3,743,991,489	15,043	#N/A	Newly sequenced	#N/A
<i>Chiloscyllium plagiosum</i>	Orectolobiformes	Chromosome	3,776,550,609	72,140,027	37,239	GCA_004010195.1	2019/1/10
<i>Chiloscyllium punctatum</i>	Orectolobiformes	Scaffold	3,375,698,300	1,963,924	21,276	GCA_003427335.1	2018/8/2
<i>Rhincodon typus</i>	Orectolobiformes	Scaffold	2,931,599,576	144,422	144,422	GCA_001642345.2	2017/3/3
<i>Cirrhoscyllium formosanum</i>	Orectolobiformes	Scaffold	2,483,794,455	3,446	#N/A	Newly sequenced	#N/A
<i>Alopias pelagicus</i>	Lamniformes	Scaffold	3,209,220,808	7,803	#N/A	Newly sequenced	#N/A
<i>Carcharhinidae sp.</i>	Carcharhiniformes	Scaffold	2,991,061,848	4,010,668	#N/A	Newly sequenced	#N/A
<i>Hemitriakis japonica</i>	Carcharhiniformes	Scaffold	3,491,787,993	377,205	#N/A	Newly sequenced	#N/A
<i>Mustelus kanekonis</i>	Carcharhiniformes	Scaffold	3,325,085,774	106,421	#N/A	Newly sequenced	#N/A
<i>Scoliodon sorrakowah</i>	Carcharhiniformes	Scaffold	2,990,483,579	763,652	#N/A	Newly sequenced	#N/A
<i>Scyliorhinus torazame</i>	Carcharhiniformes	Scaffold	4,470,980,620	282,424	7,146	GCA_003427355.1	2018/8/2
<i>Carcharhinus melanopterus</i>	Carcharhiniformes	Scaffold	2,757,214,068	25,271	#N/A	Newly sequenced	#N/A
<i>Hemipristis elongatus</i>	Carcharhiniformes	Scaffold	3,064,718,838	69,096	#N/A	Newly sequenced	#N/A
<i>Scyliorhinus canicula</i>	Carcharhiniformes	Chromosome	4,220,406,627	198,790,641	1,862,211	GCA_902713615.1	2020/12/17
<i>Centrophorus acus</i>	Squaliformes	Scaffold	2,330,407,065	4,140	#N/A	Newly sequenced	#N/A
<i>Squatina nebulosa</i>	Squatiniiformes	Scaffold	2,019,620,021	1,578	#N/A	Newly sequenced	#N/A
<i>Squatina squatina</i>	Squatiniiformes	Scaffold	2,991,636,947	2,819	#N/A	Newly sequenced	#N/A
<i>Pristiophorus japonicus</i>	Pristiophoriformes	Scaffold	2,490,874,508	2,034	#N/A	Newly sequenced	#N/A
<i>Latimeria chalumnae</i>	Coelacanthiformes	Scaffold	2,736,328,441	331,424	11,394	GCA_000325985.2	2012/9/21

<i>Protopterus amphibius</i>	Ceratodontiformes	Chromosome	40,054,324,612	2,813,524,174	#N/A	CNP0001614	#N/A
<i>Erpetoichthys calabaricus</i>	Polypteriformes	Chromosome	3,811,038,701	199,226,436	1,143,051	GCA_900747795.2	2019/3/26
<i>Polypterus senegalus</i>	Polypteriformes	Chromosome	3,673,880,163	189,689,183	4,528,139	GCA_016835505.1	2021/2/9
<i>Polyodon spathula</i>	Acipenseriformes	Chromosome	1,542,083,420	48,906,729	3,441,286	GCA_017654505.1	2021/3/31
<i>Acipenser ruthenus</i>	Acipenseriformes	Scaffold	1,830,501,248	42,361,471	601,142	GCA_010645085.1	2020/2/19
<i>Amia calva</i>	Amiiformes	Chromosome	831,019,635	41,163,702	21,103	GCA_017591415.1	2021/3/30
<i>Atractosteus spatula</i>	Lepisosteiformes	Scaffold	1,055,216,122	1,355,703	21,205	GCA_016984175.1	2021/2/26
<i>Lepisosteus oculatus</i>	Lepisosteiformes	Chromosome	945,878,036	6,928,108	68,337	GCA_000242695.1	2012/1/13
<i>Arapaima gigas</i>	Osteoglossiformes	Scaffold	667,351,951	285,171	64,954	GCA_007844225.1	2019/8/8
<i>Campylomormyrus elephas</i>	Osteoglossiformes	Scaffold	904,220,811	1,611,125	#N/A	Newly sequenced	#N/A
<i>Gymnarchus niloticus</i>	Osteoglossiformes	Scaffold	653,962,364	1,179,548	54,271	GCA_009936485.1	2020/1/27
<i>Heterotis niloticus</i>	Osteoglossiformes	Chromosome	670,838,590	32,427,226	4,571,806	GCA_018136845.1	2021/4/23
<i>Mormyrus caschive</i>	Osteoglossiformes	Scaffold	819,549,647	1,531,429	#N/A	Newly sequenced	#N/A
<i>Paramormyrops kingsleyae</i>	Osteoglossiformes	Scaffold	799,421,083	1,731,158	37,656	GCA_002872115.1	2018/1/10
<i>Scleropages formosus</i>	Osteoglossiformes	Chromosome	784,563,014	31,084,684	9,102,216	GCA_900964775.1	2019/4/11
<i>Brevimyrus niger</i>	Osteoglossiformes	Scaffold	749,215,367	12,376,973	23,012	GCA_015164015.1	2020/11/1
<i>Mormyrus lacerda</i>	Osteoglossiformes	Scaffold	796,858,152	13,297,468	23,165	GCA_015163995.1	2020/11/1
<i>Mormyrus iriodes</i>	Osteoglossiformes	Scaffold	756,950,169	38,654	31,067	GCA_011763525.1	2020/3/27
<i>Megalops atlanticus</i>	Elopiformes	Chromosome	990,029,801	39,402,000	17,756,893	GCA_019176425.1	2021/7/9
<i>Megalops cyprinoides</i>	Elopiformes	Chromosome	959,258,777	38,251,172	26,850,664	GCA_013368585.1	2020/6/23
<i>Albula glossodonta</i>	Albuliformes	Scaffold	1,049,070,812	7,972,964	3,575,230	GCA_019788955.1	2021/8/27
<i>Anguilla obscura</i>	Anguilliformes	Scaffold	882,390,062	54,849	12,681	GCA_901111295.1	2020/2/10
<i>Anguilla japonica</i>	Anguilliformes	Scaffold	966,917,315	36,264,158	11,014	GCA_003597225.1	2018/9/14
<i>Anguilla anguilla</i>	Anguilliformes	Scaffold	1,018,701,900	59,657	2,544	GCA_000695075.1	2014/5/20
<i>Anguilla marmorata</i>	Anguilliformes	Scaffold	882,006,954	64,942	16,529	GCA_901111315.1	2020/2/10
<i>Anguilla megastoma</i>	Anguilliformes	Scaffold	877,765,645	61,910	14,572	GCA_901111305.1	2020/2/10
<i>Anguilla rostrata</i>	Anguilliformes	Scaffold	1,413,032,609	86,641	7,355	GCA_001606085.1	2016/4/4
<i>Gymnothorax sp.</i>	Anguilliformes	Scaffold	1,875,072,257	35,693	#N/A	Newly sequenced	#N/A
<i>Ophichthys apicalis</i>	Anguilliformes	Scaffold	1,448,325,526	6,331	#N/A	Newly sequenced	#N/A
<i>Clupea harengus</i>	Clupeiformes	Chromosome	725,686,888	30,022,480	1,151,065	GCA_900700415.1	2019/4/16
<i>Coilia nasus</i>	Clupeiformes	Chromosome	812,159,898	2,096,086	1,632,759	GCA_007927625.1	2019/8/16
<i>Denticeps clupeoides</i>	Clupeiformes	Chromosome	567,401,054	22,793,177	3,059,612	GCA_900700375.1	2019/3/16
<i>Tenualosa ilisha</i>	Clupeiformes	Scaffold	815,647,530	188,026	129,889	GCA_003651195.1	2018/10/15

<i>Alosa alosa</i>	Clupeiformes	Chromosome	854,597,291	35,353,468	1,188,980	GCA_017589495.1	2021/3/30
<i>Alosa sapidissima</i>	Clupeiformes	Chromosome	903,564,947	38,440,066	1,565,278	GCA_018492685.1	2021/5/25
<i>Limnothrissa miodon</i>	Clupeiformes	Scaffold	580,924,283	455,492	31,140	GCA_017657215.1	2021/4/2
<i>Sardina pilchardus</i>	Clupeiformes	Scaffold	949,617,276	96,617	9,398	GCA_900499035.1	2018/9/24
<i>Chanos chanos</i>	Gonorynchiformes	Chromosome	656,929,047	50,300,218	23,134,051	GCA_902362185.1	2019/8/14
<i>Cobitis sinensis</i>	Cypriniformes	Scaffold	1,374,041,473	807,927	#N/A	Newly sequenced	#N/A
<i>Acrossocheilus fasciatus</i>	Cypriniformes	Scaffold	907,023,967	2,232,654	#N/A	Newly sequenced	#N/A
<i>Anabarilius grahami</i>	Cypriniformes	Scaffold	991,887,266	4,459,447	36,058	GCA_003731715.1	2018/11/15
<i>Carassioides cantonensis</i>	Cypriniformes	Scaffold	1,623,183,999	112,635	#N/A	Newly sequenced	#N/A
<i>Carassius auratus</i>	Cypriniformes	Chromosome	1,820,635,050	22,763,433	821,153	GCA_003368295.1	2018/8/9
<i>Ctenopharyngo don idellus</i>	Cypriniformes	Chromosome	900,506,596	6,456,983	40,781	Yaping Wang et al.,	1900/1/0
<i>Culter alburnus</i>	Cypriniformes	Scaffold	1,017,706,352	3,669,388	121,415	GCA_009869775.1	2020/1/14
<i>Cyprinus carpio</i>	Cypriniformes	Chromosome	1,713,658,011	7,828,959	75,080	GCA_000951615.2	2014/11/3
<i>Danio aesculapii</i>	Cypriniformes	Chromosome	1,381,485,597	55,996,547	1,517,104	GCA_903798145.1	2020/6/27
<i>Danio albolineatus</i>	Cypriniformes	Scaffold	1,463,018,096	4,632,361	18,826	GCA_903798035.1	2020/6/16
<i>Danio choprae</i>	Cypriniformes	Scaffold	1,102,162,037	5,707,536	13,929	GCA_903798125.1	2020/6/16
<i>Danio jaintianensis</i>	Cypriniformes	Scaffold	1,122,503,336	18,165,534	22,545	GCA_903798115.1	2020/6/16
<i>Danio kyathit</i>	Cypriniformes	Chromosome	1,708,446,541	64,228,523	1,264,270	GCA_903798195.1	2020/6/27
<i>Danio rerio</i>	Cypriniformes	Chromosome	1,373,454,788	7,379,053	1,422,317	GCA_000002035.4	2017/5/9
<i>Danio tinwini</i>	Cypriniformes	Scaffold	1,497,358,328	13,583,820	16,685	GCA_903798205.1	2020/6/16
<i>Danionella dracula</i>	Cypriniformes	Scaffold	665,208,374	10,287,669	2,300,191	GCA_900490495.1	2018/7/1
<i>Danionella translucida</i>	Cypriniformes	Scaffold	735,303,417	340,819	133,131	GCA_007224835.1	2019/7/22
<i>Garra orientalis</i>	Cypriniformes	Scaffold	1,117,649,848	233,130	#N/A	Newly sequenced	#N/A
<i>Puntigrus tetrazona</i>	Cypriniformes	Chromosome	730,818,536	25,085,836	1,423,374	GCA_018831695.1	2021/6/10
<i>Hemibarbus labeo</i>	Cypriniformes	Scaffold	1,068,447,525	1,006,686	#N/A	Newly sequenced	#N/A
<i>Labeo rohita</i>	Cypriniformes	Scaffold	1,484,730,970	1,959,535	522,833	GCA_004120215.1	2019/1/30
<i>Leuciscus waleckii</i>	Cypriniformes	Scaffold	752,538,629	21,959,719	38,877	GCA_900092035.1	2016/6/28
<i>Macrochirichthys macrochirius</i>	Cypriniformes	Scaffold	553,794,382	9,685,639	#N/A	Newly sequenced	#N/A
<i>Sinocyclocheilus maitianheensis</i>	Cypriniformes	Scaffold	1,680,372,801	1,361,887	28,587	GCA_018148995.1	2021/4/28
<i>Megalobrama amblycephala</i>	Cypriniformes	Scaffold	1,087,899,576	1,403,021	322,873	GCA_009869865.1	2020/1/14
<i>Onychostoma macrolepis</i>	Cypriniformes	Chromosome	886,569,841	34,067,727	10,809,623	GCA_012432095.1	2020/4/17
<i>Oxygymnocypris stewartii</i>	Cypriniformes	Scaffold	1,849,224,471	257,093	257,093	GCA_003573665.1	2018/9/18

