

比较基因组学在生物进化研究中的应用

徐玉林¹, 李志强², 陈俐洁³, 陈兵⁴, 蔡颖^{3*}, 王璐^{5*}

¹峨眉山世界文化和自然遗产研究院, 四川 乐山

²峨眉山林业所, 四川 乐山

³乐山市林业科学研究院, 四川 乐山

⁴乐山市林业局, 四川 乐山

⁵乐山师范学院生命科学学院, 西南山地濒危鸟类保护四川省高等学校重点实验室, 四川 乐山

收稿日期: 2026年7月31日; 录用日期: 2026年8月26日; 发布日期: 2026年9月3日

摘要

生物进化研究长期关注物种起源、系统发育关系、演化历史及适应性性状形成机制。基因组学的发展使研究者能够从全基因组水平解析物种分化和性状演化的分子基础。随着高通量测序、长读长测序、Hi-C和光学图谱等技术不断成熟, 大量动植物基因组被测序、组装和注释, 为比较基因组学和进化基因组学研究提供了重要数据支撑。通过基因组测序、组装、注释及同源基因家族分析, 可揭示物种间基因组成、结构变异、基因家族扩张与收缩等差异; 基于全基因组数据构建系统发育关系, 有助于提高物种演化关系推断的可靠性; 结合选择压力分析, 还可筛选与环境适应和表型演化相关的候选基因。本文概述基因组测序技术的发展、基因组组装与注释方法, 并总结其在系统发育重建、同源基因家族分析和适应性进化研究中的应用, 以期为生物进化机制研究提供参考。

关键词

基因组学, 比较基因组学, 基因组测序, 基因组组装, 系统发育, 适应性进化

Applications of Comparative Genomics in Evolutionary Biology

Yulin Xu¹, Zhiqiang Li², Lijie Chen³, Bing Chen⁴, Ying Cai^{3*}, Lu Wang^{5*}

¹Institute of Mount Emei World Cultural and Natural Heritage, Leshan Sichuan

²Emeishan Forestry Institute, Leshan Sichuan

³Leshan Academy of Forestry Sciences, Leshan Sichuan

⁴Leshan Forestry Bureau, Leshan Sichuan

⁵Key Laboratory of Sichuan Province for the Conservation of Endangered Birds in Southwest Mountains, College of Life Sciences, Leshan Normal University, Leshan Sichuan

*通讯作者。

文章引用: 徐玉林, 李志强, 陈俐洁, 陈兵, 蔡颖, 王璐. 比较基因组学在生物进化研究中的应用[J]. 生物过程, 2026, 16(3): 110-119. DOI: 10.12677/bp.2026.163013

Abstract

Research on biological evolution has long focused on species origins, phylogenetic relationships, evolutionary history, and the mechanisms underlying the formation of adaptive traits. The development of genomics has enabled researchers to investigate the molecular basis of species divergence and trait evolution at the whole-genome level. With the continuous advancement of high-throughput sequencing, long-read sequencing, Hi-C, optical mapping, and other technologies, the genomes of numerous animals and plants have been sequenced, assembled, and annotated, providing important data support for comparative genomics and evolutionary genomics studies. Through genome sequencing, assembly, annotation, and homologous gene family analysis, differences among species in gene composition, structural variation, gene family expansion, and gene family contraction can be revealed. Phylogenetic reconstruction based on whole-genome data helps improve the reliability of inferences regarding evolutionary relationships among species. In combination with selection pressure analysis, candidate genes associated with environmental adaptation and phenotypic evolution can also be identified. This paper reviews the development of genome sequencing technologies, genome assembly and annotation methods, and summarizes their applications in phylogenetic reconstruction, homologous gene family analysis, and adaptive evolution research, with the aim of providing a reference for studies on the mechanisms of biological evolution.

Keywords

Genomics, Comparative Genomics, Genome Sequencing, Genome Assembly, Phylogeny, Adaptive Evolution

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

生物进化是生命科学领域长期关注的重要问题，研究内容涵盖系统发育关系、物种演化历史、适应性性状形成及其遗传机制等多个方面。基因组通常指生物体单倍体细胞中所含的全部遗传物质，包括染色体 DNA 序列及其携带的遗传信息。遗传信息通过转录和翻译参与蛋白质合成，并进一步影响个体发育、生理功能和表型特征。因此，解析基因组序列是理解生命活动和生物多样性形成机制的重要基础。基因组学作为一门以全基因组序列为研究对象、并通过实验技术与生物信息学方法挖掘遗传信息的新兴学科逐渐形成[1]。该领域通常以全基因组测序为基础，结合序列拼接、基因预测、功能注释和比较分析等方法，解析基因结构、功能调控、遗传变异及其进化规律[2]。

人类基因组计划开启(1990 年) [3]到人类基因组图谱发表(2001 年) [4]，基因组学研究进入快速发展阶段。随后，大肠杆菌、线虫、果蝇和小鼠等模式物种[5]，以及大量非模式动植物的基因组相继完成测序和组装，推动了医学领域[6]、动植物驯化[7][8]、珍稀濒危物种保护[9]、进化历史[10]等研究方向。其中，比较基因组学是基因组学的重要分支，主要通过比较不同物种或群体间的基因组结构、基因组成、功能元件和序列变异，揭示基因功能、调控机制及物种演化规律[1]。该领域推动了基因组结构变异、进

