



群体基因组学前沿探索: 发展、技术革新与跨学科融合

陈俊豪, 张林, 陆剑*

北京大学生命科学学院, 基因功能研究与操控全国重点实验室, 北京 100871

* 联系人, Email: luj@pku.edu.cn

收稿日期: 2025-01-08; 接受日期: 2025-03-08; 网络版发表日期: 2025-04-17

国家自然科学基金(批准号: 32070597)、国家重点研发计划政府间国际科技创新合作专项(批准号: 2022YFE0132000)资助

摘要 群体基因组学通过结合高通量测序技术和统计方法, 从基因组层面研究群体内和群体间遗传变异. 通过探讨包括突变、重组、遗传漂变、自然选择和迁移在内的群体演化过程, 群体基因组学揭示基因组变异的来源及其在进化和适应中的作用. 在过去的20年中, 随着数据规模的快速扩张和分析方法的不断迭代, 群体基因组学经历飞速发展, 取得很多重要成果, 包括揭示人类迁徙历史、自然群体适应、农业物种驯化的遗传基础以及病原体的进化轨迹等. 这些研究极大地加深了人们对物种演化历史和环境适应机制的理解. 本文简要回顾群体基因组学的发展历程与主要成就, 并探讨新时代群体基因组学所面临的机遇与挑战.

关键词 群体基因组学, 遗传多样性, 演化历史, 适应

群体遗传学(population genetics)是演化生物学的重要分支, 主要研究突变、重组、遗传漂变、自然选择及迁移等因素对群体的遗传组成和变化的影响. 群体遗传学的基本理论框架发源自20世纪初. 1908年, 英国数学家Godfrey Harold Hardy和德国医生Wilhelm Weinberg分别提出遗传平衡定律, 建立理想群体中等位基因和基因型频率变化的零假设. 20世纪20~30年代, Ronald Fisher, John Burdon Sanderson Haldane和Sewall Wright将达尔文进化理论和孟德尔遗传学定律结合, 为群体遗传学的理论框架和数学计算奠定基础, 被称为群体遗传学的“三驾马车”. 1953年, DNA双螺旋结构的发现标志着分子生物学时代的到来, 促进人们对遗传密码的认识. 1966年, Lewontin和Hubby^[1,2]通过蛋

白质凝胶电泳第一次揭示拟暗果蝇(*Drosophila pseudoobscura*)自然群体中存在大量的多态性, 开创分子层面研究遗传变异的先河. 在逐渐积累的蛋白序列的基础上, 20世纪60~70年代, Kimura^[3]提出的分子演化的中性理论和Ohta^[4]在此基础上提出的近中性理论, 进一步完善群体遗传学的理论框架. 1982年, Kingman^[5]提出溯祖理论(coalescent theory)的数学模型, 为研究群体演化提供强有力的工具. 1983年, Kreitman^[6]完成第一个群体水平的DNA序列多态性分析, 并开发McDonald-Kreitman test方法检测到果蝇中*Adh*基因的正选择信号.

在群体遗传学发展的早期, 关于自然群体多态性的数据非常有限, 只有例如微卫星长度多态性、限制

引用格式: 陈俊豪, 张林, 陆剑. 群体基因组学前沿探索: 发展、技术革新与跨学科融合. 中国科学: 生命科学, 2025, 55: 854–870

Chen J H, Zhang L, Lu J. Frontiers in population genomics: development, technological innovation, and interdisciplinary integration (in Chinese). Sci Sin Vitae, 2025, 55: 854–870, doi: [10.1360/SSV-2024-0311](https://doi.org/10.1360/SSV-2024-0311)

性片段长度多态性(restriction fragment length polymorphism, RFLP)、随机扩增多态性DNA (random amplified polymorphic DNA, RAPD)等分子标记, 以及少量的线粒体基因和核基因序列, 缺乏基因组水平的多态性信息, 很多理论和假说难以得到真实数据的检验和支撑. 在过去二十多年中, 伴随着高通量测序技术的发展和数据分析方法的不断迭代, 群体基因组学(population genomics)应运而生, 并经历飞速发展. 群体基因组学是对群体遗传学的延伸与拓展, 随着测序成本的急剧下降, 对大量个体的数百万个多态性位点进行基因分型成为可能, 使研究者能够更全面准确地解析群体内或群体间的遗传变异及演化过程. 结合新出现的其他多组学技术, 群体基因组学可以在全基因组水平研究遗传变异与基因表达、表观遗传状态、蛋白质功能等的关系. 伴随着组学数据的积累, 新的计算方法、统计模型和分析软件也被开发出来, 这使得群体基因组学在回答长期存在的经典科学问题或者新问题时拥有前所未有的效力和准确性, 极大地促进研究者对进化

生物学、保护生物学、生态学、农学以及人类健康的理解^[7]. 在本文中, 将简要回顾群体遗传学和群体基因组学的发展和标志性事件(图1), 并讨论新时代的群体基因组学的机遇和挑战.