<i>Pseudobrama simoni</i>	Cypriniformes	Scaffold	929,130,599	13,799,189	#N/A	Newly sequenced	#N/A
<i>Rhodeus ocellatus</i>	Cypriniformes	Scaffold	902,427,846	4,198,183	#N/A	Newly sequenced	#N/A
<i>Scaphesthes lepturus</i>	Cypriniformes	Scaffold	986,829,913	1,093,417	#N/A	Newly sequenced	#N/A
<i>Sinocyclocheilus anshuiensis</i>	Cypriniformes	Scaffold	1,632,718,266	1,284,143	17,271	GCA_001515605.1	2015/12/14
<i>Sinocyclocheilus grahami</i>	Cypriniformes	Scaffold	1,750,287,761	1,156,368	29,353	GCA_001515645.1	2015/12/16
<i>Sinocyclocheilus rhinoceros</i>	Cypriniformes	Scaffold	1,655,786,410	945,738	18,758	GCA_001515625.1	2015/12/14
<i>Tor tambroides</i>	Cypriniformes	Scaffold	1,235,136,976	44,262	35,829	GCA_021397915.1	2022/1/10
<i>Carassius gibelio</i>	Cypriniformes	Scaffold	1,930,848,899	29,263,521	2,176,597	GCA_019843895.1	2021/9/1
<i>Nemacheilidae sp.</i>	Cypriniformes	Scaffold	826,914,543	774,240	#N/A	Newly sequenced	#N/A
<i>Triplophysa siluroides</i>	Cypriniformes	Scaffold	583,428,323	2,872,994	2,549,348	GCA_006030095.1	2019/6/5
<i>Triplophysa tibetana</i>	Cypriniformes	Chromosome	652,930,317	24,885,746	2,833,080	GCA_008369825.1	2019/9/12
<i>Triplophysa dalaica</i>	Cypriniformes	Chromosome	607,923,607	23,627,495	9,273,388	GCA_015846415.1	2020/12/2
<i>Beaufortia kweichowensis</i>	Cypriniformes	Chromosome	448,565,924	17,950,327	5,533,578	GCA_019155185.1	2021/7/7
<i>Paracanthobrama guichenoti</i>	Cypriniformes	Chromosome	1,087,759,805	38,753,313	1,965,784	GCA_018749465.1	2021/6/8
<i>Gobio gobio</i>	Cypriniformes	Chromosome	1,550,158,453	55,293,470	3,305,918	GCA_021464655.1	2022/1/18
<i>Gobiocypris rarus</i>	Cypriniformes	Chromosome	959,117,570	36,586,440	5,333,633	GCA_018491645.1	2021/5/26
<i>Hypophthalmichthys molitrix</i>	Cypriniformes	Scaffold	1,104,676,189	314,181	2,130	GCA_004764525.1	2019/4/9
<i>Hypophthalmichthys nobilis</i>	Cypriniformes	Scaffold	1,012,063,666	83,012	4,840	GCA_004193235.1	2019/2/12
<i>Ctenopharyngodon idella</i>	Cypriniformes	Chromosome	893,198,495	35,662,356	19,319,444	GCA_019924925.1	2021/9/14
<i>Hemiculter leuciscus</i>	Cypriniformes	Scaffold	1,110,025,803	85,604	#N/A	Newly sequenced	#N/A
<i>Pimephales promelas</i>	Cypriniformes	Scaffold	1,219,326,373	60,380	7,468	GCA_000700825.1	2014/6/10
<i>Rutilus rutilus</i>	Cypriniformes	Scaffold	1,096,109,301	42,584,803	12,790,461	GCA_021464565.1	2022/1/14
<i>Squalius pyrenaicus</i>	Cypriniformes	Contig	48,139,320	1,710	1,710	GCA_001403095.1	2015/7/8
<i>Leuciscus idus</i>	Cypriniformes	Chromosome	1,104,046,756	42,071,081	6,028,738	GCA_021554675.1	2022/1/23
<i>Gyrinocheilus sp.</i>	Cypriniformes	Scaffold	459,729,959	1,614,508	#N/A	Newly sequenced	#N/A
<i>Myxocyprinus asiaticus</i>	Cypriniformes	Chromosome	2,581,058,468	48,473,654	4,188,572	GCA_019703515.1	2021/8/25
<i>Acestrorhynchus altus</i>	Characiformes	Scaffold	825,326,853	10,033,129	#N/A	Newly sequenced	#N/A
<i>Astyanax mexicanus</i>	Characiformes	Chromosome	1,335,239,194	35,377,769	1,767,240	GCA_000372685.2	2017/9/25
<i>Salminus brasiliensis</i>	Characiformes	Scaffold	941,724,108	2,363,637	#N/A	Newly sequenced	#N/A
<i>Metynnus lippincottianus</i>	Characiformes	Scaffold	1,306,190,897	1,106,485	#N/A	Newly sequenced	#N/A
<i>Distichodus sexfasciatus</i>	Characiformes	Scaffold	814,107,174	1,049,286	#N/A	Newly sequenced	#N/A
<i>Hepsetus odoe</i>	Characiformes	Chromosome	862,056,194	25,843,955	476,524	GCA_017165825.1	2021/3/9
<i>Hydrocynus vittatus</i>	Characiformes	Scaffold	1,006,223,410	3,257,534	#N/A	Newly sequenced	#N/A

<i>Phenacogrammus interruptus</i>	Characiformes	Scaffold	690,662,898	3,698	#N/A	Newly sequenced	#N/A
<i>Hydrolycus armatus</i>	Characiformes	Scaffold	805,985,529	3,692,813	#N/A	Newly sequenced	#N/A
<i>Pygocentrus nattereri</i>	Characiformes	Scaffold	1,285,352,492	1,440,044	57,732	GCA_001682695.1	2016/7/12
<i>Colossoma macropomum</i>	Characiformes	Scaffold	2,314,107,020	25,040	18,011	GCA_012026605.1	2020/4/3
<i>Semaprochilodus insignis</i>	Characiformes	Scaffold	1,178,389,447	1,777,306	#N/A	Newly sequenced	#N/A
<i>Hoplias aimara</i>	Characiformes	Scaffold	824,217,961	9,528	#N/A	Newly sequenced	#N/A
<i>Megaleporinus macrocephalus</i>	Characiformes	Chromosome	1,282,030,338	45,034,219	5,013,076	GCA_021613375.1	2022/1/31
<i>Leporinus frederici</i>	Characiformes	Scaffold	1,007,572,843	8,353	#N/A	Newly sequenced	#N/A
<i>Ageneiosus marmoratus</i>	Siluriformes	Scaffold	1,030,000,983	223,139	7,741	GCA_003347165.1	2018/7/31
<i>Ameiurus melas</i>	Siluriformes	Chromosome	868,496,679	32,284,220	7,408,031	GCA_012411365.1	2020/4/16
<i>Auchenoglanis occidentalis</i>	Siluriformes	Scaffold	767,001,076	6,631,916	#N/A	Newly sequenced	#N/A
<i>Bagarius yarrelli</i>	Siluriformes	Scaffold	570,806,968	3,129,371	1,854,961	GCA_005784505.1	2019/5/22
<i>Ictalurus punctatus</i>	Siluriformes	Chromosome	1,002,389,428	26,676,597	2,695,784	GCA_004006655.2	2019/1/25
<i>Malapterurus electricus</i>	Siluriformes	Scaffold	796,449,416	3,578,488	#N/A	Newly sequenced	#N/A
<i>Panaque nigrolineatus</i>	Siluriformes	Scaffold	723,887,262	3,515,394	#N/A	Newly sequenced	#N/A
<i>Pangasianodon hypophthalmus</i>	Siluriformes	Scaffold	715,760,110	14,288,580	62,522	GCA_003671635.1	2018/10/23
<i>Tachysurus fulvidraco</i>	Siluriformes	Scaffold	713,810,725	3,653,474	980,445	GCA_003724035.1	2018/11/13
<i>Clarias batrachus</i>	Siluriformes	Scaffold	821,750,104	361,123	24,893	GCA_003987875.1	2018/12/31
<i>Clarias macrocephalus</i>	Siluriformes	Scaffold	883,399,353	80,802	47,837	GCA_011419295.1	2020/3/18
<i>Clarias magur</i>	Siluriformes	Scaffold	941,297,321	1,316,675	1,226,249	GCA_013621035.1	2020/7/24
<i>Corydoras maculifer</i>	Siluriformes	Scaffold	636,368,989	1,083,835	43,875	GCA_019802435.1	2021/8/30
<i>Hemibagrus wyckioides</i>	Siluriformes	Chromosome	789,799,567	26,010,843	21,700,000	GCA_019097595.1	2021/7/6
<i>Ompok bimaculatus</i>	Siluriformes	Scaffold	718,108,919	81,583	81,474	GCA_009108245.1	2019/10/15
<i>Silurus meridionalis</i>	Siluriformes	Chromosome	742,579,920	28,049,309	13,195,069	GCA_014805685.1	2020/9/30
<i>Brachyplatystoma tigrinum</i>	Siluriformes	Scaffold	740,881,822	11,539	#N/A	Newly sequenced	#N/A
<i>Hexanematichthys leptaspis</i>	Siluriformes	Scaffold	184,588,161	1,374	#N/A	Newly sequenced	#N/A
<i>Synodontis flavitaeniatus</i>	Siluriformes	Scaffold	437,903,111	2,015	#N/A	Newly sequenced	#N/A
<i>Apteronotus albifrons</i>	Gymnotiformes	Scaffold	564,030,374	452,978	#N/A	Newly sequenced	#N/A
<i>Electrophorus electricus</i>	Gymnotiformes	Scaffold	551,880,868	613,956	37,141	GCA_003665695.2	2018/10/30
<i>Brachyhypopomus occidentalis</i>	Gymnotiformes	Scaffold	540,332,535	5,414,873	5,180,106	GCA_020368025.1	2021/10/7
<i>Coregonus sp.</i>	Salmoniformes	Scaffold	2,068,074,305	52,020,451	#N/A	De-Kayne et al., 2020	#N/A
<i>Hucho hucho</i>	Salmoniformes	Scaffold	2,487,549,814	287,338	37,639	GCA_003317085.1	2018/7/13
<i>Oncorhynchus keta</i>	Salmoniformes	Chromosome	1,853,087,674	656,204	50,313	GCA_012931545.1	2020/4/29

<i>Oncorhynchus kisutch</i>	Salmoniformes	Chromosome	2,369,932,239	1,266,128	58,118	GCA_002021735.1	2017/3/6
<i>Oncorhynchus mykiss</i>	Salmoniformes	Chromosome	2,178,999,613	1,670,138	13,827	GCA_002163495.1	2017/6/2
<i>Oncorhynchus nerka</i>	Salmoniformes	Chromosome	1,927,141,915	1,058,586	329,583	GCA_006149115.1	2019/6/7
<i>Oncorhynchus tshawytscha</i>	Salmoniformes	Chromosome	2,425,713,975	1,728,323	133,169	GCA_002872995.1	2018/1/16
<i>Salmo salar</i>	Salmoniformes	Chromosome	2,966,890,203	1,366,254	57,618	GCA_000233375.4	2015/6/10
<i>Salmo trutta</i>	Salmoniformes	Chromosome	2,371,880,186	52,209,666	1,703,178	GCA_901001165.1	2019/6/2
<i>Salvelinus alpinus</i>	Salmoniformes	Chromosome	2,169,553,147	1,018,695	55,619	GCA_002910315.2	2018/2/12
<i>Thymallus thymallus</i>	Salmoniformes	Chromosome	1,564,834,359	32,985,317	31,774	GCA_004348285.1	2019/3/19
<i>Coregonus clupeaformis</i>	Salmoniformes	Chromosome	2,682,612,311	57,494,107	5,728,271	GCA_018398675.1	2021/5/25
<i>Oncorhynchus gorbuscha</i>	Salmoniformes	Chromosome	2,533,396,071	1,769,953	1,753,068	GCA_017355495.1	2021/3/24
<i>Salvelinus namaycush</i>	Salmoniformes	Chromosome	2,345,513,378	44,976,251	1,804,090	GCA_016432855.1	2021/1/6
<i>Esox lucius</i>	Esociformes	Chromosome	940,906,975	37,550,661	3,396,779	GCA_004634155.1	2019/4/17
<i>Esox niger</i>	Esociformes	Scaffold	859,849,040	35,161	30,390	GCA_016801105.1	2021/2/2
<i>Argyrolepeus hemigymnus</i>	Stomiiformes	Scaffold	730,531,183	6,543	#N/A	Newly sequenced	#N/A
<i>Borostomias antarcticus</i>	Stomiiformes	Scaffold	430,362,827	5,368	3,928	GCA_900323325.1	2018/3/26
<i>Mallotus villosus</i>	Osmeriformes	Scaffold	490,877,849	236,994	229,616	GCA_903064625.1	2020/5/18
<i>Protosalanx chinensis</i>	Osmeriformes	Scaffold	466,693,640	5,188,763	103,007	GCA_010882115.1	2020/2/20
<i>Hypomesus transpacificus</i>	Osmeriformes	Chromosome	437,273,953	14,850,352	412,669	GCA_021917145.1	2022/2/4
<i>Plecoglossus altivelis</i>	Osmeriformes	Scaffold	449,104,997	4,315,844	113,526	GCA_019972215.1	2021/7/20
<i>Aplocheilichthys taeniatus</i>	Galaxiiformes	Chromosome	461,946,147	20,392,628	956,540	GCA_017639675.1	2021/3/30
<i>Harpadon nehereus</i>	Aulopiformes	Scaffold	1,203,945,501	13,596	#N/A	Newly sequenced	#N/A
<i>Parasudis fraserbrunneri</i>	Aulopiformes	Scaffold	707,987,062	6,391	4,177	GCA_900302295.1	2018/3/21
<i>Guentherus altivela</i>	Ateleopodiformes	Scaffold	539,598,795	3,201	2,928	GCA_900312595.1	2018/3/19
<i>Benthoosema glaciale</i>	Myctophiformes	Scaffold	676,314,385	6,111	4,393	GCA_900323375.1	2018/3/26
<i>Polymixia japonica</i>	Polymixiiformes	Scaffold	554,895,936	9,571	5,803	GCA_900302305.1	2018/3/19
<i>Percopsis transmontana</i>	Percopsiformes	Scaffold	458,089,168	15,180	8,161	GCA_900302285.1	2018/3/20
<i>Typhlichthys subterraneus</i>	Percopsiformes	Scaffold	555,559,596	9,654	7,314	GCA_900302405.1	2018/3/19
<i>Cyttopsis rosea</i>	Zeiformes	Scaffold	546,506,150	7,082	4,843	GCA_900302355.1	2018/3/19
<i>Zeus faber</i>	Zeiformes	Scaffold	610,433,400	6,332	4,642	GCA_900323335.1	2018/3/26
<i>Stylephorus chordatus</i>	Stylephoriformes	Scaffold	488,488,587	4,684	3,373	GCA_900312615.1	2018/3/20
<i>Coryphaenoides rupestris</i>	Gadiformes	Scaffold	829,208,733	159,738	20,848	GCA_002895965.1	2018/1/23
<i>Gadus morhua</i>	Gadiformes	Chromosome	669,966,409	28,732,775	1,015,663	GCA_902167405.1	2019/7/16
<i>Melanogrammus aeglefinus</i>	Gadiformes	Scaffold	652,790,733	209,126	77,605	GCA_900291075.1	2018/3/19