化历史重建、系统发育关系解析、基因家族演化和适应性进化等方向[11]。

比较基因组学研究通常包括基因组数据获取、序列质量控制、基因组组装与注释、同源基因识别、系统发育分析、基因家族演化分析和适应性进化分析等环节。其核心目标并非单纯获得基因组序列，而是通过跨物种或跨群体的比较，识别与物种分化、表型演化和环境适应相关的遗传变异及其功能意义。

2. 基因组获得的方法论框架

2.1. 基因组测序与质量控制

获得高质量基因组序列是比较基因组学研究的前提。目前常用的测序策略主要包括从头测序和全基因组重测序。从头测序不依赖参考基因组，适用于新物种或缺乏参考基因组的类群；重测序则以已有参考基因组为基础，主要用于不同个体、群体或品种之间的遗传变异分析。由于成本较低、效率较高，重测序在群体基因组学和适应性进化研究中应用广泛。

DNA 测序技术大致经历了三代发展。第一代测序技术以化学降解法[12]和双脱氧末段终止法(酶法或 Sanger 测序法)为代表[13]。例如，双脱氧末段终止法作为早期主流技术，在人类基因组计划[14]和噬菌体基因组拼接[15]中发挥了重要作用。但该技术通量低、成本高，难以满足大规模基因组研究需求[16]。随后，第二代高通量测序技术快速发展，代表平台包括 Roche 454、Illumina 和 ABI SOLiD 等[17]-[20]。二代测序显著提高了测序通量，降低了测序成本，推动了大规模生物基因组研究[20]-[32]。例如，人类基因组的提升[20]、大熊猫(*Ailuropda melanleuca*) [21]、狸藻(*Utricularia gibba*) [22]、野生稻(*Oryza brachyantha*) [23]、辣椒(*Capsicum annuum*) [24]等物种均借助二代测序技术完成了基因组测序和组装。然而，二代测序读长较短，且部分平台需进行 PCR 扩增，可能影响基因组复杂重复区域和结构变异区域的准确组装[19]。因此，三代高通量测序技术应运而生，其以单分子、高速度、长读长测序(高于二代测序 200 多倍)为主要特点，且不需 PCR 过程[33]。代表平台包括 Helicos BioSciences 公司的 tSMSTM 平台[34]；Life Technologies 公司的 FRET 技术[35]；Pacific Biosciences 公司的 SMRT [36]；Oxford Nanopore Nechnologies 公司的纳米孔单分子技术[37]。目前，长读长测序已广泛用于复杂基因组从头组装和旧版本基因组升级[38]-[43]，包括最连续的人类基因组的更新[38]。除常规测序平台外，BioNano 光学图谱、10X Genomics linked-read 测序和 Hi-C 技术也常用于辅助基因组组装[44]-[46]。这些技术与短读长、长读长测序结合，已成为构建高质量参考基因组的重要策略。

2.2. 基因组组装与质量评估

基因组组装是将测序获得的短读长或长读长序列拼接为连续序列，并尽可能还原真实染色体结构的过程，通常包括 contig 构建、scaffold 连接、序列纠错和染色体挂载等步骤[42]。组装算法和策略需要根据测序数据类型、基因组大小、杂合度、重复序列比例和倍性等因素选择[47]-[52]。此外，整合短读长、长读长、光学图谱和 Hi-C 数据的联合组装策略逐渐成为主流。例如，山羊基因组研究整合单细胞实时、Illumina、BioNano 和 Hi-C 测序数据，获得染色体水平基因组重复片段更长，更完整[42]。

组装质量评估通常从连续性、完整性和准确性三个方面进行。一般而言，N50 值越大，组装连续性越好[53]。但 N50 不能完全代表组装质量，还需结合测序读段回比对率、覆盖深度、转录组比对率等指标综合判断[53]。此外，BUSCO [54]和 CEGMA [55]等工具可通过检测保守单拷贝直系同源基因或核心真核基因完整性，评估基因组的完整程度。

2.3. 基因组注释

基因组注释主要包括重复序列注释、非编码 RNA 注释、蛋白编码基因预测和功能注释等[56]。

真核生物基因组中含有大量重复序列,主要包括串联重复序列和散在重复序列[57],其不仅影响基因组大小和结构,也可参与基因调控、基因复制和基因组重排,在物种演化中具有重要作用[58]。重复序列注释通常结合从头预测和同源预测[59][60]。非编码 RNA 通常也占基因组极小比例,通过 tRNAscan-SE 等软件实现预测[61]-[64]。蛋白编码基因注释通常结合从头预测、同源预测和转录组证据进行,以提高基因结构预测的准确性[65]-[67]。功能注释则依赖 KEGG、GO、InterPro 等数据库,用于推断基因可能参与的代谢通路、生物过程和蛋白结构域[68]-[70]。通过多数据库联合注释,可较为全面地揭示基因的潜在功能及其生物学意义。