<i>Phycis blennoides</i>	Gadiformes	Scaffold	416,766,999	10,640	4,532	GCA_900302315.1	2018/3/18
<i>Trachyrincus murrayi</i>	Gadiformes	Scaffold	452,416,606	19,998	6,231	GCA_900323305.1	2018/3/26
<i>Lampris guttatus</i>	Lampriformes	Scaffold	849,277,706	5,222	4,051	GCA_900302545.1	2018/3/30
<i>Regalecus glesne</i>	Lampriformes	Scaffold	656,003,707	9,773	6,781	GCA_900302585.1	2018/3/20
<i>Gephyroberyx darwinii</i>	Trachichthyiformes	Scaffold	535,045,983	18,145	8,514	GCA_900660455.1	2019/6/18
<i>Hoplostethus atlanticus</i>	Trachichthyiformes	Scaffold	520,173,038	15,562	7,812	GCA_900660355.1	2019/6/18
<i>Monocentris japonica</i>	Trachichthyiformes	Scaffold	556,023,515	18,672	8,046	GCA_900323365.1	2018/3/26
<i>Acanthochaenus luetkenii</i>	Beryciformes	Scaffold	545,759,480	8,444	5,636	GCA_900312575.1	2018/3/19
<i>Rondeletia loricata</i>	Beryciformes	Scaffold	568,597,941	7,469	5,112	GCA_900302605.1	2018/3/19
<i>Myripristis murdjan</i>	Holocentriformes	Chromosome	835,254,674	34,950,760	14,475,636	GCA_902150065.1	2019/7/3
<i>Sargocentron spiniferum</i>	Holocentriformes	Scaffold	791,903,703	2,676,641	#N/A	Newly sequenced	#N/A
<i>Neoniphon sammara</i>	Holocentriformes	Scaffold	659,225,728	21,768	#N/A	GCA_900302535.1	#N/A
<i>Holocentrus rufus</i>	Holocentriformes	Scaffold	649,757,301	21,389	9,243	GCA_900302615.1	2018/3/19
<i>Myripristis jacobus</i>	Holocentriformes	Scaffold	720,396,841	21,306	9,816	GCA_900302555.1	2018/3/21
<i>Brotula barbata</i>	Ophidiiformes	Scaffold	485,061,200	45,752	17,578	GCA_900303265.1	2018/3/20
<i>Carapus acus</i>	Ophidiiformes	Scaffold	387,834,307	16,922	9,554	GCA_900312935.1	2018/3/18
<i>Lamprogrammus exutus</i>	Ophidiiformes	Scaffold	492,850,272	5,470	4,213	GCA_900312555.1	2018/3/20
<i>Lucifuga dentata</i>	Ophidiiformes	Scaffold	630,651,555	120,915	62,305	GCA_014773175.1	2020/9/28
<i>Thalassophryne amazonica</i>	Batrachoidiformes	Chromosome	2,446,592,878	118,421,481	2,329,598	GCA_902500255.1	2019/9/30
<i>Apogonidae sp.</i>	Kurtiformes	Scaffold	1,043,886,260	2,084,352	#N/A	Newly sequenced	#N/A
<i>Sphaeramia orbicularis</i>	Kurtiformes	Chromosome	1,342,662,642	57,165,184	2,360,121	GCA_902148855.1	2019/7/3
<i>Siphamia tubifer</i>	Kurtiformes	Chromosome	1,200,845,361	51,162,488	2,340,513	GCA_020466265.1	2021/10/15
<i>Bostrychus sinensis</i>	Gobiiformes	Scaffold	865,256,608	37,614,318	101,659	GCA_018703735.1	2021/6/3
<i>Eleotris oxycephala</i>	Gobiiformes	Scaffold	502,138,739	13,146,699	#N/A	Newly sequenced	#N/A
<i>Oxyeleotris marmorata</i>	Gobiiformes	Scaffold	502,138,739	13,146,699	#N/A	Newly sequenced	#N/A
<i>Periophthalmus magnuspinnatus</i>	Gobiiformes	Scaffold	701,696,780	296,161	28,254	GCA_000787105.1	2014/12/2
<i>Periophthalmus modestus</i>	Gobiiformes	Scaffold	854,419,050	32,909,307	576,039	GCA_019594935.1	2021/8/13
<i>Periophthalmodon schlosseri</i>	Gobiiformes	Scaffold	679,761,122	39,308	16,946	GCA_000787095.1	2014/12/2
<i>Boleophthalmus pectinirostris</i>	Gobiiformes	Scaffold	955,752,150	2,375,582	20,437	GCA_000788275.1	2014/12/2
<i>Rhinogobius giurinus</i>	Gobiiformes	Chromosome	890,097,861	41,510,926	1,569,660	GCA_019453435.1	2021/8/10
<i>Chaenogobius annularis</i>	Gobiiformes	Scaffold	747,409,675	951,958	803,221	GCA_015082035.1	2020/10/21
<i>Lythrypnus dalli</i>	Gobiiformes	Scaffold	830,691,505	2,577,483	8,114	GCA_011763505.1	2020/3/27
<i>Mugilogobius chulae</i>	Gobiiformes	Chromosome	1,003,362,100	42,457,289	701,221	GCA_016735935.1	2021/1/27

<i>Neogobius melanostomus</i>	Gobiiformes	Contig	1,003,738,541	2,817,412	2,817,412	GCA_007210695.1	2019/7/22
<i>Proterorhinus semilunaris</i>	Gobiiformes	Chromosome	864,621,867	39,007,679	7,924,916	GCA_021464625.1	2022/1/18
<i>Rhinogobius similis</i>	Gobiiformes	Chromosome	890,097,861	41,510,926	1,569,660	GCA_019453435.1	2021/8/10
<i>Lesueurigobius sanzi</i>	Gobiiformes	Scaffold	810,626,388	11,480	6,729	GCA_900303255.1	2018/3/21
<i>Gobiidae sp.</i>	Gobiiformes	Scaffold	1,338,042,597	133,471	#N/A	Newly sequenced	#N/A
<i>Gempylus serpens</i>	Scombriformes	Scaffold	791,026,746	3,391,346	#N/A	Newly sequenced	#N/A
<i>Euthynnus affinis</i>	Scombriformes	Scaffold	758,243,246	29,176,745	84,203	GCA_019973915.1	2021/9/9
<i>Scomber colias</i>	Scombriformes	Scaffold	814,072,382	4,190,967	3,315,783	GCA_021039105.1	2021/12/3
<i>Thunnus albacares</i>	Scombriformes	Scaffold	728,212,003	46,920	16,808	GCA_900302625.1	2018/3/21
<i>Thunnus maccoyii</i>	Scombriformes	Chromosome	782,423,979	33,761,101	26,803,536	GCA_910596095.1	2021/7/7
<i>Thunnus orientalis</i>	Scombriformes	Contig	684,497,465	8,235	8,235	GCA_000418415.1	2013/6/27
<i>Synchiropus splendidus</i>	Callionymiformes	Scaffold	482,926,376	12,187,484	66,195	GCA_019776025.1	2021/8/26
<i>Callionymus lyra</i>	Callionymiformes	Scaffold	568,707,486	26,698,546	2,201,294	GCA_016630915.1	2021/1/13
<i>Dactylopterus volitans</i>	Syngnathiformes	Scaffold	576,982,700	17,099	8,271	GCA_901007715.1	2020/4/15
<i>Hippocampus comes</i>	Syngnathiformes	Scaffold	493,775,940	2,034,572	39,546	GCA_001891065.1	2016/12/8
<i>Hippocampus erectus</i>	Syngnathiformes	Scaffold	457,759,912	1,971,390	-	Lin et al., 2017	1900/1/0
<i>Syngnathus acus</i>	Syngnathiformes	Chromosome	324,331,233	14,974,571	11,959,915	GCA_901709675.1	2019/5/29
<i>Syngnathus floridae</i>	Syngnathiformes	Scaffold	367,727,558	1,379,020	11,028	GCA_010014945.1	2020/1/30
<i>Syngnathus typhle</i>	Syngnathiformes	Scaffold	314,580,961	3,046,963	25,809	GCA_901007915.1	2020/4/15
<i>Parupeneus multifasciatus</i>	Syngnathiformes	Scaffold	727,107,922	61,794	#N/A	Newly sequenced	#N/A
<i>Aeoliscus strigatus</i>	Syngnathiformes	Scaffold	381,315,519	115,785	15,935	GCA_901007665.1	2020/4/15
<i>Dunckerocampus dactyliophorus</i>	Syngnathiformes	Scaffold	618,747,239	75,189	27,594	GCA_901007775.1	2020/4/15
<i>Fistularia tabacaria</i>	Syngnathiformes	Scaffold	593,175,436	107,234	17,669	GCA_901007705.1	2020/4/15
<i>Hippocampus abdominalis</i>	Syngnathiformes	Chromosome	444,667,964	21,766,359	8,410,403	GCA_018466805.1	2021/5/25
<i>Hippocampus whitei</i>	Syngnathiformes	Scaffold	432,755,118	40,759	10,333	GCA_901007805.1	2020/4/15
<i>Macroramphosus scolopax</i>	Syngnathiformes	Scaffold	417,659,696	41,786	13,365	GCA_901007825.1	2020/4/15
<i>Phyllopteryx taeniolatus</i>	Syngnathiformes	Chromosome	668,103,038	30,975,377	12,818,866	GCA_019802545.1	2021/9/9
<i>Syngnathoides biaculeatus</i>	Syngnathiformes	Chromosome	636,519,225	31,867,959	18,018,000	GCA_019802595.1	2021/9/9
<i>Syngnathus rostellatus</i>	Syngnathiformes	Scaffold	283,186,098	87,573	14,947	GCA_901007895.1	2020/4/15
<i>Hippocampus kuda</i>	Syngnathiformes	Scaffold	444,947,940	31,191	10,411	GCA_901007745.1	2020/4/15
<i>Corythoichthys haematopterus</i>	Syngnathiformes	Scaffold	957,794,659	8,523	#N/A	Newly sequenced	#N/A
<i>Mastacembelus armatus</i>	Synbranchiformes	Chromosome	591,935,101	25,090,313	8,014,513	GCA_900324485.2	2019/3/16

<i>Monopterus albus</i>	Synbranchiformes	Scaffold	684,144,148	2,106,322	22,239	GCA_001952655.1	2017/1/17
<i>Anabas testudineus</i>	Anabantiformes	Chromosome	555,658,001	25,063,394	7,055,436	GCA_900324465.2	2018/12/19
<i>Betta splendens</i>	Anabantiformes	Chromosome	441,388,503	20,129,463	2,497,747	GCA_900634795.2	2019/3/20
<i>Channa argus</i>	Anabantiformes	Chromosome	644,129,659	27,759,500	11,786,852	GCA_004786185.1	2019/4/15
<i>Helostoma temminckii</i>	Anabantiformes	Scaffold	599,808,025	71,786	17,055	GCA_900302695.1	2018/3/19
<i>Channa maculata</i>	Anabantiformes	Chromosome	618,823,650	28,367,461	21,727,292	GCA_020496755.1	2021/10/14
<i>Brachirus orientalis</i>	Pleuronectiformes	Scaffold	629,956,282	5,477,449	352,130	GCA_016785285.1	2021/2/1
<i>Cynoglossus semilaevis</i>	Pleuronectiformes	Chromosome	470,199,494	509,861	27,008	GCA_000523025.1	2014/1/28
<i>Hippoglossus hippoglossus</i>	Pleuronectiformes	Chromosome	596,792,676	26,312,018	7,020,318	GCA_009819705.1	2019/12/31
<i>Hippoglossus stenolepis</i>	Pleuronectiformes	Chromosome	594,145,200	24,985,257	17,826,919	GCA_013339905.1	2020/6/16
<i>Paralichthys olivaceus</i>	Pleuronectiformes	Chromosome	545,775,252	3,817,360	30,544	GCA_001904815.2	2017/3/15
<i>Reinhardtius hippoglossoides</i>	Pleuronectiformes	Scaffold	677,540,803	17,640,195	775,256	GCA_006182925.2	2019/7/25
<i>Scophthalmus maximus</i>	Pleuronectiformes	Chromosome	524,979,463	24,811,384	54,836	GCA_003186165.1	2018/6/5
<i>Verasper variegatus</i>	Pleuronectiformes	Chromosome	556,789,779	23,832,094	175,150	GCA_012393435.1	2020/4/16
<i>Chascanopsetta lugubris</i>	Pleuronectiformes	Scaffold	577,766,669	66,028	14,706	GCA_016785325.1	2021/2/1
<i>Colistium nudipinnis</i>	Pleuronectiformes	Scaffold	427,535,253	739,841	141,808	GCA_016785305.1	2021/2/1
<i>Paraplagusia blochii</i>	Pleuronectiformes	Scaffold	555,100,837	178,074	37,510	GCA_016785265.1	2021/2/1
<i>Platichthys stellatus</i>	Pleuronectiformes	Chromosome	609,990,026	25,101,570	33,203	GCA_016801935.1	2021/2/4
<i>Pseudopleuronectes yokohamae</i>	Pleuronectiformes	Contig	547,831,023	1,994	1,994	GCA_000787555.1	2014/11/24
<i>Pseudorhombus dupliciocellatus</i>	Pleuronectiformes	Scaffold	395,092,285	2,687,910	538,669	GCA_016785235.1	2021/2/1
<i>Solea senegalensis</i>	Pleuronectiformes	Chromosome	603,535,267	26,091,242	1,188,264	GCA_019176455.1	2021/7/9
<i>Trinectes maculatus</i>	Pleuronectiformes	Scaffold	450,194,497	373,139	103,915	GCA_016785345.1	2021/2/1
<i>Eleutheronema tetradactylum</i>	#N/A	Scaffold	610,497,648	56,314	30,813	GCA_014490765.1	2020/9/8
<i>Polydactylus sextarius</i>	#N/A	Chromosome	564,623,279	22,256,062	18,418,211	GCA_016801845.1	2021/2/4
<i>Lates calcarifer</i>	#N/A	Scaffold	668,481,366	1,191,366	1,066,117	GCA_001640805.1	2016/5/9
<i>Echeneis naucrates</i>	Carangiformes	Chromosome	544,229,245	23,287,306	12,371,513	GCA_900963305.1	2019/4/11
<i>Caranx melampygus</i>	Carangiformes	Scaffold	581,342,446	2,168,377	1,469,537	Newly sequenced	2021/6/29
<i>Seriola dumerili</i>	Carangiformes	Scaffold	677,686,174	5,812,906	249,509	GCA_002260705.1	2017/8/3
<i>Seriola lalandi</i>	Carangiformes	Scaffold	766,364,468	411,616	37,711	GCA_003054885.1	2018/4/17
<i>Seriola quinqueradiata</i>	Carangiformes	Scaffold	639,269,536	5,610,255	872,227	GCA_002217815.1	2017/7/5
<i>Seriola rivoliana</i>	Carangiformes	Scaffold	666,141,578	9,509,606	740,108	GCA_002994505.1	2018/3/12
<i>Trachinotus ovatus</i>	Carangiformes	Scaffold	648,062,395	29,494,812	1,846,793	GCA_900607315.1	2018/12/2
<i>Caranx ignobilis</i>	Carangiformes	Scaffold	605,128,561	46,304,415	7,719,230	GCA_021432705.1	2022/1/10

<i>Trachurus trachurus</i>	Carangiformes	Scaffold			1,547,897	GCA_905171655.1	2021/1/28
<i>Selene dorsalis</i>	Carangiformes	Scaffold	818,939,194	1,610,964	11,209	GCA_900303245.1	2018/3/19
			528,779,420	32,464			
<i>Seriola dorsalis</i>	Carangiformes	Scaffold			#N/A	GCF_002814215.1	#N/A
			732,509,836	1,269,737			
<i>Istiophorus platypterus</i>	Istiophoriformes	Chromosome			11,851,605	GCA_016859345.1	2021/2/12
<i>Xiphias gladius</i>	Istiophoriformes	Chromosome	614,147,335	26,283,175	5,252,707	GCA_016859285.1	2021/2/12
			691,780,295	29,220,096			
<i>Toxotes chatareus</i>	#N/A	Chromosome			8,430,381	GCA_016801885.1	2021/2/4
<i>Toxotes jaculatrix</i>	#N/A	Chromosome	618,729,764	23,928,265	15,555,772	GCA_017976425.1	2021/4/15
<i>Gouania willdenowi</i>	Gobiesociformes	Chromosome	621,503,431	26,859,436	1,838,341	GCA_900634775.1	2019/2/6
<i>Omobranchus fasciolatoceps</i>			937,150,793	38,978,045			
<i>Salarias fasciatus</i>	Blenniiformes	Scaffold			#N/A	Newly sequenced	#N/A
<i>Parablennius parvicornis</i>	Blenniiformes	Chromosome	667,507,060	1,478,370	2,597,836	GCA_902148845.1	2019/7/3
<i>Planiliza haematocheila</i>	Blenniiformes	Scaffold	797,507,141	32,729,575	7,343	GCA_900302745.1	2018/3/19
<i>Parambassis ranga</i>	Mugiliformes	Contig	599,249,148	16,796	3,973,280	GCA_005024645.1	2019/5/1
<i>Amphiprion clarkii</i>			747,342,729	3,973,280			
<i>Stegastes partitus</i>	#N/A	Chromosome			5,080,925	GCA_900634625.1	2019/2/6
<i>Abudefduf vaigiensis</i>	#N/A	Scaffold	551,012,959	22,993,012	#N/A	Newly sequenced	#N/A
<i>Acanthochromis polyacanthus</i>	#N/A	Scaffold	856,815,181	575,009	43,010	GCA_000690725.1	2014/5/13
<i>Amphiprion ocellaris</i>	#N/A	Scaffold	800,491,834	411,659	#N/A	Newly sequenced	#N/A
<i>Amphiprion percula</i>	#N/A	Scaffold	825,522,809	6,382,270	16,099	GCA_002109545.1	2017/4/25
<i>Amphiprion sebae</i>	#N/A	Scaffold	991,584,656	334,400	324,210	GCA_002776465.1	2017/11/16
<i>Chromis chromis</i>	#N/A	Chromosome	880,720,895	401,715	3,123,421	GCA_003047355.2	2018/11/28
<i>Amphiprion akallopisos</i>	#N/A	contig	908,955,932	38,416,550	#N/A	Marcionetti et al., 2019	#N/A
<i>Amphiprion bicinctus</i>	#N/A	Scaffold	799,224,790	245,870	8,509	GCA_900302755.1	2018/3/21
<i>Amphiprion melanopus</i>	#N/A	Contig	834,429,411	14,223	#N/A	Marcionetti et al., 2019	#N/A
<i>Amphiprion nigripes</i>	#N/A	Contig	800,983,251	246,052	#N/A	Marcionetti et al., 2019	#N/A
<i>Amphiprion perideraion</i>	#N/A	Contig	799,222,757	246,127	#N/A	Marcionetti et al., 2019	#N/A
<i>Amphiprion polymnus</i>	#N/A	Contig	802,629,797	245,703	#N/A	Marcionetti et al., 2019	#N/A
<i>Pomacentrus moluccensis</i>	#N/A	Contig	799,696,425	246,124	#N/A	Marcionetti et al., 2019	#N/A
<i>Premnas biaculeatus</i>	#N/A	Contig	800,833,675	246,037	#N/A	Marcionetti et al., 2019	#N/A
<i>Pseudochromis fuscus</i>	#N/A	Contig	799,914,227	246,119	#N/A	Marcionetti et al., 2019	#N/A
<i>Amphilophus citrinellus</i>	#N/A	Contig	794,059,034	246,470	#N/A	Marcionetti et al., 2019	#N/A
<i>Archocentrus centrarchus</i>	#N/A	Contig	796,587,812	247,121	#N/A	Marcionetti et al., 2019	#N/A
<i>Astatotilapia burtoni</i>	#N/A	Scaffold			12,029	GCA_900323345.1	2018/3/26
			657,041,210	24,689	23,456	GCA_000751415.1	2014/9/7
			844,902,565	1,216,136	2,146,538	GCA_007364275.2	2019/8/1
			932,930,362	35,590,001	21,886	GCA_000239415.1	2011/12/22
			831,411,547	1,194,190			