3. 比较基因组学下游分析及在生物进化中的应用

3.1. 同源基因家族演化与性状分化

同源基因家族分析是揭示物种性状分化和功能创新的重要方法。祖先基因可能因复制、分化和物种分化事件形成多个具有相似序列或相关功能的基因,共同构成同源基因家族[71]。直系同源基因主要由物种分化产生,通常保留相似结构和功能;旁系同源基因则多由基因复制形成,经过长期演化后可能出现功能分化、新功能获得或功能丧失[72]。

基因家族扩张和收缩能够反映不同谱系在演化过程中功能需求的变化。扩张基因家族往往提示某类功能在特定谱系中得到强化;收缩基因家族则可能反映功能简化或生态适应过程中的基因丢失[73]。例如,牦牛基因组中与感觉感知、能量代谢和低氧应激相关的基因家族存在扩张或功能富集,这可能与其长期适应青藏高原低氧、寒冷和强紫外环境有关[28]。该研究并非仅获得牦牛参考基因组,而是通过与低海拔近缘家牛比较,识别高原环境压力下基因家族组成和功能模块的变化,从而揭示大型哺乳动物高海拔适应的基因组基础。而在植物的研究中,抗病相关基因家族是比较基因组学研究的经典对象。比较基因组研究表明, *NBS-LRR* 基因常以簇状形式分布于基因组中,并通过串联复制、片段复制、重组和基因转换等机制快速演化[74]。

诸如此类动植物研究均表明,性状分化和生态适应常常并非由单个基因变化决定,而是由基因复制、基因丢失、表达调控、水平基因转移和功能分化共同塑造。未来研究需要从“基因家族成员数比较”进一步转向“基因家族功能演化机制解析”,结合转录组、蛋白结构、代谢通路和实验验证,判断扩张或收缩基因家族是否真正参与适应性表型形成。

3.2. 系统发育关系重建与物种演化历史解析

系统发育关系重建是比较基因组学在生物进化研究中的核心内容之一。早期分子系统学常利用线粒体基因片段、核基因片段或线粒体基因组等分子标记推断物种间亲缘关系,但单个或少数分子标记所包含的信息量有限,且容易受到谱系不完全分选、基因渗入等因素影响,不同标记所得系统发育关系有时存在差异甚至冲突[75]。随着全基因组数据的快速积累,系统发育基因组学通过整合大量单拷贝直系同源基因、全基因组序列特征和稀有基因组变化,为复杂类群系统发育关系解析提供了更高分辨率的数据基础[76]-[78]。

基因组尺度的系统发育分析主要可分为基于序列的方法和基于非序列特征的方法[77]。基于序列的方法通常以多个物种共有的单拷贝直系同源基因为基础。串联法将所有直系同源基因序列比对后连接成超级矩阵,再利用最大似然法、贝叶斯法等方法构建系统发育树。并联法则先分别基于每个基因构建基因树,再通过整合多个基因树估计物种树[77]。在基于非序列的分析方法中,利用全基因组特征或稀有基因组变化构建系统发育关系。全基因组特征包括基因组成、基因顺序、低聚核苷酸频率和基因组结构特征等[77];稀有基因组变化则包括插入缺失、内含子得失、逆转座子插入、基因融合和基因分裂等[78]。

综合多种基因组信息,有助于提高系统发育关系推断的稳定性和可信度。

比较基因组学在系统发育重建中的价值不仅在于获得更稳定的物种树,还在于揭示系统发育冲突背后的演化过程。例如,地山雀长期因形态特征被认为与地鸦关系亲缘较近,但 Qu 等[26]通过地山雀基因组测序并结合近缘鸟类重测序数据发现,地山雀实际上应归入山雀科。这一案例说明,基因组尺度数据能够纠正仅依赖形态分类可能产生的误判。植物类群中,比较基因组学同样显著提升了系统发育和物种演化历史解析能力。以水稻属为例,Chen 等[23]通过野生稻基因组分析揭示了水稻属基因组演化过程中的染色体结构变化和转座元件动态,为理解栽培稻与野生近缘种之间的演化关系提供了重要依据。类似地,在辣椒基因组研究中,Kim 等[24]通过比较不同茄科植物基因组,分析了辣椒属特异性演化和辣味形成相关基因的演化,为理解作物特殊性状起源提供了系统发育背景。

总之,系统发育基因组学的核心进展已从单纯回答“物种之间谁与谁亲缘关系更近”,转向解释“不同基因组区域为何支持不同演化历史”。因此,未来系统发育基因组学研究需要结合物种树、基因树、群体基因组和生态证据,以更全面地揭示物种形成和演化历史。

3.3. 适应性进化与环境响应机制

适应性进化研究关注自然选择如何塑造基因组变异,并最终影响物种表型、生态位和分布格局[79]。比较基因组学为适应性进化研究提供了全基因组尺度的分析框架。通过比较不同物种间基因组差异,可以识别正选择基因、选择扫荡区域、趋同演化位点和与环境因子显著相关的遗传变异,从而推断环境压力与基因组演化之间的关系[79]-[81]。