<i>Astatotilapia calliptera</i>	Cichliformes	Chromosome	880,445,564	38,669,361	4,438,245	GCA_900246225.3	2018/5/30
<i>Crenicichla johanna</i>	Cichliformes	Scaffold	825,230,989	2,353,102	#N/A	Newly sequenced	#N/A
<i>Haplochromis nyererei</i>	Cichliformes	Scaffold	830,133,247	2,525,540	22,622	GCA_000239375.1	2011/12/22
<i>Maylandia zebra</i>	Cichliformes	Chromosome	957,485,262	32,660,920	1,407,748	GCA_000238955.5	2018/4/10
<i>Neolamprologus brichardi</i>	Cichliformes	Scaffold	847,910,432	4,430,025	13,047	GCA_000239395.1	2011/12/22
<i>Oreochromis aureus</i>	Cichliformes	Scaffold	918,937,175	1,102,239	60,340	GCA_005870065.1	2019/5/28
<i>Oreochromis niloticus</i>	Cichliformes	Chromosome	1,005,681,550	38,839,487	2,923,640	GCA_001858045.3	2018/6/29
<i>Petrochromis trewavasae</i>	Cichliformes	Scaffold	917,573,940	1,826,695	46,763	GCA_902810495.1	2020/4/15
<i>Simochromis diagramma</i>	Cichliformes	Scaffold	848,827,444	8,960,300	2,231,376	GCA_900408965.1	2018/6/18
<i>Tropheus moorii</i>	Cichliformes	Scaffold	911,126,605	1,635,562	32,181	GCA_902810505.1	2020/4/15
<i>Haplochromis chilotes</i>	Cichliformes	Scaffold	798,403,686	1,540,223	20,893	GCA_018403475.1	2021/4/6
<i>Haplochromis sauvagei</i>	Cichliformes	Scaffold	805,051,152	1,207,037	21,828	GCA_018403495.1	2021/4/6
<i>Oreochromis spilurus</i>	Cichliformes	Contig	764,974,731	11,851	11,851	GCA_008269305.1	2019/9/3
<i>Pundamilia nyererei</i>	Cichliformes	Scaffold	830,133,247	2,525,540	22,622	GCA_000239375.1	2011/12/22
<i>Metriaclimma zebra</i>	Cichliformes	Chromosome	957,200,631	32,647,892	#N/A	GCF_000238955.4	#N/A
<i>Glossolepis incisus</i>	Atheriniformes	Scaffold	888,160,701	1,420,359	#N/A	Newly sequenced	#N/A
<i>Menidia beryllina</i>	Atheriniformes	Scaffold	564,723,020	4,493,486	40,093	GCA_013363375.1	2020/6/19
<i>Neostethus bicornis</i>	Atheriniformes	Chromosome	778,440,790	29,119,677	23,089,617	GCA_902685375.1	2019/11/26
<i>Melanotaenia boesemani</i>	Atheriniformes	Chromosome	865,592,505	37,878,235	9,299,978	GCA_017639745.1	2021/3/30
<i>Menidia menidia</i>	Atheriniformes	Scaffold	508,735,381	1,562,819	7,957	GCA_905332855.1	2021/3/21
<i>Odontesthes bonariensis</i>	Atheriniformes	Scaffold	870,329,685	60,945	14,340	GCA_014825785.1	2020/9/30
<i>Oryzias javanicus</i>	Beloniformes	Chromosome	809,679,899	35,390,520	3,558,013	GCA_003999625.1	2019/1/8
<i>Oryzias latipes</i>	Beloniformes	Chromosome	734,057,086	31,218,526	2,530,934	GCA_002234675.1	2017/7/27
<i>Oryzias melastigma</i>	Beloniformes	Scaffold	779,469,774	23,737,187	30,057	GCA_002922805.1	2018/2/7
<i>Hirundichthys speculiger</i>	Beloniformes	Scaffold	1,043,046,751	1,152,470	1,092,820	GCA_018398805.1	2021/5/19
<i>Oryzias celebensis</i>	Beloniformes	Scaffold	688,104,639	28,928,686	6,114,092	GCA_014656515.1	2020/9/12
<i>Oryzias sinensis</i>	Beloniformes	Scaffold	813,986,518	991,358	28,180	GCA_008586565.1	2019/9/19
<i>Xenotodon cancila</i>	Beloniformes	Chromosome	736,214,936	28,679,447	2,195,850	GCA_014839995.1	2020/9/30
<i>Aphyosemion australe</i>	Cyprinodontiformes	Scaffold	868,348,903	1,435,212	119,465	GCA_006937985.1	2019/7/12
<i>Aplocheilichthys lineatus</i>	Cyprinodontiformes	Scaffold	495,092,281	106,765	#N/A	Newly sequenced	#N/A
<i>Austrofundulus limnaeus</i>	Cyprinodontiformes	Scaffold	866,963,281	1,098,383	8,097	GCA_001266775.1	2015/7/28
<i>Callopanchax toddi</i>	Cyprinodontiformes	Scaffold	853,385,936	1,656,350	52,012	GCA_006937965.1	2019/7/12
<i>Cyprinodon variegatus</i>	Cyprinodontiformes	Scaffold	1,035,184,475	835,301	20,803	GCA_000732505.1	2015/10/1

<i>Fundulus heteroclitus</i>	Cyprinodontiformes	Scaffold	1,021,898,560	1,252,252	16,688	GCA_000826765.1	2015/1/21
<i>Gambusia affinis</i>	Cyprinodontiformes	Scaffold	598,663,367	6,651,460	17,511	GCA_003097735.1	2018/5/8
<i>Kryptolebias hermaphroditus</i>	Cyprinodontiformes	Chromosome	683,986,837	27,459,464	46,411	GCA_007896545.1	2019/8/8
<i>Nothobranchius furzeri</i>	Cyprinodontiformes	Chromosome	1,242,518,059	15,858,201	19,950	GCA_001465895.2	2015/12/11
<i>Nothobranchius kuhntae</i>	Cyprinodontiformes	Scaffold	1,122,656,415	1,178,460	22,234	GCA_006942095.1	2019/7/12
<i>Orestias ascotanensis</i>	Cyprinodontiformes	Scaffold	696,366,514	2,226,781	43,756	GCA_012931525.1	2020/4/28
<i>Pachypanchax playfairii</i>	Cyprinodontiformes	Scaffold	669,774,067	3,173,794	76,641	GCA_006937955.1	2019/7/12
<i>Poecilia formosa</i>	Cyprinodontiformes	Scaffold	748,923,461	1,574,226	57,472	GCA_000485575.1	2013/10/28
<i>Poecilia latipinna</i>	Cyprinodontiformes	Scaffold	815,144,743	279,200	33,278	GCA_001443285.1	2015/11/13
<i>Poecilia mexicana</i>	Cyprinodontiformes	Scaffold	801,711,499	275,316	39,840	GCA_001443325.1	2015/11/13
<i>Poecilia reticulata</i>	Cyprinodontiformes	Chromosome	731,622,281	5,270,359	41,908	GCA_000633615.2	2014/4/28
<i>Poeciliopsis occidentalis</i>	Cyprinodontiformes	Scaffold	724,239,826	1,540,873	56,887	GCA_010883035.1	2020/2/20
<i>Poeciliopsis retropinna</i>	Cyprinodontiformes	Scaffold	621,858,691	21,601,530	15,063,430	GCA_010277075.1	2020/2/6
<i>Poeciliopsis turrubarensis</i>	Cyprinodontiformes	Scaffold	596,996,646	4,228,945	101,021	GCA_010277015.1	2020/2/6
<i>Xiphophorus couchianus</i>	Cyprinodontiformes	Chromosome	688,541,509	30,550,352	15,315,838	GCA_001444195.3	2018/9/19
<i>Xiphophorus hellerii</i>	Cyprinodontiformes	Chromosome	733,126,988	26,465,357	7,119,835	GCA_003331165.1	2018/7/23
<i>Xiphophorus maculatus</i>	Cyprinodontiformes	Chromosome	704,321,165	31,535,491	9,181,372	GCA_002775205.2	2017/12/7
<i>Alfaro cultratus</i>	Cyprinodontiformes	Scaffold	639,540,724	57,399	56,524	GCA_903066825.1	2021/9/22
<i>Anableps anableps</i>	Cyprinodontiformes	Chromosome	867,632,993	39,321,252	17,460,000	GCA_014839685.1	2020/9/30
<i>Brachyrhaphis roseni</i>	Cyprinodontiformes	Scaffold	646,914,198	41,687	41,649	GCA_903066815.1	2021/9/22
<i>Cyprinodon brontotheroides</i>	Cyprinodontiformes	Scaffold	1,162,749,926	30,093,578	71,034	GCA_018398635.1	2021/5/19
<i>Cyprinodon nevadensis</i>	Cyprinodontiformes	Scaffold	1,011,849,000	83,166	1,416	GCA_000776015.1	2014/11/13
<i>Cyprinodon tularosa</i>	Cyprinodontiformes	Contig	2,173,780,990	1,362,150	1,362,150	GCA_016077235.1	2020/12/15
<i>Girardinichthys multiradiatus</i>	Cyprinodontiformes	Chromosome	1,109,672,912	49,294,359	15,927,784	GCA_021462225.1	2022/1/13
<i>Girardinus metallicus</i>	Cyprinodontiformes	Scaffold	674,367,140	41,813	41,812	GCA_903067095.1	2021/9/22
<i>Nematolebias whitei</i>	Cyprinodontiformes	Scaffold	1,218,316,901	49,984,095	19,984	GCA_014905685.1	2020/10/22
<i>Poeciliopsis infans</i>	Cyprinodontiformes	Scaffold	582,637,719	59,589	59,269	GCA_903067105.1	2021/9/22
<i>Poeciliopsis paucimaculata</i>	Cyprinodontiformes	Scaffold	643,974,302	65,100	65,094	GCA_903067055.1	2021/9/22
<i>Fundulus catenatus</i>	Cyprinodontiformes	Scaffold	1,165,204,350	435,700	#N/A	Johnson et al., 2020	#N/A
<i>Fundulus notti</i>	Cyprinodontiformes	Scaffold	1,101,220,061	2,749,221	#N/A	Johnson et al., 2020	#N/A
<i>Fundulus olivaceus</i>	Cyprinodontiformes	Scaffold	1,221,672,488	2,716,042	#N/A	Johnson et al., 2020	#N/A
<i>Fundulus xenicus</i>	Cyprinodontiformes	Scaffold	1,095,621,269	904,156	#N/A	Johnson et al., 2020	#N/A
<i>Cephalopholis argus</i>	Serranidae	Scaffold	1,103,168,673	115,718	#N/A	Newly sequenced	#N/A