在蛋白编码基因中,自然选择压力常通过非同义替换率与同义替换率的比值 ω 进行衡量[80],非同义替换可能改变氨基酸序列并影响蛋白功能,而同义替换通常不改变氨基酸序列。当 $\omega < 1$ 时,基因通常受到纯化选择,说明多数非同义突变不利于功能维持;当 $\omega = 1$ 时,基因可能接近中性演化;当 $\omega > 1$ 时,提示非同义替换积累速度高于同义替换,该基因可能受到正选择作用[81]。PAML 和 HyPhy 等软件常用于基于系统发育框架的选择压力分析[82]。

通过全基因组范围所鉴定的正选择基因,往往与物种适应新环境、形成特殊表型或增强特定生理功能有关。动物高海拔适应是适应性进化研究中的代表性案例。Qiu 等[28]发现牦牛中与低氧响应和能量代谢相关的基因或蛋白结构域存在选择信号和功能富集,提示能量代谢重塑可能是大型高原哺乳动物适应低氧环境的重要机制。Qu 等[26]对地山雀基因组的研究则发现,高海拔鸟类适应不仅涉及低氧响应,也可能与形态变化、能量代谢和生态行为转变相关。而在人类中,秘鲁安第斯高原人群 *PDE1B*、*PPP1R1A* 和 *RASGEF1B* 的基因组区域受到近期正选择,并与血红蛋白浓度相关,提示这些基因可能参与高海拔低氧适应[83]。综合这些研究可以看出,不同类群虽然面临相似低氧压力,但其适应性变异并不完全相同。这提示高海拔适应可能存在多条遗传路径,既包括共同的低氧响应机制,也包括群体特异性的生理调节方式。

植物适应性进化研究则更多关注驯化、逆境胁迫等。棉花基因组研究显示,棉花驯化和改良过程中受到选择的区域不仅与纤维长度、强度和产量相关,也涉及胁迫响应、细胞壁生物合成和激素信号调控等功能[84]。微生物适应性进化常更突出地表现为基因获得、基因丢失、同源重组和可变基因组组成变化。以大肠杆菌为例,比较基因组分析发现,该物种具有庞大的泛基因组和相对较小的核心基因组,并可重建大量基因获得与丢失事件;部分谱系中的代谢多样化具有重要适应意义,而基因组流动主要集中在染色体中少数保守位置,说明大肠杆菌适应性演化与有组织的可变基因组动态密切相关[85]。这与动植物相较于动植物中常通过蛋白编码基因正选择来解释适应性演化的模式存在一定差异,说明比较基因组学在不同生物类群中需要采用不同解释框架。

综上,适应性进化研究正在从寻找“单个候选基因”转向解析“多基因、多通路和多层次调控网络”。未来,适应性进化研究应进一步结合基因组、转录组、表观基因组、蛋白质组、代谢组和功能实验,从候选基因筛选走向因果机制验证。

4. 总结与展望

基因组学的发展推动了生物进化研究从单基因或少数分子标记层面转向全基因组层面。通过测序、组装、注释和比较分析,研究者能够更加系统地解析物种间基因组差异、基因家族演化、系统发育关系和适应性进化机制。未来,随基因组学研究将进一步向高连续性、高准确性和多层次整合方向发展。泛基因组和图基因组研究将突破单一参考基因组的限制,更全面地展示物种内部遗传多样性和结构变异;单细胞组学、表观基因组学、转录组学、蛋白质组学和代谢组学等多组学整合,将有助于从基因结构、表达调控和功能实现等多个层面解析复杂性状演化机制。同时,基因组学在濒危物种保护、资源生物开发、生态适应研究和生物多样性保护中的作用也将更加突出。总体而言,基因组学不仅为进化生物学提供了强有力的数据和方法支撑,也将继续促进生命科学对物种起源、适应和多样化机制的深入理解。

基金项目

感谢以下项目资助:大学生创新创业训练项目(202510649003);乐山市科学技术局项目(24YYJC0018);乐山师范学院科研平台基金(203250087; 205240070)。

参考文献

- [1] 李伟, 印莉萍. 基因组学相关概念及其研究进展[J]. 生物学通报, 2000, 35(11): 1-3.
- [2] Chen, R. (2021) Early Bioinformatics Research in China. *Quantitative Biology*, **9**, 242-250. <https://doi.org/10.15302/j-qb-021-0255>
- [3] Watson, J.D. (1990) The Human Genome Project: Past, Present, and Future. *Science*, **248**, 44-49. <https://doi.org/10.1126/science.2181665>
- [4] 赵晋平, 徐平丽, 孟静静, 等. 从结构基因组学到功能基因组学[J]. 生命科学研究, 2006, 10(2): 57-61.
- [5] 吴学军, 柴建华. 比较基因组学和人类基因组研究[J]. 生物工程进展, 2000, 20(1): 57-59.
- [6] Schon, K.R., Horvath, R., Wei, W., Calabrese, C., Tucci, A., Ibañez, K., *et al.* (2021) Use of Whole Genome Sequencing to Determine Genetic Basis of Suspected Mitochondrial Disorders: Cohort Study. *BMJ*, **375**, e066288. <https://doi.org/10.1136/bmj-2021-066288>
- [7] Carneiro, M., Rubin, C., Di Palma, F., Albert, F.W., Alföldi, J., Barrio, A.M., *et al.* (2014) Rabbit Genome Analysis Reveals a Polygenic Basis for Phenotypic Change during Domestication. *Science*, **345**, 1074-1079. <https://doi.org/10.1126/science.1253714>
- [8] Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., *et al.* (2020) The Origin of Domestication Genes in Goats. *Science Advances*, **6**, eaaz5216. <https://doi.org/10.1126/sciadv.aaz5216>
- [9] Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., *et al.* (2022) The Era of Reference Genomes in Conservation Genomics. *Trends in Ecology & Evolution*, **37**, 197-202. <https://doi.org/10.1016/j.tree.2021.11.008>
- [10] Sands, T.R. (2019) Evolutionary Genomics: The Fruits of Genomic Approaches Applied to Evolutionary Biology. *Genome Biology*, **20**, Article No. 10. <https://doi.org/10.1186/s13059-018-1615-x>
- [11] Das, S. and Hirano, M. (2012) Comparative Genomics and Genome Evolution. *Current Genomics*, **13**, 85.
- [12] Maxam, A.M. and Gilbert, W. (1977) A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences*, **74**, 560-564. <https://doi.org/10.1073/pnas.74.2.560>
- [13] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences*, **74**, 5463-5467. <https://doi.org/10.1073/pnas.74.12.5463>
- [14] Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., *et al.* (2001) The Sequence of the Human Genome. *Science*, **291**, 1304-1351. <https://doi.org/10.1126/science.1058040>

- [15] Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., *et al.* (1977) Nucleotide Sequence of Bacteriophage ϕ X174 DNA. *Nature*, **265**, 687-695. <https://doi.org/10.1038/265687a0>
- [16] 田李, 张颖, 赵云峰. 新一代测序技术的发展和应[J]. 生物技术报, 2015, 31(11): 1-8.
- [17] Eisenstein, M. (2012) The Battle for Sequencing Supremacy. *Nature Biotechnology*, **30**, 1023-1026. <https://doi.org/10.1038/nbt.2412>
- [18] 杨晓玲, 施华, 唐恬. 新一代测序技术的发展和应前景[J]. 生物技术通报, 2010(10): 76-81.
- [19] Goodwin, S., McPherson, J.D. and McCombie, W.R. (2016) Coming of Age: Ten Years of Next-Generation Sequencing Technologies. *Nature Reviews Genetics*, **17**, 333-351. <https://doi.org/10.1038/nrg.2016.49>
- [20] Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., *et al.* (2008) The Complete Genome of an Individual by Massively Parallel DNA Sequencing. *Nature*, **452**, 872-876. <https://doi.org/10.1038/nature06884>
- [21] Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., *et al.* (2010) The Sequence and De Novo Assembly of the Giant Panda Genome. *Nature*, **463**, 311-317. <https://doi.org/10.1038/nature08696>
- [22] Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C.A., Carretero-Paulet, L., Chang, T., *et al.* (2013) Architecture and Evolution of a Minute Plant Genome. *Nature*, **498**, 94-98. <https://doi.org/10.1038/nature12132>