<i>Cromileptes altivelis</i>	Serranidae	Chromosome	1,013,370,389	43,466,551	18,269,829	GCA_013133815.1	2020/5/20
<i>Epinephelus akaara</i>	Serranidae	Scaffold	1,135,726,210	46,067,422	#N/A	Ge et al., 2019	#N/A
<i>Epinephelus coioides</i>	Serranidae	Scaffold	1,023,559,032	2,567,038	2,494,056	GCA_900536245.1	2019/12/5
<i>Epinephelus fuscoguttatus</i>	Serranidae	Chromosome	1,046,995,804	44,419,306	13,857,438	GCA_011397635.1	2020/2/20
<i>Epinephelus lanceolatus</i>	Serranidae	Chromosome	1,087,399,367	46,227,939	159,800	GCA_005281545.1	2019/5/13
<i>Cephalopholis sonnerati</i>	Serranidae	Scaffold	1,043,730,103	43,987,100	2,482,587	GCA_019142605.1	2021/7/6
<i>Plectropomus leopardus</i>	Serranidae	Chromosome	787,600,493	34,616,500	1,139,749	GCA_011397275.1	2020/2/20
<i>Hypoplectrus puella</i>	Serranidae	Scaffold	612,290,098	24,210,077	22,581	GCA_900610375.1	2018/12/12
<i>Serranus cabrilla</i>	Serranidae	Scaffold	630,578,932	610,905	21,564	GCA_903364295.1	2020/6/5
<i>Etheostoma cragini</i>	#N/A	Chromosome	643,078,674	27,593,271	45,088	GCA_013103735.1	2020/5/15
<i>Etheostoma spectabile</i>	#N/A	Chromosome	854,806,606	30,497,795	26,023	GCA_008692095.1	2019/10/7
<i>Perca flavescens</i>	#N/A	Chromosome	877,456,336	37,412,490	4,268,950	GCA_004354835.1	2019/3/19
<i>Perca fluviatilis</i>	#N/A	Scaffold	958,225,486	6,260,519	18,196	GCA_003412525.1	2018/8/22
<i>Percina caprodes</i>	#N/A	Scaffold	1,011,959,657	1,429,287	1,429,287	GCA_013375575.1	2020/6/29
<i>Sander lucioperca</i>	#N/A	Scaffold	900,461,225	4,929,547	4,695,595	GCA_008315115.1	2019/9/6
<i>Platycephalidae</i>	#N/A	Scaffold	689,669,459	1,019,513	#N/A	Newly sequenced	#N/A
<i>sp.</i>	#N/A	Scaffold	635,682,126	8,111,519	#N/A	Newly sequenced	#N/A
<i>Inimicus sinensis</i>	#N/A	Scaffold	492,122,949	1,105,004	#N/A	Newly sequenced	#N/A
<i>Uranoscopus</i>	Trachiniformes	Scaffold	612,774,492	5,717,598	41,757	GCA_004355925.1	2019/3/8
<i>ocellatus</i>	Scorpaeniformes	Scaffold	699,326,415	5,156	5,156	GCA_000499045.1	2013/11/25
<i>Anoplopoma fimbria</i>	Scorpaeniformes	Contig	615,153,901	29,138,660	13,037,006	GCA_910589615.1	2021/7/7
<i>Taurulus bubalis</i>	Scorpaeniformes	Chromosome	572,885,595	23,863,342	4,950,682	GCA_009769545.1	2019/12/23
<i>Cyclopterus lumpus</i>	Scorpaeniformes	Chromosome	428,916,835	18,096,870	5,710,868	GCA_021346845.1	2022/1/3
<i>Apeltes quadracus</i>	Scorpaeniformes	Chromosome	599,537,287	17,641,857	544,634	GCA_014132575.1	2020/7/29
<i>Gasterosteus nipponicus</i>	Scorpaeniformes	Scaffold	467,452,432	3,715,221	38,090	GCA_006229165.1	2019/6/11
<i>Gasterosteus aculeatus</i>	Scorpaeniformes	Scaffold	441,089,565	302,682	14,136	GCA_003399555.1	2018/8/16
<i>Pungitius pungitius</i>	Scorpaeniformes	Scaffold	635,567,917	5,092,707	31,240	GCA_004358465.1	2019/3/8
<i>Ophiodon elongatus</i>	Scorpaeniformes	Scaffold	498,979,456	375,216	9,903	GCA_006348945.1	2019/6/14
<i>Liparis tanakae</i>	Scorpaeniformes	Scaffold	588,696,950	25,364,761	3,519,450	GCA_910591455.2	2021/11/18
<i>Pholis gunnellus</i>	Scorpaeniformes	Chromosome	832,469,038	53,184	29,796	GCA_916700685.1	2021/10/9
<i>Adelosebastes latens</i>	Scorpaeniformes	Scaffold	834,246,755	41,075	23,430	GCA_916700935.1	2021/10/9
<i>Hozukius emblemarius</i>	Scorpaeniformes	Scaffold	725,092,264	16,662	16,662	GCA_004335335.1	2019/3/8
<i>Sebastes koreanus</i>	Scorpaeniformes	Contig	724,045,237	11,290	11,290	GCA_004335365.1	2019/3/8
<i>Sebastes nudus</i>	Scorpaeniformes	Contig					

<i>Sebastes schlegelii</i>	Scorpaeniformes	Contig	728,476,695	14,246	14,246	GCA_004335315.1	2019/3/8
<i>Sebastes umbrosus</i>	Scorpaeniformes	Chromosome	800,904,020	34,912,176	11,445,908	GCA_015220745.1	2020/11/3
<i>Sebastes albobasiscus</i>	Scorpaeniformes	Scaffold	787,252,963	66,236	41,071	GCA_916701655.1	2021/10/9
<i>Sebastes tertius</i>	Scorpaeniformes	Scaffold	772,441,018	60,826	42,036	GCA_916701665.1	2021/11/13
<i>Sebastolobus alascanus</i>	Scorpaeniformes	Scaffold	858,864,611	78,318	35,928	GCA_916701645.1	2021/10/9
<i>Sebastolobus altivelis</i>	Scorpaeniformes	Scaffold	851,925,253	56,458	29,583	GCA_916701675.1	2021/10/9
<i>Sebastes norvegicus</i>	Scorpaeniformes	Scaffold	717,740,616	16,564	9,467	GCA_900302655.1	2018/3/21
<i>Sebastes rubrivinctus</i>	Scorpaeniformes	Scaffold	756,296,653	30,046	13,541	GCA_000475215.1	2013/10/22
<i>Sebastes aleutianus</i>	Scorpaeniformes	Scaffold	899,650,391	340,062	10,838	GCA_001910805.2	2016/12/20
<i>Sebastes nigrocinctus</i>	Scorpaeniformes	Scaffold	746,044,620	116,274	13,471	GCA_000475235.3	2016/12/20
<i>Cebidichthys violaceus</i>	Scorpaeniformes	Scaffold	593,001,491	6,717,604	5,507,153	GCA_008087265.1	2019/8/26
<i>Gymnodraco acuticeps</i>	Notothenioidei	Scaffold	996,888,495	1,921,211	534,678	GCA_902827175.1	2020/4/15
<i>Parachaenichthys charcoti</i>	Notothenioidei	Scaffold	797,053,978	178,902	#N/A	Ahn et al., 2017	#N/A
<i>Cottopeca gobio</i>	Notothenioidei	Chromosome	609,391,784	25,156,145	6,330,900	GCA_900634415.1	2019/2/6
<i>Chionodraco hamatus</i>	Notothenioidei	Scaffold	829,466,164	51,057	2,261	GCA_009756495.1	2019/12/16
<i>Chionodraco myersi</i>	Notothenioidei	Scaffold	1,120,102,150	48,322	42,918	GCA_009756435.1	2019/12/16
<i>Eleginops maclovinus</i>	Notothenioidei	Scaffold	744,404,235	704,474	#N/A	Chen et al., 2019	#N/A
<i>Harpagifer antarcticus</i>	Notothenioidei	Scaffold	941,503,312	4,997,231	1,048,452	GCA_902827135.1	2020/4/15
<i>Dissostichus mawsoni</i>	Notothenioidei	Scaffold	926,363,279	36,983,948	3,015,437	GCA_011823955.1	2020/4/1
<i>Notothenia coriiceps</i>	Notothenioidei	Scaffold	636,613,682	217,655	17,492	GCA_000735185.1	2014/7/29
<i>Trematomus bernacchii</i>	Notothenioidei	Scaffold	867,125,071	8,748,152	1,402,669	GCA_902827165.1	2020/4/15
<i>Trematomus loennbergii</i>	Notothenioidei	Scaffold	1,210,523,816	15,527,446	968,532	GCA_012861695.1	2020/4/27
<i>Gerres acinaces</i>	#N/A	Scaffold	544,497,604	5,361,428	#N/A	Newly sequenced	#N/A
<i>Cheilinus undulatus</i>	Labriformes	Chromosome	1,173,472,134	51,466,935	16,477,222	GCA_018320785.1	2021/5/12
<i>Tautoglabrus adspersus</i>	Labriformes	Chromosome	723,582,185	31,691,308	7,919,333	GCA_020745685.1	2021/11/5
<i>Thalassoma bifasciatum</i>	Labriformes	Scaffold	1,095,910,316	155,821	122,955	GCA_008086565.1	2019/8/2
<i>Symphodus melops</i>	Labriformes	Scaffold	533,823,763	21,275	9,362	GCA_900323315.1	2018/3/26
<i>Labrus bergylta</i>	Labriformes	Scaffold	805,480,521	794,648	703,847	GCA_900080235.1	2016/5/3
<i>Notolabrus celidotus</i>	Labriformes	Chromosome	846,760,650	37,090,664	3,748,462	GCA_009762535.1	2019/12/18
<i>Scaridae sp.</i>	Labriformes	Scaffold	698,084,642	14,207,675	#N/A	Newly sequenced	#N/A
<i>Lateolabrax maculatus</i>	#N/A	Scaffold	729,326,463	927,298	#N/A	GCA_004023545.1	#N/A
<i>Therapon oxyrhynchus</i>	Centrarchiformes	Scaffold	724,392,045	11,679,792	#N/A	Newly sequenced	#N/A
<i>Kuhlia marginata</i>	Centrarchiformes	Scaffold	674,444,388	4,755,420	#N/A	Newly sequenced	#N/A

<i>Oplegnathus fasciatus</i>	Centrarchiformes	Scaffold	766,301,214	1,126,915	29,992	GCA_003416845.1	2018/8/22
<i>Maccullochella peelii</i>	Centrarchiformes	Scaffold	633,241,041	109,974	70,439	GCA_002120245.1	2017/5/9
<i>Macquaria ambigua</i>	Centrarchiformes	Scaffold	661,429,323	252,501	243,852	GCA_008360985.1	2019/9/9
<i>Macquaria australasica</i>	Centrarchiformes	Scaffold	675,976,139	845,515	678,975	GCA_005408345.1	2019/5/16
<i>Micropterus dolomieu</i>	Centrarchiformes	Chromosome	829,587,215	34,173,225	46,775	GCA_021292245.1	2022/1/11
<i>Micropterus floridanus</i>	Centrarchiformes	Contig	1,001,521,525	10,978	10,978	GCA_002592385.1	2017/10/19
<i>Micropterus salmoides</i>	Centrarchiformes	Scaffold	963,613,914	36,481,429	1,227,323	GCA_014851395.1	2020/10/5
<i>Coreoperca whiteheadi</i>	Centrarchiformes	Chromosome	712,451,466	28,603,870	76,587	GCA_011952105.1	2020/4/9
<i>Siniperca scherzeri</i>	Centrarchiformes	Chromosome	736,162,835	30,166,107	83,753	GCA_011952095.1	2020/4/9
<i>Siniperca chuatsi</i>	Centrarchiformes	Chromosome	755,045,244	30,508,166	12,191,788	GCA_011952085.1	2020/4/9
<i>Siniperca kneri</i>	Centrarchiformes	Chromosome	723,564,461	29,777,865	77,506	GCA_011952075.1	2020/4/9
<i>Siganus vulpinus</i>	#N/A	Scaffold	580,888,565	1,195,464	#N/A	Newly sequenced	#N/A
<i>Lophius piscatorius</i>	Lophiiformes	Contig	747,363,853	104,538	104,538	GCA_009660295.1	2019/11/13
<i>Antennarius striatus</i>	Lophiiformes	Scaffold	441,856,641	9,774	6,086	GCA_900303275.1	2018/3/19
<i>Antennarius maculatus</i>	Lophiiformes	Chromosome	549,949,511	22,928,215	10,114,706	GCA_013358685.1	2020/6/12
<i>Scatophagus argus</i>	#N/A	Chromosome	570,802,272	24,451,451	17,189,422	GCA_020382885.1	2021/10/13
<i>Balistidae sp.</i>	Tetraodontiformes	Scaffold	595,235,161	1,575,972	#N/A	Newly sequenced	#N/A
<i>Sufflamen fraenatus</i>	Tetraodontiformes	Scaffold	585,083,749	9,046,690	#N/A	Newly sequenced	#N/A
<i>Melichthys vidua</i>	Tetraodontiformes	Scaffold	595,633,782	3,626,994	#N/A	Newly sequenced	#N/A
<i>Diodon liturosus</i>	Tetraodontiformes	Scaffold	654,040,692	7,140,086	#N/A	Newly sequenced	#N/A
<i>Diodon holocanthus</i>	Tetraodontiformes	Scaffold	650,024,253	8,129,349	#N/A	Newly sequenced	#N/A
<i>Mola mola</i>	Tetraodontiformes	Scaffold	639,451,992	8,766,736	23,239	GCA_001698575.1	2016/8/8
<i>Cantherhines dumerili</i>	Tetraodontiformes	Scaffold	480,562,930	773,060	#N/A	Newly sequenced	#N/A
<i>Monacanthus chinensis</i>	Tetraodontiformes	Scaffold	433,808,282	9,128,308	#N/A	Newly sequenced	#N/A
<i>Thamnaconus septentrionalis</i>	Tetraodontiformes	Chromosome	474,309,635	23,049,015	22,457,894	GCA_009823395.1	2020/1/2
<i>Arothron hispidus</i>	Tetraodontiformes	Scaffold	379,525,681	2,258,198	#N/A	Newly sequenced	#N/A
<i>Monotremus palembangensis</i>	Tetraodontiformes	Scaffold	359,745,401	5,700,007	#N/A	Newly sequenced	#N/A
<i>Takifugu bimaculatus</i>	Tetraodontiformes	Chromosome	371,675,663	16,786,025	1,398,332	GCA_004026145.1	2019/1/16
<i>Takifugu flavidus</i>	Tetraodontiformes	Chromosome	366,286,831	15,676,631	4,357,567	GCA_003711565.2	2019/8/16
<i>Takifugu rubripes</i>	Tetraodontiformes	Chromosome	384,126,662	16,705,553	3,136,617	GCA_901000725.2	2019/6/5
<i>Tetraodon fluviatilis</i>	Tetraodontiformes	Scaffold	326,729,102	7,010,526	#N/A	Newly sequenced	#N/A
<i>Tetraodon nigroviridis</i>	Tetraodontiformes	Scaffold	342,403,326	734,039	29,054	GCA_000180735.1	2004/5/14
<i>Tetraodon palembangensis</i>	Tetraodontiformes	Scaffold	359,409,263	7,036,088	#N/A	Newly sequenced	#N/A

<i>Pao palembangensis</i>	Tetraodontiformes	Chromosome	356,041,988	15,808,960	3,269,158	GCA_015343265.1	2020/11/10
<i>Arothron firmamentum</i>	Tetraodontiformes	Scaffold	689,581,009	15,567	7,930	GCA_016586285.1	2020/12/9
<i>Triacanthus biaculeatus</i>	Tetraodontiformes	Scaffold	493,070,606	6,475,552	#N/A	Newly sequenced	#N/A
<i>Sillaginidae sp.</i>	#N/A	Scaffold	554,949,161	1,681,751	#N/A	Newly sequenced	#N/A
<i>Sillago japonica</i>	#N/A	Scaffold	601,147,888	4,030,735	#N/A	Newly sequenced	#N/A
<i>Morone chrysops</i>	#N/A	Scaffold	620,984,155	51,932	20,318	GCA_003610055.1	2018/10/2
<i>Dicentrarchus labrax</i>	#N/A	Scaffold	675,917,103	26,439,989	54,134	GCA_000689215.1	2014/4/9
<i>Morone saxatilis</i>	#N/A	Scaffold	598,109,547	25,942,274	17,211	GCA_004916995.1	2019/4/26
<i>Collichthys lucidus</i>	#N/A	Chromosome	877,615,421	35,919,243	1,098,566	GCA_004119915.1	2019/1/30
<i>Johnius grypotus</i>	#N/A	Scaffold	588,004,829	7,836,149	#N/A	Newly sequenced	#N/A
<i>Larimichthys crocea</i>	#N/A	Chromosome	657,939,657	27,037,660	277,487	GCA_000972845.2	2018/11/27
<i>Larimichthys polyactis</i>	#N/A	Scaffold	694,446,753	146,739	8,079	GCA_010119295.1	2020/2/3
<i>Nibea albiflora</i>	#N/A	Scaffold	574,466,150	2,154,052	55,145	GCA_900327885.1	2018/12/5
<i>Argyrosomus japonicus</i>	#N/A	Contig	334,905,208	134,966	17,030	GCA_015710095.1	2020/12/2
<i>Sciaenops ocellatus</i>	#N/A	Chromosome	684,633,790	25,657,701	151,930	GCA_014183145.1	2020/8/13
<i>Miichthys miiuy</i>	#N/A	Scaffold	619,300,777	1,145,539	81,271	GCA_001593715.1	2016/3/17
<i>Acanthopagrus latus</i>	#N/A	Chromosome	806,072,371	30,688,393	1,174,323	GCA_013436515.1	2020/7/21
<i>Spondyliosoma cantharus</i>	#N/A	Scaffold	680,472,139	28,198	11,633	GCA_900302685.1	2018/3/21
<i>Diplodus sargus</i>	#N/A	Scaffold	784,049,296	3,406,122	27,464	GCA_903131615.1	2020/6/5
<i>Sparus aurata</i>	#N/A	Chromosome	833,595,063	35,791,275	2,862,625	GCA_900880675.1	2019/7/30
<i>Lutjanus erythropterus</i>	#N/A	Chromosome	969,885,659	40,645,435	2,781,023	GCA_020091685.1	2021/9/23
<i>Lutjanus cyanopterus</i>	#N/A	Scaffold	1,220,328,034	540,867	#N/A	Newly sequenced	#N/A
<i>Lutjanus bohar</i>	#N/A	Scaffold	1,090,887,124	8,770,032	#N/A	Newly sequenced	#N/A
<i>Lutjanus gibbus</i>	#N/A	Scaffold	1,036,077,630	11,472,890	#N/A	Newly sequenced	#N/A
<i>Lutjanus johni</i>	#N/A	Scaffold	990,881,675	12,969,207	#N/A	Newly sequenced	#N/A
<i>Lutjanus kasmira</i>	#N/A	Scaffold	972,793,156	12,954,483	#N/A	Newly sequenced	#N/A
<i>Lutjanus rivulatus</i>	#N/A	Scaffold	1,044,188,900	15,208,033	#N/A	Newly sequenced	#N/A
<i>Lutjanus fulviflamma</i>	#N/A	Scaffold	1,063,792,642	12,380,625	#N/A	Newly sequenced	#N/A
<i>Datnioides pulcher</i>	#N/A	Scaffold	594,964,832	9,730,178	#N/A	Newly sequenced	#N/A
<i>Datnioides undecimradiatus</i>	#N/A	Scaffold	595,034,438	9,689,676	3,802,431	GCA_008933995.1	2019/10/10
<i>Leiognathus brevirostris</i>	Acanthuriformes	Scaffold	762,281,169	2,661,087	#N/A	Newly sequenced	#N/A
<i>Centropyge bicolor</i>	Acanthuriformes	Scaffold	681,285,455	4,424,004	#N/A	Newly sequenced	#N/A