- [23] Chen, J., Huang, Q., Gao, D., Wang, J., Lang, Y., Liu, T., *et al.* (2013) Whole-Genome Sequencing of *Oryza Brachyantha* Reveals Mechanisms Underlying *Oryza* Genome Evolution. *Nature Communications*, **4**, Article No. 1595. <https://doi.org/10.1038/ncomms2596>
- [24] Kim, S., Park, M., Yeom, S., Kim, Y., Lee, J.M., Lee, H., *et al.* (2014) Genome Sequence of the Hot Pepper Provides Insights into the Evolution of Pungency in *Capsicum* Species. *Nature Genetics*, **46**, 270-278. <https://doi.org/10.1038/ng.2877>
- [25] Wu, C., Zhang, D., Kan, M., Lv, Z., Zhu, A., Su, Y., *et al.* (2014) The Draft Genome of the Large Yellow Croaker Reveals Well-Developed Innate Immunity. *Nature Communications*, **5**, Article No. 5227. <https://doi.org/10.1038/ncomms6227>
- [26] Qu, Y., Zhao, H., Han, N., Zhou, G., Song, G., Gao, B., *et al.* (2013) *Nature Communications*, **4**, Article No. 2071. <https://doi.org/10.1038/ncomms3071>
- [27] Cho, Y.S., Hu, L., Hou, H., Lee, H., Xu, J., Kwon, S., *et al.* (2013) The Tiger Genome and Comparative Analysis with Lion and Snow Leopard Genomes. *Nature Communications*, **4**, Article No. 2433. <https://doi.org/10.1038/ncomms3433>
- [28] Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., *et al.* (2012) The Yak Genome and Adaptation to Life at High Altitude. *Nature Genetics*, **44**, 946-949. <https://doi.org/10.1038/ng.2343>
- [29] Laing, R., Kikuchi, T., Martinelli, A., Tsai, I.J., Beech, R.N., Redman, E., *et al.* (2013) The Genome and Transcriptome of *Haemonchus contortus*, a Key Model Parasite for Drug and Vaccine Discovery. *Genome Biology*, **14**, R88. <https://doi.org/10.1186/gb-2013-14-8-r88>
- [30] Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A., Moreland, R.T., Simmons, D.K., *et al.* (2013) The Genome of the Ctenophore *Mnemiopsis leidyi* and Its Implications for Cell Type Evolution. *Science*, **342**, Article ID: 1242592. <https://doi.org/10.1126/science.1242592>
- [31] Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., *et al.* (2014) The Locust Genome Provides Insight into Swarm Formation and Long-Distance Flight. *Nature Communications*, **5**, Article No. 2957. <https://doi.org/10.1038/ncomms3957>
- [32] Park, D., Jung, J.W., Choi, B., Jayakodi, M., Lee, J., Lim, J., *et al.* (2015) Uncovering the Novel Characteristics of Asian Honey Bee, *Apis Cerana*, by Whole Genome Sequencing. *BMC Genomics*, **16**, Article No. 1. <https://doi.org/10.1186/1471-2164-16-1>
- [33] 张得芳, 马秋月, 尹佟明, 等. 第三代测序技术及其应[J]. 中国生物工程杂志, 2013, 33(5): 125-131.
- [34] Bowers, J., Mitchell, J., Beer, E., Buzby, P.R., Causey, M., Efcavitch, J.W., *et al.* (2009) Virtual Terminator Nucleotides for Next-Generation DNA Sequencing. *Nature Methods*, **6**, 593-595. <https://doi.org/10.1038/nmeth.1354>
- [35] Roy, R., Hohng, S. and Ha, T. (2008) A Practical Guide to Single-Molecule FRET. *Nature Methods*, **5**, 507-516. <https://doi.org/10.1038/nmeth.1208>
- [36] Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P. and Barron, A.E. (2011) Landscape of Next-Generation Sequencing Technologies. *Analytical Chemistry*, **83**, 4327-4341. <https://doi.org/10.1021/ac2010857>
- [37] Clarke, J., Wu, H., Jayasinghe, L., Patel, A., Reid, S. and Bayley, H. (2009) Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing. *Nature Nanotechnology*, **4**, 265-270. <https://doi.org/10.1038/nnano.2009.12>
- [38] Seo, J., Rhie, A., Kim, J., Lee, S., Sohn, M., Kim, C., *et al.* (2016) De Novo Assembly and Phasing of a Korean Human Genome. *Nature*, **538**, 243-247. <https://doi.org/10.1038/nature20098>

- [39] Badouin, H., Gouzy, J., Grassa, C.J., Murat, F., Staton, S.E., Cottret, L., *et al.* (2017) The Sunflower Genome Provides Insights into Oil Metabolism, Flowering and Asterid Evolution. *Nature*, **546**, 148-152. <https://doi.org/10.1038/nature22380>
- [40] Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B., *et al.* (2017) Improved Maize Reference Genome with Single-Molecule Technologies. *Nature*, **546**, 524-527. <https://doi.org/10.1038/nature22971>
- [41] Wang, K., Wang, J., Zhu, C., Yang, L., Ren, Y., Ruan, J., *et al.* (2021) African Lungfish Genome Sheds Light on the Vertebrate Water-to-Land Transition. *Cell*, **184**, 1362-1376.e18. <https://doi.org/10.1016/j.cell.2021.01.047>
- [42] Bickhart, D.M., Rosen, B.D., Koren, S., Sayre, B.L., Hastie, A.R., Chan, S., *et al.* (2017) Single-Molecule Sequencing and Chromatin Conformation Capture Enable De Novo Reference Assembly of the Domestic Goat Genome. *Nature Genetics*, **49**, 643-650. <https://doi.org/10.1038/ng.3802>
- [43] Larsen, P.A., Harris, R.A., Liu, Y., Murali, S.C., Campbell, C.R., Brown, A.D., *et al.* (2017) Hybrid De Novo Genome Assembly and Centromere Characterization of the Gray Mouse Lemur (*Microcebus murinus*). *BMC Biology*, **15**, Article No. 110. <https://doi.org/10.1186/s12915-017-0439-6>
- [44] Barseghyan, H., Tang, W., Wang, R.T., Almalvez, M., Segura, E., Bramble, M.S., *et al.* (2017) Next-Generation Mapping: A Novel Approach for Detection of Pathogenic Structural Variants with a Potential Utility in Clinical Diagnosis. *Genome Medicine*, **9**, Article No. 90. <https://doi.org/10.1186/s13073-017-0479-0>
- [45] Mostovoy, Y., Levy-Sakin, M., Lam, J., Lam, E.T., Hastie, A.R., Marks, P., *et al.* (2016) A Hybrid Approach for De Novo Human Genome Sequence Assembly and Phasing. *Nature Methods*, **13**, 587-590. <https://doi.org/10.1038/nmeth.3865>
- [46] Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., *et al.* (2017) De Novo Assembly of the *Aedes aegypti* Genome Using Hi-C Yields Chromosome-Length Scaffolds. *Science*, **356**, 92-95. <https://doi.org/10.1126/science.aal3327>
- [47] Roberts, M., Zimin, A.V., Hayes, W., Hunt, B.R., Ustun, C., White, J.R., *et al.* (2008) Improving Phrap-Based Assembly of the Rat Using “Reliable” Overlaps. *PLOS ONE*, **3**, e1836. <https://doi.org/10.1371/journal.pone.0001836>
- [48] He, Y., Zhang, Z., Peng, X., Wu, F. and Wang, J. (2013) De Novo Assembly Methods for Next Generation Sequencing Data. *Tsinghua Science and Technology*, **18**, 500-514. <https://doi.org/10.1109/tst.2013.6616523>
- [49] Ye, C., Hill, C.M., Wu, S., Ruan, J. and Ma, Z. (2016) DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Scientific Reports*, **6**, Article No. 31900. <https://doi.org/10.1038/srep31900>
- [50] Vollger, M.R., Logsdon, G.A., Audano, P.A., Sulovari, A., Porubsky, D., Peluso, P., *et al.* (2020) Improved Assembly and Variant Detection of a Haploid Human Genome Using Single-Molecule, High-Fidelity Long Reads. *Annals of Human Genetics*, **84**, 125-140. <https://doi.org/10.1111/ahg.12364>
- [51] Zimin, A.V. and Salzberg, S.L. (2020) The Genome Polishing Tool POLCA Makes Fast and Accurate Corrections in Genome Assemblies. *PLOS Computational Biology*, **16**, e1007981. <https://doi.org/10.1371/journal.pcbi.1007981>
- [52] Zhang, X., Zhang, S., Zhao, Q., Ming, R. and Tang, H. (2019) Assembly of Allele-Aware, Chromosomal-Scale Autopolyploid Genomes Based on Hi-C Data. *Nature Plants*, **5**, 833-845. <https://doi.org/10.1038/s41477-019-0487-8>
- [53] Meader, S., Hillier, L.W., Locke, D., Ponting, C.P. and Lunter, G. (2010) Genome Assembly Quality: Assessment and Improvement Using the Neutral Indel Model. *Genome Research*, **20**, 675-684. <https://doi.org/10.1101/gr.096966.109>
- [54] Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics*, **31**, 3210-3212. <https://doi.org/10.1093/bioinformatics/btv351>
- [55] Parra, G., Bradnam, K. and Korf, I. (2007) CEGMA: A Pipeline to Accurately Annotate Core Genes in Eukaryotic Genomes. *Bioinformatics*, **23**, 1061-1067. <https://doi.org/10.1093/bioinformatics/btm071>
- [56] Stein, L. (2001) Genome Annotation: From Sequence to Biology. *Nature Reviews Genetics*, **2**, 493-503. <https://doi.org/10.1038/35080529>
- [57] Lu, J.Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., *et al.* (2020) Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Reports*, **30**, 3296-3311.e5. <https://doi.org/10.1016/j.celrep.2020.02.048>
- [58] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., *et al.* (2001) Initial Sequencing and Analysis of the Human Genome. *Nature*, **409**, 860-921. <https://doi.org/10.1038/35057062>
- [59] Price, A.L., Jones, N.C. and Pevzner, P.A. (2005) De Novo Identification of Repeat Families in Large Genomes. *Bioinformatics*, **21**, i351-i358. <https://doi.org/10.1093/bioinformatics/bti1018>
- [60] Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J. (2005) Repbase Update, a Database of Eukaryotic Repetitive Elements. *Cytogenetic and Genome Research*, **110**, 462-467. <https://doi.org/10.1159/000084979>