<i>Chaetodon auriga</i>	Acanthuriformes	Scaffold	636,774,498	4,684,412	#N/A	Newly sequenced	#N/A
<i>Chaetodon sp.</i>	Acanthuriformes	Scaffold	636,530,979	2,934,856	#N/A	Newly sequenced	#N/A
<i>Chaetodon trifasciatus</i>	Acanthuriformes	Scaffold	667,687,180	9,959,163	#N/A	Newly sequenced	#N/A
<i>Chelmon rostratus</i>	Acanthuriformes	Chromosome	644,208,416	27,778,429	16,959,746	GCA_017976325.1	2021/4/15
<i>Naso vlamingii</i>	Acanthuriformes	Scaffold	860,946,575	5,734,092	#N/A	Newly sequenced	#N/A
<i>Zebrasoma veliferum</i>	Acanthuriformes	Scaffold	681,534,212	2,076,120	#N/A	Newly sequenced	#N/A
<i>Naso lituratus</i>	Acanthuriformes	Scaffold	655,661,147	58,364	#N/A	Newly sequenced	#N/A
<i>Zebrasoma flavescens</i>	Acanthuriformes	Scaffold	681,683,760	2,234,300	#N/A	Newly sequenced	#N/A
<i>Monotaxis grandoculis</i>	Lethrinidae	Scaffold	1,234,128,523	7,399,509	#N/A	Newly sequenced	#N/A
<i>Lethrinus olivaceus</i>	Lethrinidae	Scaffold	1,088,771,447	5,239,893	#N/A	Newly sequenced	#N/A
<i>Plectorhinchus chaetodonoides</i>	Haemulidae	Scaffold	651,761,722	3,190,179	#N/A	Newly sequenced	#N/A

附表 4. 已公布的海洋四足动物基因组基本信息

Species	Class	Order	Family	Genus	Accession	Assembly level	Total length (bp)	Scaffold N50 (bp)	Contig N50 (bp)	GC (%)	Release Date
<i>Merops pusillus</i>	Aves	Coraciiformes	Meropidae	Merops	GCA_025762695.1	Scaffold	942,069,649	56,230	21,654	42.50	2022/10/19
<i>Fringilla coelebs</i>	Aves	Passeriformes	Fringillidae	Fringilla	GCA_015532645.2	Chromosome	994,865,760	69,727,445	66,810	42.00	2021/6/16
<i>Oenanthe oenanthe</i>	Aves	Passeriformes	Muscicapidae	Oenanthe	GCA_013399995.1	Scaffold	1,001,836,003	507,914	44,998	41.50	2020/7/10
<i>Anhinga rufa</i>	Aves	Pelecaniformes	Anhingidae	Anhinga	GCA_013400875.1	Scaffold	1,014,091,971	16,986	12,297	42.50	2020/7/10
<i>Parus major</i>	Aves	Passeriformes	Paridae	Parus	GCA_001522545.3	Chromosome	1,020,310,769	71,365,269	148,693	41.50	2019/1/15
<i>Alauda arvensis</i>	Aves	Passeriformes	Alaudidae	Alauda	GCA_902810485.1	Scaffold	1,028,664,229	1,430,931	79,699	42.00	2020/3/13
<i>Corvus cornix</i>	Aves	Passeriformes	Corvidae	Corvus	GCA_000738735.6	Chromosome	1,030,881,721	73,732,264	8,553,235	42.00	2022/4/1
<i>Sturnus vulgaris</i>	Aves	Passeriformes	Sturnidae	Sturnus	GCA_001447265.1	Scaffold	1,036,755,994	3,416,708	151,865	41.50	2015/11/20
<i>Passer montanus</i>	Aves	Passeriformes	Passeridae	Passer	GCA_014805655.1	Scaffold	1,039,949,367	11,160,574	1,083,540	42.00	2020/9/30
<i>Carduelis carduelis</i>	Aves	Passeriformes	Fringillidae	Carduelis	GCA_032274745.1	Scaffold	1,043,792,359	72,471,425	35,533	42.00	2023/10/2
<i>Sylvia borin</i>	Aves	Passeriformes	Sylviidae	Sylvia	GCA_014839755.1	Chromosome	1,045,652,180	72,750,092	31,914,676	42.00	2020/10/2
<i>Protonotaria citrea</i>	Aves	Passeriformes	Parulidae	Protonotaria	GCA_026259555.1	Scaffold	1,048,280,936	71,442,492	1,472	44.50	2022/11/21
<i>Chloroceryle aenea</i>	Aves	Coraciiformes	Alcedinidae	Chloroceryle	GCA_013399075.1	Scaffold	1,062,127,307	324,840	34,835	41.50	2020/7/10
<i>Merops nubicus</i>	Aves	Coraciiformes	Meropidae	Merops	GCA_000691845.1	Scaffold	1,062,961,556	48,089	24,675	41.50	2014/5/16
<i>Sylvia atricapilla</i>	Aves	Passeriformes	Sylviidae	Sylvia	GCA_009819655.1	Chromosome	1,066,786,587	72,978,054	7,345,957	42.00	2019/12/31
<i>Motacilla alba</i>	Aves	Passeriformes	Motacillidae	Motacilla	GCA_015832195.1	Chromosome	1,072,653,902	72,386,170	8,515,530	42.50	2020/12/8
<i>Acrocephalus scirpaceus</i>	Aves	Passeriformes	Sylviidae	Acrocephalus	GCA_910950805.1	Chromosome	1,075,083,815	74,438,198	12,742,779	42.00	2021/8/16
<i>Geothlypis trichas</i>	Aves	Passeriformes	Parulidae	Geothlypis	GCA_009764595.1	Chromosome	1,078,128,490	72,541,719	3,247,339	42.0886	2019/12/19
<i>Somateria mollissima</i>	Aves	Anseriformes	Anatidae	Somateria	GCA_951416345.1	Chromosome	1,079,261,109	78,490,558	9,245,871	43.00	2023/6/27
<i>Aix galericulata</i>	Aves	Anseriformes	Anatidae	Aix	GCA_024635365.1	Scaffold	1,079,840,820	78,198,423	77,396	41.50	2022/8/16
<i>Muscicapa striata</i>	Aves	Passeriformes	Muscicapidae	Muscicapa	GCA_030060385.1	Contig	1,083,165,776	3,249,550	3,249,550	42.00	2023/5/22
<i>Erithacus rubecula</i>	Aves	Passeriformes	Saxicolidae	Erithacus	GCA_903797595.2	Chromosome	1,086,738,418	60,684,490	5,585,877	42.0925	2021/5/3
<i>Pitta sordida</i>	Aves	Passeriformes	Pittidae	Pitta	GCA_013390025.1	Scaffold	1,088,830,872	176,673	41,696	42.50	2020/7/6
<i>Corvus brachyrhynchos</i>	Aves	Passeriformes	Corvidae	Corvus	GCA_000691975.1	Scaffold	1,091,312,783	6,953,989	29,093	42.00	2014/6/2
<i>Riparia riparia</i>	Aves	Passeriformes	Hirundinidae	Riparia	GCA_020917445.1	Scaffold	1,093,155,110	13,690,488	43,283	41.00	2021/11/22
<i>Lophonetta specularioides</i>	Aves	Anseriformes	Anatidae	Lophonetta	GCA_032165235.1	Scaffold	1,093,695,053	79,729,295	50,151	41.50	2023/9/28
<i>Tachyeres brachypterus</i>	Aves	Anseriformes	Anatidae	Tachyeres	GCA_032357605.1	Scaffold	1,094,081,176	79,483,301	29,025	41.50	2023/10/4

<i>Anas georgica</i>	Aves	Anseriformes	Anatidae	Anas	GCA_032275265.1	Scaffold	1,098,703,473	80,592,087	29,010	41.50	2023/10/2
<i>Apus apus</i>	Aves	Apodiformes	Apodidae	Apus	GCA_020740795.1	Chromosome	1,100,361,224	115,508,478	25,561,769	42.00	2021/11/4
<i>Hirundo rustica</i>	Aves	Passeriformes	Hirundinidae	Hirundo	GCA_015227805.3	Chromosome	1,105,955,550	76,187,387	2,794,774	42.50	2021/11/30
<i>Fratercula cirrhata</i>	Aves	Charadriiformes	Alcidae	Fratercula	GCA_014363165.1	Scaffold	1,105,963,355	54,950	19,680	42.00	2020/8/22
<i>Anas capensis</i>	Aves	Anseriformes	Anatidae	Anas	GCA_032357685.1	Scaffold	1,106,937,217	64,715,642	9,950	41.00	2023/10/4
<i>Chloroceryle amazona</i>	Aves	Coraciiformes	Alcedinidae	Chloroceryle	GCA_027560015.1	Scaffold	1,110,454,235	60,666	27,367	42.00	2023/1/3
<i>Anser brachyrhynchus</i>	Aves	Anseriformes	Anatidae	Anser	GCA_002592135.1	Scaffold	1,116,992,394	4,974,386	97,462	42.00	2017/10/19
<i>Chaetura pelagica</i>	Aves	Apodiformes	Apodidae	Chaetura	GCA_000747805.2	Scaffold	1,118,724,105	3,841,852	30,767	41.50	2014/9/4
<i>Ceuthmochares aereus</i>	Aves	Cuculiformes	Cuculidae	Ceuthmochares	GCA_013398935.1	Scaffold	1,120,248,821	163,031	29,269	41.00	2020/7/10
<i>Cygnus atratus</i>	Aves	Anseriformes	Anatidae	Cygnus	GCA_013377495.2	Chromosome	1,124,217,020	84,552,148	28,995,010	42.00	2022/8/3
<i>Clangula hyemalis</i>	Aves	Anseriformes	Anatidae	Clangula	GCA_029619115.1	Scaffold	1,130,355,540	548,271	47,528	41.00	2023/4/7
<i>Haliaeetus albicilla</i>	Aves	Falconiformes	Accipitridae	Haliaeetus	GCA_000691405.1	Scaffold	1,133,549,865	57,319	25,143	41.00	2014/5/15
<i>Falco tinnunculus</i>	Aves	Falconiformes	Falconidae	Falco	GCA_024430995.1	Scaffold	1,134,479,392	10,701,775	107,537	41.50	2022/7/25
<i>Podiceps cristatus</i>	Aves	Podicipediformes	Podicipedidae	Podiceps	GCA_000699545.1	Scaffold	1,134,922,578	30,087	17,412	41.50	2014/6/11
<i>Aythya marila</i>	Aves	Anseriformes	Anatidae	Aythya	GCA_029042245.1	Chromosome	1,135,194,460	85,439,399	32,460,729	42.00	2023/3/15
<i>Sula dactylatra</i>	Aves	Pelecaniformes	Sulidae	Sula	GCA_013389905.1	Scaffold	1,136,665,023	156,242	38,906	42.00	2020/7/6
<i>Phalacrocorax carbo</i>	Aves	Pelecaniformes	Phalacrocoracidae	Phalacrocorax	GCA_000708925.1	Scaffold	1,138,967,842	48,427	17,343	41.50	2014/6/17
<i>Fulmarus glacialis</i>	Aves	Procellariiformes	Procellariidae	Fulmarus	GCA_000690835.1	Scaffold	1,141,395,646	47,208	25,926	41.00	2014/5/13
<i>Gallinago hardwickii</i>	Aves	Charadriiformes	Scolopacidae	Gallinago	GCA_026122665.1	Scaffold	1,142,750,823	77,540,074	25,428	42.00	2022/11/10
<i>Tringa solitaria</i>	Aves	Charadriiformes	Scolopacidae	Tringa	GCA_026413425.1	Scaffold	1,150,739,019	79,619,486	15,233	42.00	2022/11/27
<i>Phaetusa simplex</i>	Aves	Charadriiformes	Laridae	Phaetusa	GCA_013399535.1	Scaffold	1,151,692,184	520,939	16,901	42.50	2020/7/10
<i>Cathartes aura</i>	Aves	Accipitriformes	Cathartidae	Cathartes	GCA_000699945.1	Scaffold	1,152,571,117	36,359	15,248	41.00	2014/6/11
<i>Phaethon lepturus</i>	Aves	Pelecaniformes	Phaethontidae	Phaethon	GCA_000687285.1	Scaffold	1,152,958,507	47,896	22,941	41.50	2014/5/8
<i>Gallinago paraguaiae</i>	Aves	Charadriiformes	Scolopacidae	Gallinago	GCA_025728075.1	Scaffold	1,153,592,092	77,818,824	15,517	42.50	2022/10/17
<i>Athene cunicularia</i>	Aves	Strigiformes	Strigidae	Athene	GCA_003259725.1	Scaffold	1,157,069,330	42,147,404	64,091	41.00	2018/6/21
<i>Stercorarius parasiticus</i>	Aves	Charadriiformes	Stercorariidae	Stercorarius	GCA_013396915.1	Scaffold	1,159,854,254	644,604	107,045	42.00	2020/7/10
<i>Pelecanus crispus</i>	Aves	Pelecaniformes	Pelecanidae	Pelecanus	GCA_000687375.1	Scaffold	1,160,924,693	43,364	21,679	41.50	2014/5/9
<i>Phoebastria albatrus</i>	Aves	Procellariiformes	Diomedidae	Phoebastria	GCA_014363385.1	Scaffold	1,165,200,156	87,989	43,134	42.00	2020/8/22
<i>Progne subis</i>	Aves	Passeriformes	Hirundinidae	Progne	GCA_022316685.1	Scaffold	1,165,951,862	6,129,949	3,084,734	43.00	2022/2/17
<i>Aramus guarauna</i>	Aves	Gruiformes	Aramidae	Aramus	GCA_013400195.1	Scaffold	1,170,510,309	810,609	62,981	42.00	2020/7/10
<i>Tachycineta bicolor</i>	Aves	Passeriformes	Hirundinidae	Tachycineta	GCA_025960845.1	Scaffold	1,172,536,026	15,594,027	58,357	42.50	2022/11/2

<i>Glareola pratincola</i>	Aves	Charadriiformes	Glareolidae	Glareola	GCA_013401295.1	Scaffold	1,172,993,239	411,631	39,898	42.50	2020/7/10
<i>Fregetta grallaria</i>	Aves	Procellariiformes	Hydrobatidae	Fregetta	GCA_013399335.1	Scaffold	1,173,305,012	342,041	15,821	42.00	2020/7/10
<i>Xenus cinereus</i>	Aves	Charadriiformes	Scolopacidae	Xenus	GCA_025774025.1	Scaffold	1,173,667,304	80,336,233	22,265	42.00	2022/10/20
<i>Chionis minor</i>	Aves	Charadriiformes	Chionidae	Chionis	GCA_013398295.1	Scaffold	1,174,003,224	592,381	128,427	42.50	2020/7/10
<i>Larus smithsonianus</i>	Aves	Charadriiformes	Laridae	Larus	GCA_013400295.1	Scaffold	1,174,553,315	82,245	59,826	42.50	2020/7/10
<i>Fregata magnificens</i>	Aves	Pelecaniformes	Fregatidae	Fregata	GCA_013389945.1	Scaffold	1,177,430,334	145,412	35,511	42.00	2020/7/6
<i>Rynchops niger</i>	Aves	Charadriiformes	Rynchopidae	Rynchops	GCA_013400035.1	Scaffold	1,178,085,471	523,074	57,549	42.50	2020/7/10
<i>Oceanites oceanicus</i>	Aves	Procellariiformes	Hydrobatidae	Oceanites	GCA_013396615.1	Scaffold	1,178,378,986	406,319	28,090	41.50	2020/7/10
<i>Haliaeetus leucocephalus</i>	Aves	Falconiformes	Accipitridae	Haliaeetus	GCA_000737465.1	Scaffold	1,178,409,481	9,145,499	105,493	41.50	2014/8/4
<i>Alca torda</i>	Aves	Charadriiformes	Alcidae	Alca	GCA_008658365.1	Chromosome	1,178,492,568	84,526,827	18,221,265	42.50	2019/10/10
<i>Nycticorax nycticorax</i>	Aves	Ciconiiformes	Ardeidae	Nycticorax	GCA_023375905.1	Scaffold	1,179,045,135	3,016,563	73,971	42.50	2022/5/16
<i>Uria lomvia</i>	Aves	Charadriiformes	Alcidae	Uria	GCA_002289315.1	Scaffold	1,179,358,714	15,847,591	24,899	42.50	2017/9/11
<i>Cuculus canorus</i>	Aves	Cuculiformes	Cuculidae	Cuculus	GCA_017976375.1	Chromosome	1,180,156,273	78,325,852	20,797,936	42.00	2021/4/15
<i>Uria aalge</i>	Aves	Charadriiformes	Alcidae	Uria	GCA_013401055.1	Scaffold	1,181,532,998	11,658,233	23,216	42.50	2020/7/10
<i>Pluvianellus socialis</i>	Aves	Charadriiformes	Pluvianellidae	Pluvianellus	GCA_013400795.1	Scaffold	1,185,839,265	121,952	48,629	42.50	2020/7/10
<i>Anas platyrhynchos</i>	Aves	Anseriformes	Anatidae	Anas	GCA_015476345.1	Chromosome	1,188,516,685	76,269,206	5,679,408	42.50	2020/11/23
<i>Agelaius phoeniceus</i>	Aves	Passeriformes	Fringillidae	Agelaius	GCA_020745825.3	Chromosome	1,188,530,404	36,435,137	22,833,505	43.00	2023/3/15
<i>Falco columbarius</i>	Aves	Falconiformes	Falconidae	Falco	GCA_027564205.1	Scaffold	1,188,729,421	90,981,074	39,527	42.00	2023/1/3
<i>Oceanodroma tethys</i>	Aves	Procellariiformes	Hydrobatidae	Oceanodroma	GCA_013397025.1	Scaffold	1,189,005,779	372,486	54,514	41.50	2020/7/10
<i>Accipiter cooperii</i>	Aves	Falconiformes	Accipitridae	Accipiter	GCA_024222105.1	Scaffold	1,190,545,496	35,883,487	5,194	42.00	2022/7/12
<i>Accipiter nisus</i>	Aves	Falconiformes	Accipitridae	Accipiter	GCA_004320145.1	Scaffold	1,190,649,881	91,241	35,336	42.00	2019/2/16
<i>Dromas ardeola</i>	Aves	Charadriiformes	Dromadidae	Dromas	GCA_013398915.1	Scaffold	1,192,853,411	314,166	40,991	42.00	2020/7/10
<i>Recurvirostra avosetta</i>	Aves	Charadriiformes	Recurvirostridae	Recurvirostra	GCA_004023745.1	Scaffold	1,193,282,475	96,778	10,462	41.9	2019/1/15
<i>Butorides virescens</i>	Aves	Ciconiiformes	Ardeidae	Butorides	GCA_017310255.1	Contig	1,194,621,758	222,988	222,988	43.00	2021/3/9
<i>Oceanodroma leucorhoa</i>	Aves	Procellariiformes	Hydrobatidae	Oceanodroma	GCA_030449065.1	Scaffold	1,195,518,234	8,693,377	195,584	42.00	2023/7/13
<i>Falco rusticolus</i>	Aves	Falconiformes	Falconidae	Falco	GCA_015220075.1	Chromosome	1,195,847,766	91,094,845	15,323,486	42.00	2020/11/3
<i>Himantopus himantopus</i>	Aves	Charadriiformes	Recurvirostridae	Himantopus	GCA_013400095.1	Scaffold	1,196,502,159	712,820	79,248	42.00	2020/7/10
<i>Podilymbus podiceps</i>	Aves	Podicipediformes	Podicipedidae	Podilymbus	GCA_013399565.1	Scaffold	1,197,106,122	692,269	35,810	42.00	2020/7/10
<i>Cephus grylle</i>	Aves	Charadriiformes	Alcidae	Cephus	GCA_013401065.1	Scaffold	1,197,650,647	16,187,446	39,689	42.50	2020/7/10

<i>Numenius phaeopus</i>	Aves	Charadriiformes	Scolopacidae	Numenius	GCA_030770645.1	Scaffold	1,199,643,465	3,118,629	106,225	42.00	2023/8/17
<i>Gallinago delicata</i>	Aves	Charadriiformes	Scolopacidae	Gallinago	GCA_027564035.1	Scaffold	1,201,500,499	77,961,257	6,412	42.50	2023/1/3
<i>Phalacrocorax harrisi</i>	Aves	Pelecaniformes	Phalacrocoracidae	Phalacrocorax	GCA_002173475.1	Scaffold	1,202,994,831	4,648,443	100,243	42.50	2017/6/6
<i>Psilopogon haemacephalus</i>	Aves	Piciformes	Megalaimidae	Psilopogon	GCA_013396835.1	Scaffold	1,204,560,348	297,144	36,269	44.50	2020/7/10
<i>Pelecanoides urinatrix</i>	Aves	Procellariiformes	Pelecanoididae	Pelecanoides	GCA_013400755.1	Scaffold	1,205,479,880	250,258	93,910	42.00	2020/7/10
<i>Egretta garzetta</i>	Aves	Ciconiiformes	Ardeidae	Egretta	GCA_000687185.1	Scaffold	1,206,484,573	3,067,157	29,019	42.00	2014/5/9
<i>Busarellus nigricollis</i>	Aves	Falconiformes	Accipitridae	Busarellus	GCA_026873725.1	Scaffold	1,206,635,864	36,361,711	37,081	41.50	2022/12/13
<i>Buteo lagopus</i>	Aves	Falconiformes	Accipitridae	Buteo	GCA_025727815.1	Scaffold	1,209,271,270	36,009,240	31,269	41.50	2022/10/17
<i>Buteo platypterus</i>	Aves	Falconiformes	Accipitridae	Buteo	GCA_025340775.1	Scaffold	1,209,786,575	36,168,614	41,170	41.50	2022/9/20
<i>Spheniscus mendiculus</i>	Aves	Sphenisciformes	Spheniscidae	Spheniscus	GCA_003264655.1	Scaffold	1,209,845,584	5,226,231	22,339	41.50	2018/6/25
<i>Phalacrocorax pelagicus</i>	Aves	Pelecaniformes	Phalacrocoracidae	Phalacrocorax	GCA_002173435.1	Scaffold	1,210,663,266	1,719,903	67,731	42.50	2017/6/6
<i>Eudypetes sclateri</i>	Aves	Sphenisciformes	Spheniscidae	Eudypetes	GCA_010078445.1	Scaffold	1,211,402,317	1,921,244	83,178	42.00	2020/1/31
<i>Haliaeetus leucogaster</i>	Aves	Falconiformes	Accipitridae	Haliaeetus	GCA_025740775.1	Scaffold	1,211,516,626	36,561,209	31,650	41.50	2022/10/18
<i>Fratercula arctica</i>	Aves	Charadriiformes	Alcidae	Fratercula	GCA_947846985.1	Scaffold	1,215,066,582	82,413,236	16,421,795	42.50	2023/2/3
<i>Pygoscelis adeliae</i>	Aves	Sphenisciformes	Spheniscidae	Pygoscelis	GCA_000699105.1	Scaffold	1,216,600,033	5,118,896	22,195	41.50	2014/6/6
<i>Puffinus mauretanicus</i>	Aves	Procellariiformes	Procellariidae	Puffinus	GCA_023333565.1	Scaffold	1,218,519,395	2,125,936	2,082,814	42.50	2022/5/11
<i>Strix aluco</i>	Aves	Strigiformes	Strigidae	Strix	GCA_031877785.1	Chromosome	1,219,048,950	89,677,789	13,783,470	42.50	2023/9/25
<i>Charadrius vociferus</i>	Aves	Charadriiformes	Charadriidae	Charadrius	GCA_000708025.2	Scaffold	1,219,859,583	3,657,050	39,278	42.00	2014/8/5
<i>Burhinus oedicnemus</i>	Aves	Charadriiformes	Burhinidae	Burhinus	GCA_008921705.1	Scaffold	1,228,108,199	1,937,420	93,678	40.8	2019/10/8
<i>Sterna hirundo</i>	Aves	Charadriiformes	Sternidae	Sterna	GCA_009819605.1	Chromosome	1,229,972,541	85,487,239	22,002,498	43.07	2019/12/31
<i>Charadrius alexandrinus</i>	Aves	Charadriiformes	Charadriidae	Charadrius	GCA_008711295.1	Scaffold	1,231,940,710	3,290,793	129,786	42.50	2019/9/27
<i>Gallinago gallinago</i>	Aves	Charadriiformes	Scolopacidae	Gallinago	GCA_027564015.1	Scaffold	1,232,026,665	81,304,857	5,018	42.00	2023/1/3
<i>Aquila chrysaetos</i>	Aves	Falconiformes	Accipitridae	Aquila	GCA_900496995.4	Chromosome	1,233,723,095	46,934,974	21,924,221	42.00	2021/3/26
<i>Buteo lineatus</i>	Aves	Falconiformes	Accipitridae	Buteo	GCA_025435075.1	Scaffold	1,239,587,557	33,451,961	42,301	41.50	2022/9/27
<i>Macheiramphus alcinus</i>	Aves	Falconiformes	Accipitridae	Macheiramphus	GCA_025727915.1	Scaffold	1,241,501,536	37,279,510	14,262	41.50	2022/10/17
<i>Eudypetes robustus</i>	Aves	Sphenisciformes	Spheniscidae	Eudypetes	GCA_010085315.1	Scaffold	1,241,676,055	367,542	83,286	42.00	2020/1/31
<i>Phalacrocorax auritus</i>	Aves	Pelecaniformes	Phalacrocoracidae	Phalacrocorax	GCA_002173455.1	Scaffold	1,246,050,644	1,990,987	97,559	42.50	2017/6/6
<i>Phoenicopterus ruber</i>	Aves	Phoenicopteriformes	Phoenicopteridae	Phoenicopterus	GCA_009819775.1	Chromosome	1,246,226,932	85,472,901	14,700,712	43.50	2019/12/31
<i>Phylloscopus collybita</i>	Aves	Passeriformes	Sylviidae	Phylloscopus	GCA_02786175.1	Contig	1,247,501,225	28,093,532	28,093,532	43.50	2023/1/17
<i>Pluvialis apricaria</i>	Aves	Charadriiformes	Charadriidae	Pluvialis	GCA_017639485.1	Chromosome	1,247,767,512	87,527,607	15,203,797	42.7679	2021/3/30