- [61] Palazzo, A.F. and Lee, E.S. (2015) Non-Coding RNA: What Is Functional and What Is Junk? *Frontiers in Genetics*, **6**, 1-11. <https://doi.org/10.3389/fgene.2015.00002>
- [62] Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*, **25**, 955-964. <https://doi.org/10.1093/nar/25.5.955>
- [63] Nawrocki, E.P., Kolbe, D.L. and Eddy, S.R. (2009) Infernal 1.0: Inference of RNA Alignments. *Bioinformatics*, **25**, 1713-1713. <https://doi.org/10.1093/bioinformatics/btp326>
- [64] Shukla, V., Varghese, V.K., Kabekkodu, S.P., Mallya, S. and Satyamoorthy, K. (2017) A Compilation of Web-Based Research Tools for miRNA Analysis. *Briefings in Functional Genomics*, **16**, 249-273. <https://doi.org/10.1093/bfgp/ew042>
- [65] Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O. and Grau, J. (2018) Combining RNA-seq Data and Homology-Based Gene Prediction for Plants, Animals and Fungi. *BMC Bioinformatics*, **19**, Article No. 189. <https://doi.org/10.1186/s12859-018-2203-5>
- [66] Goodswen, S.J., Kennedy, P.J. and Ellis, J.T. (2012) Evaluating High-Throughput *Ab Initio* Gene Finders to Discover Proteins Encoded in Eukaryotic Pathogen Genomes Missed by Laboratory Techniques. *PLOS ONE*, **7**, e50609. <https://doi.org/10.1371/journal.pone.0050609>
- [67] Humann, J.L., Lee, T., Ficklin, S. and Main, D. (2019) Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. In: Kollmar, M., Ed., *Gene Prediction: Methods and Protocols, Methods in Molecular Biology*, Springer, 29-51. https://doi.org/10.1007/978-1-4939-9173-0_3
- [68] Kanehisa, M. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **28**, 27-30. <https://doi.org/10.1093/nar/28.1.27>
- [69] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al. (2000) Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, **25**, 25-29. <https://doi.org/10.1038/75556>
- [70] Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., et al. (2019) Interpro in 2019: Improving Coverage, Classification and Access to Protein Sequence Annotations. *Nucleic Acids Research*, **47**, D351-D360. <https://doi.org/10.1093/nar/gky1100>
- [71] Harris, R.M. and Hofmann, H.A. (2015) Seeing Is Believing: Dynamic Evolution of Gene Families. *Proceedings of the National Academy of Sciences*, **112**, 1252-1253. <https://doi.org/10.1073/pnas.1423685112>
- [72] Hu, X. and Friedberg, I. (2019) SwiftOrtho: A Fast, Memory-Efficient, Multiple Genome Orthology Classifier. *Giga-Science*, **8**, giz118. <https://doi.org/10.1093/gigascience/giz118>
- [73] Prachumwat, A. and Li, W. (2008) Gene Number Expansion and Contraction in Vertebrate Genomes with Respect to Invertebrate Genomes. *Genome Research*, **18**, 221-232. <https://doi.org/10.1101/gr.7046608>
- [74] Leister, D. (2004) Tandem and Segmental Gene Duplication and Recombination in the Evolution of Plant Disease Resistance Genes. *Trends in Genetics*, **20**, 116-122. <https://doi.org/10.1016/j.tig.2004.01.007>
- [75] Yu, L. (2006) Phylogenomics—An Attractive Avenue to Reconstruct “Tree of Life”. *Hereditas*, **11**, Article No. 1445. <https://doi.org/10.1360/yc-006-1445>
- [76] Thornton, J.W. and DeSalle, R. (2000) Gene Family Evolution and Homology: Genomics Meets Phylogenetics. *Annual Review of Genomics and Human Genetics*, **1**, 41-73. <https://doi.org/10.1146/annurev.genom.1.1.41>
- [77] Delsuc, F., Brinkmann, H. and Philippe, H. (2005) Phylogenomics and the Reconstruction of the Tree of Life. *Nature Reviews Genetics*, **6**, 361-375. <https://doi.org/10.1038/nrg1603>
- [78] Rokas, A. and Holland, P.W.H. (2000) Rare Genomic Changes as a Tool for Phylogenetics. *Trends in Ecology & Evolution*, **15**, 454-459. [https://doi.org/10.1016/s0169-5347\(00\)01967-4](https://doi.org/10.1016/s0169-5347(00)01967-4)
- [79] Fan, Y. and Fu, X. (2017) The Complete Mitochondrial Genome of the Firefly, *Pteroptyx maipo* (Coleoptera: Lampyridae). *Mitochondrial DNA Part B*, **2**, 795-796. <https://doi.org/10.1080/23802359.2017.1398598>
- [80] Yang, Z. (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, **24**, 1586-1591. <https://doi.org/10.1093/molbev/msm088>
- [81] Yuan, M.L., Zhang, Q.L., Zhang, L., Jia, C., Li, X., Yang, X., et al. (2018) *Molecular Phylogenetics and Evolution*, **122**, 116-124. <https://doi.org/10.1016/j.ympev.2018.01.016>
- [82] Pond, S.L.K., Frost, S.D.W. and Muse, S.V. (2005) HyPhy: Hypothesis Testing Using Phylogenies. *Bioinformatics*, **21**, 676-679. <https://doi.org/10.1093/bioinformatics/bti079>
- [83] Zhu, K., Jorgensen, K.C., Weinstein, J., Kiyamu, M., Elías, G., Isherwood, J.L., et al. (2026) Genomic Evidence for Natural Selection Underlying High-Altitude Adaptive Hemoglobin Levels among Peruvian Andeans. *Genome Biology and Evolution*, **18**, evag164. <https://doi.org/10.1093/gbe/evag164>
- [84] Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., et al. (2017) Genomic Analyses in Cotton Identify Signatures of

Selection and Loci Associated with Fiber Quality and Yield Traits. *Nature Genetics*, **49**, 1089-1098.
<https://doi.org/10.1038/ng.3887>

- [85] Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., *et al.* (2009) Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. *PLOS Genetics*, **5**, e1000344.
<https://doi.org/10.1371/journal.pgen.1000344>