<i>Phylloscopus trochilus</i>	Aves	Passeriformes	Sylviidae	Phylloscopus	GCA_016584745.2	Scaffold	1,249,322,907	34,133,506	24,608,944	44.00	2022/11/18
<i>Aptenodytes patagonicus</i>	Aves	Sphenisciformes	Spheniscidae	Aptenodytes	GCA_010087175.1	Scaffold	1,252,291,219	2,912,628	133,461	42.00	2020/1/31
<i>Aptenodytes forsteri</i>	Aves	Sphenisciformes	Spheniscidae	Aptenodytes	GCA_000699145.1	Scaffold	1,254,347,440	5,071,598	31,730	41.50	2014/6/6
<i>Grus americana</i>	Aves	Gruiformes	Gruidae	Grus	GCA_028858705.1	Chromosome	1,257,821,324	82,903,739	28,324,902	43.00	2023/2/24
<i>Calidris melanotos</i>	Aves	Charadriiformes	Scolopacidae	Calidris	GCA_028566615.2	Scaffold	1,260,162,641	60,089,602	3,168	41.50	2023/5/25
<i>Spheniscus magellanicus</i>	Aves	Sphenisciformes	Spheniscidae	Spheniscus	GCA_010076225.1	Scaffold	1,260,229,131	12,679,469	101,702	42.00	2020/1/31
<i>Pygoscelis antarcticus</i>	Aves	Sphenisciformes	Spheniscidae	Pygoscelis	GCA_010078415.1	Scaffold	1,260,447,050	6,271,244	136,952	42.00	2020/1/31
<i>Gallinago media</i>	Aves	Charadriiformes	Scolopacidae	Gallinago	GCA_025630025.2	Scaffold	1,265,690,539	67,977,983	3,241	41.50	2023/5/25
<i>Passerculus sandwichensis</i>	Aves	Passeriformes	Emberizidae	Passerculus	GCA_031885435.1	Chromosome	1,270,940,266	64,395,891	19,816,254	43.50	2023/9/22
<i>Spheniscus demersus</i>	Aves	Sphenisciformes	Spheniscidae	Spheniscus	GCA_010077935.1	Scaffold	1,275,043,525	15,386,364	109,372	42.00	2020/1/31
<i>Phalacrocorax aristotelis</i>	Aves	Pelecaniformes	Phalacrocoracidae	Phalacrocorax	GCA_949628215.1	Chromosome	1,279,134,750	78,889,319	11,420,000	43.00	2023/4/8
<i>Pygoscelis papua</i>	Aves	Sphenisciformes	Spheniscidae	Pygoscelis	GCA_030674165.1	Scaffold	1,283,624,771	14,305,032	138,660	42.00	2023/8/7
<i>Morus bassanus</i>	Aves	Pelecaniformes	Sulidae	Morus	GCA_031468815.1	Chromosome	1,294,804,326	85,624,757	10,178,591	42.50	2023/9/15
<i>Eudyptes pachyrhynchus</i>	Aves	Sphenisciformes	Spheniscidae	Eudyptes	GCA_010085335.1	Scaffold	1,295,505,775	8,912,950	58,247	42.50	2020/1/31
<i>Eudyptes moseleyi</i>	Aves	Sphenisciformes	Spheniscidae	Eudyptes	GCA_010082375.1	Scaffold	1,301,647,863	2,252,023	42,318	42.00	2020/1/31
<i>Eudyptes schlegeli</i>	Aves	Sphenisciformes	Spheniscidae	Eudyptes	GCA_010080425.1	Scaffold	1,308,180,354	1,894,246	42,596	42.00	2020/1/31
<i>Falco peregrinus</i>	Aves	Falconiformes	Falconidae	Falco	GCA_023634155.1	Chromosome	1,311,580,791	93,682,670	15,744,644	42.50	2022/6/2
<i>Megadyptes antipodes</i>	Aves	Sphenisciformes	Spheniscidae	Megadyptes	GCA_010078485.1	Scaffold	1,314,943,623	23,315,117	89,505	42.00	2020/1/31
<i>Gavia stellata</i>	Aves	Gaviiformes	Gaviidae	Gavia	GCA_030936135.1	Chromosome	1,321,287,916	50,637,386	13,765,203	43.00	2023/8/30
<i>Phalacrocorax brasilianus</i>	Aves	Pelecaniformes	Phalacrocoracidae	Phalacrocorax	GCA_002174335.1	Scaffold	1,346,185,690	86,910	35,201	42.50	2017/6/6
<i>Rissa tridactyla</i>	Aves	Charadriiformes	Laridae	Rissa	GCA_028500815.1	Chromosome	1,353,621,101	88,208,671	15,653,959	44.50	2023/2/9
<i>Spheniscus humboldti</i>	Aves	Sphenisciformes	Spheniscidae	Spheniscus	GCA_027474245.1	Chromosome	1,358,357,718	89,079,306	13,936,613	44.00	2022/12/28
<i>Eudyptes chrysolophus</i>	Aves	Sphenisciformes	Spheniscidae	Eudyptes	GCA_010084205.1	Scaffold	1,365,695,164	13,794,837	215,030	42.00	2020/1/31
<i>Accipiter gentilis</i>	Aves	Falconiformes	Accipitridae	Accipiter	GCA_929443795.2	Chromosome	1,398,030,142	35,025,567	17,712,854	44.00	2022/10/27
<i>Eudyptula minor</i>	Aves	Sphenisciformes	Spheniscidae	Eudyptula	GCA_010080355.1	Scaffold	1,462,615,782	21,127,646	95,403	42.00	2020/1/31
<i>Pontoporia blainvillei</i>	Mammalia	Cetartiodactyla	Pontoporiidae	Pontoporia	GCA_011754075.1	Scaffold	2,011,747,776	6,929	5,888	41.00	2020/3/25
<i>Balaenoptera bonaerensis</i>	Mammalia	Cetartiodactyla	Balaenopteridae	Balaenoptera	GCA_000978805.1	Scaffold	2,234,641,557	20,082	8,410	40.50	2015/3/18

<i>Neophocae na asiaeorient alis</i>	Mammalia	Cetartiodac tyla	Phocoenida e	Neophocae na	GCA_0030 31525.2	Scaffold	2,284,609, 912	6,341,296	86,003	41.50	2018/4/3
<i>Arctocephalus gazella</i>	Mammalia	Carnivora	Otariidae	Arctocephalus	GCA_9006 42305.1	Scaffold	2,313,585, 684	139,181,86 9	537,425	41.00	2019/2/8
<i>Ursus maritimus</i>	Mammalia	Carnivora	Ursidae	Ursus	GCA_0173 11325.1	Scaffold	2,330,485, 043	72,237,886	206,653	42.00	2021/3/9
<i>Steno bredanensi s</i>	Mammalia	Cetartiodac tyla	Delphinida e	Steno	GCA_0286 46385.1	Chromoso me	2,331,324, 105	94,897,118	59,977	41.50	2023/2/16
<i>Globicephala melas</i>	Mammalia	Cetartiodac tyla	Delphinida e	Globicephala	GCA_0065 47405.1	Scaffold	2,333,871, 032	18,102,937	332,801	41.50	2019/7/1
<i>Lagenorhynchus obliquidens</i>	Mammalia	Cetartiodac tyla	Delphinida e	Lagenorhynchus	GCA_0036 76395.1	Scaffold	2,334,450, 779	28,371,583	255,779	41.50	2018/10/23
<i>Halichoerus grypus</i>	Mammalia	Carnivora	Phocidae	Halichoerus	GCA_0123 93455.1	Scaffold	2,338,260, 283	1,033,864	1,011,981	41.50	2020/4/15
<i>Sousa chinensis</i>	Mammalia	Cetartiodac tyla	Delphinida e	Sousa	GCA_0077 60645.1	Scaffold	2,338,882, 461	19,436,979	113,766	41.50	2019/7/31
<i>Monodon monoceros</i>	Mammalia	Cetartiodac tyla	Monodonti dae	Monodon	GCA_0051 90385.3	Scaffold	2,341,954, 608	108,465,05 1	22,031,227	41.50	2022/11/9
<i>Vulpes lagopus</i>	Mammalia	Carnivora	Canidae	Vulpes	GCA_0183 45385.1	Chromoso me	2,345,533, 697	131,537,14 2	33,460,300	41.50	2021/5/12
<i>Pusa hispida</i>	Mammalia	Carnivora	Phocidae	Pusa	GCA_9470 44825.1	Contig	2,350,022, 681	3,664,905	3,664,905	41.50	2023/2/3
<i>Phoca vitulina</i>	Mammalia	Carnivora	Phocidae	Phoca	GCA_0043 48235.1	Scaffold	2,363,587, 482	41,024,070	283,887	41.50	2019/2/11
<i>Mirounga angustirost ris</i>	Mammalia	Carnivora	Phocidae	Mirounga	GCA_0212 88785.3	Chromoso me	2,366,166, 981	139,676,04 8	76,189	41.50	2022/12/21
<i>Phocoena sinus</i>	Mammalia	Cetartiodac tyla	Phocoenida e	Phocoena	GCA_0086 92025.1	Chromoso me	2,371,540, 524	115,469,29 2	20,218,762	41.50	2019/9/26
<i>Balaenoptera musculus</i>	Mammalia	Cetartiodac tyla	Balaenopte ridae	Balaenoptera	GCA_0098 73245.3	Chromoso me	2,374,868, 944	110,314,66 6	6,315,640	41.00	2020/10/2
<i>Tursiops truncatus</i>	Mammalia	Cetartiodac tyla	Delphinida e	Tursiops	GCA_0117 62595.1	Chromoso me	2,378,522, 214	108,430,13 5	9,729,386	41.50	2008/1/1
<i>Odobenus rosmarus</i>	Mammalia	Carnivora	Odobenida e	Odobenus	GCA_0003 21225.1	Scaffold	2,400,133, 628	2,616,778	89,951	41.50	2013/1/16
<i>Lontra canadensis</i>	Mammalia	Carnivora	Mustelidae	Lontra	GCA_0100 15895.1	Scaffold	2,405,748, 306	18,460,785	691,969	41.50	2020/1/27
<i>Zalophus californian us</i>	Mammalia	Carnivora	Otariidae	Zalophus	GCA_0097 62305.2	Chromoso me	2,409,685, 310	147,124,15 2	32,626,160	41.50	2020/7/14
<i>Balaenoptera physalus</i>	Mammalia	Cetartiodac tyla	Balaenopte ridae	Balaenoptera	GCA_0233 38255.1	Scaffold	2,410,646, 746	24,939,677	146,061	41.00	2022/5/11
<i>Mirounga leonina</i>	Mammalia	Carnivora	Phocidae	Mirounga	GCA_0118 00145.1	Scaffold	2,417,303, 335	54,232,831	197,952	41.50	2020/3/31
<i>Eumetopias jubatus</i>	Mammalia	Carnivora	Otariidae	Eumetopias	GCA_0040 28035.1	Scaffold	2,418,246, 527	14,018,600	242,372	41.50	2019/1/15
<i>Monachus schauinslan di</i>	Mammalia	Carnivora	Phocidae	Monachus	GCA_0022 01575.2	Chromoso me	2,419,460, 046	150,810,82 5	181,061	41.50	2021/10/20
<i>Lutra lutra</i>	Mammalia	Carnivora	Mustelidae	Lutra	GCA_9026 55055.2	Chromoso me	2,438,458, 878	149,004,80 7	30,403,456	41.50	2020/7/31
<i>Enhydra lutris</i>	Mammalia	Carnivora	Mustelidae	Enhydra	GCA_0022 88905.2	Scaffold	2,455,260, 235	38,751,465	244,548	41.50	2017/9/11
<i>Stenella coerulealba</i>	Mammalia	Cetartiodac tyla	Delphinida e	Stenella	GCA_9513 94445.1	Contig	2,485,529, 352	2,186,058	2,186,058	42.00	2023/5/13
<i>Caperea marginata</i>	Mammalia	Cetartiodac tyla	Neobalaeni dae	Caperea	GCA_0292 24305.1	Scaffold	2,510,959, 822	112,264	19,492	41.00	2023/3/16
<i>Grampus griseus</i>	Mammalia	Cetartiodac tyla	Delphinida e	Grampus	GCA_0286 46425.1	Chromoso me	2,536,411, 710	93,180,659	62,041	41.50	2023/2/16

<i>Lagenorhynchus albirostris</i>	Mammalia	Cetartiodactyla	Delphinidae	Lagenorhynchus	GCA_949774975.1	Chromosome	2,544,835,795	110,721,866	3,398,034	42.00	2023/4/8
<i>Kogia breviceps</i>	Mammalia	Cetartiodactyla	Kogiidae	Kogia	GCA_026419965.1	Chromosome	2,555,838,840	122,394,096	42,780,423	42.00	2022/11/27
<i>Delphinus delphis</i>	Mammalia	Cetartiodactyla	Delphinidae	Delphinus	GCA_949987515.1	Chromosome	2,663,533,373	107,080,983	3,627,551	42.00	2023/5/1
<i>Eubalaena japonica</i>	Mammalia	Cetartiodactyla	Balaenidae	Eubalaena	GCA_004363455.1	Scaffold	2,693,575,300	39,813	34,866	41.00	2019/3/22
<i>Phocoena phocoena</i>	Mammalia	Cetartiodactyla	Phocoenidae	Phocoena	GCA_004363495.1	Scaffold	2,701,010,810	115,969	89,111	42.00	2019/3/22
<i>Callorhinus ursinus</i>	Mammalia	Carnivora	Otariidae	Callorhinus	GCA_003265705.1	Scaffold	2,706,852,204	31,506,801	133,024	41.50	2018/6/26
<i>Balaenoptera acutorostrata</i>	Mammalia	Cetartiodactyla	Balaenopteridae	Balaenoptera	GCA_949987535.1	Chromosome	2,772,901,778	116,513,105	3,015,723	41.50	2023/5/1
<i>Hyperoodon ampullatus</i>	Mammalia	Cetartiodactyla	Ziphiidae	Hyperoodon	GCA_949752795.1	Chromosome	2,828,760,712	117,448,692	3,126,546	42.00	2023/4/8
<i>Eschrichtius robustus</i>	Mammalia	Cetartiodactyla	Eschrichtiidae	Eschrichtius	GCA_028021215.1	Chromosome	2,982,434,994	126,538,682	39,209,000	41.50	2023/1/31
<i>Trichechus manatus</i>	Mammalia	Sirenia	Trichechidae	Trichechus	GCA_000243295.1	Scaffold	3,103,791,524	14,442,683	37,750	40.50	2012/1/23
<i>Leptonychotes weddellii</i>	Mammalia	Carnivora	Phocidae	Leptonychotes	GCA_000349705.1	Scaffold	3,156,886,159	904,068	23,664	41.50	2013/3/29
<i>Dugong dugon</i>	Mammalia	Sirenia	Dugongidae	Dugong	GCA_030035585.1	Chromosome	3,158,488,110	140,670,990	57,632,671	40.50	2023/5/18
<i>Cerberus rynchops</i>	Reptilia	Squamata	Colubridae	Cerberus	GCA_032468195.1	Scaffold	1,263,809,361	72,084	3,368	39.00	2023/10/10
<i>Acrochordus granulatus</i>	Reptilia	Squamata	Acrochordidae	Acrochordus	GCA_032468055.1	Scaffold	1,324,550,020	156,840	7,023	38.00	2023/10/10
<i>Hydrophis melanocephalus</i>	Reptilia	Squamata	Elapidae	Hydrophis	GCA_004320005.1	Scaffold	1,402,639,853	59,810	7,411	39.00	2019/2/13
<i>Laticauda laticaudata</i>	Reptilia	Squamata	Elapidae	Laticauda	GCA_004320025.1	Scaffold	1,558,706,106	39,330	35,581	40.00	2019/1/9
<i>Emydocephalus ijimae</i>	Reptilia	Squamata	Elapidae	Emydocephalus	GCA_004319985.1	Scaffold	1,625,199,642	18,937	18,545	40.50	2019/1/9
<i>Hydrophis cyanocinctus</i>	Reptilia	Squamata	Elapidae	Hydrophis	GCA_019473425.1	Chromosome	1,980,712,740	264,245,889	18,985,681	40.5567	2021/8/9
<i>Laticauda colubrina</i>	Reptilia	Squamata	Elapidae	Laticauda	GCA_015471245.1	Scaffold	2,038,820,311	39,959,644	38,278	37.6	2020/11/4
<i>Crocodylus porosus</i>	Reptilia	Crocodylia	Crocodylidae	Crocodylus	GCA_001723895.1	Scaffold	2,049,519,336	84,437,661	34,073	44.00	2016/9/13
<i>Caretta caretta</i>	Reptilia	Testudines	Cheloniidae	Caretta	GCA_023653815.1	Chromosome	2,133,995,980	130,956,235	18,214,114	44.00	2022/6/7
<i>Chelonia mydas</i>	Reptilia	Testudines	Cheloniidae	Chelonia	GCA_015237465.2	Chromosome	2,134,358,617	134,428,053	39,415,510	44.00	2021/7/15
<i>Dermochelys coriacea</i>	Reptilia	Testudines	Dermochelyidae	Dermochelys	GCA_009764565.4	Chromosome	2,164,778,715	137,568,771	7,029,801	43.50	2021/8/25
<i>Malaclemys terrapin</i>	Reptilia	Testudines	Emydidae	Malaclemys	GCA_027887155.1	Chromosome	2,211,221,380	144,565,223	75,591,180	44.50	2023/1/18
<i>Eretmochelys imbricata</i>	Reptilia	Testudines	Cheloniidae	Eretmochelys	GCA_030012505.1	Chromosome	2,296,226,205	137,212,766	82,034,986	44.00	2023/5/15
<i>Alligator mississippiensis</i>	Reptilia	Crocodylia	Alligatoridae	Alligator	GCA_030867095.1	Chromosome	2,346,779,227	255,078,640	38,556,409	45.00	2023/8/24