1 人类基因组计划及人类群体基因组学

2001年, 人类基因组计划(Human Genome Project, HGP)发表第一个人类基因组草图^[8,9], 以30亿美元的预算初步完成约30亿碱基的测定, 标志着人类进入基因组时代. 虽然当时的草图并不完整, 只覆盖约90%的人类基因组, 但仍为破译人类遗传信息提供前所未有的洞见. 对基因组草图的分析产生令人惊讶的发现, 人类基因组中的蛋白质编码基因数目只有20000-25000个, 远少于此前的估计; 而且基因组中编码蛋白质的序列大约只占2%, 而转座子等重复序列占到将近一半. 这个堪比阿波罗计划和曼哈顿计划的项目是生物学中重要的里程碑, 同时也刺激测序技术的进步, 为群体基因

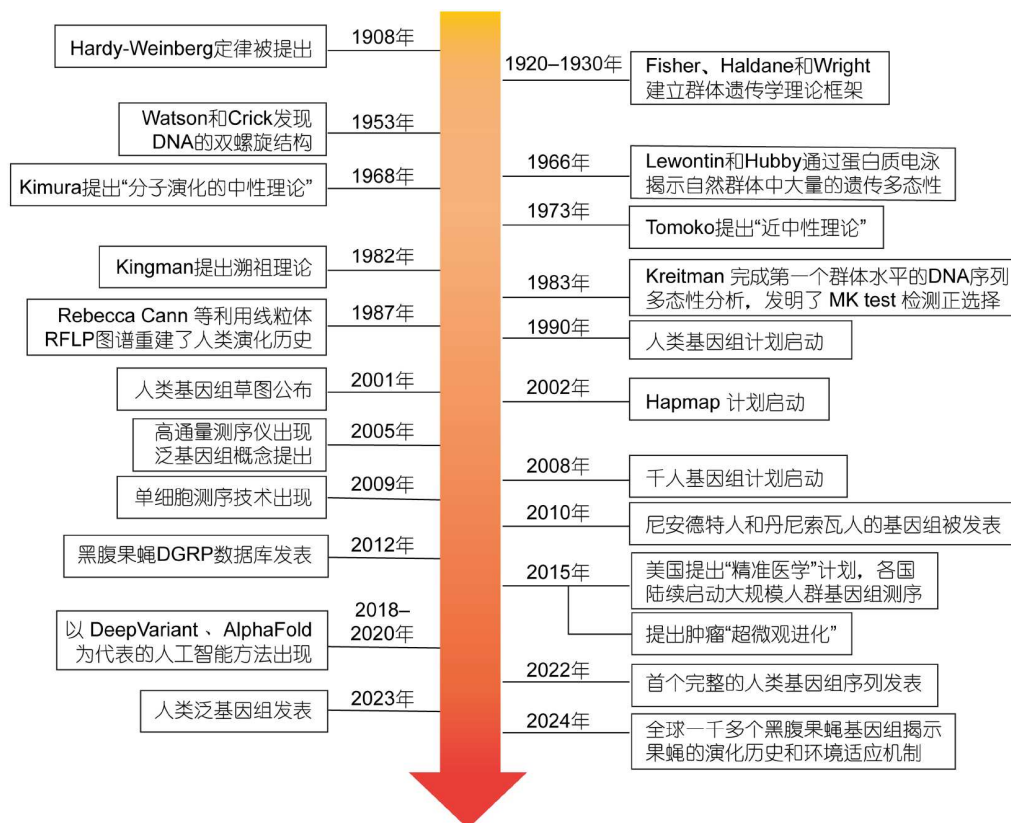


图1 群体遗传学和群体基因组学发展的标志性事件

Figure 1 Milestones in the development of population genetics and population genomics

组学的加速发展奠定基础^[10]。

人类基因组草图发布后, 人们很快认识到一个参考基因组无法代表人类群体的遗传多样性。人类个体间的基因组序列差异主要包括单核苷酸变异(single nucleotide variants, SNV)、插入或缺失(indels)和结构变异(structural variants, SVs), 其中SNV发生的概率不足千分之一。然而, 这些微小但多样的变异共同塑造人类的遗传多样性。2002年, 多国科学家发起国际人类基因组单体型图计划(HapMap计划)。HapMap的前两期研究主要通过Sanger测序和基因芯片对非、亚、欧裔的269个DNA样品进行基因型检测, 建立人类基因组中常见(等位基因频率大于5%)SNV的分布和连锁模式^[11,12]。HapMap第三期研究进一步扩大人群样本, 包括全球11个人群的1184个个体, 并补充对低频SNV的检测^[13]。从2005年开始, 高通量测序技术开始出现并蓬勃发展, Roche/454, Illumina等新一代的测序系统被推出, 极大地提高测序的通量。2008年, 中英美德等国科学家又启动国际千人基因组计划(1000 Genomes Project, 1kGP), 这一项目计划通过全基因组测序绘制非常详尽的全球人类基因组遗传多样性图谱^[14]。该项目共三期, 最终在2015年发表来自全球26个群体的2504个基因组(涵盖非洲、欧洲、亚洲和美洲), 同时还发表了包括约8500万个SNV、360万个短的插入缺失和6万个结构变异的人群多态性数据^[15,16]。1kGP是第一个真正意义上的人群全基因组测序计划, 虽然大部分的样品测序深度不高, 平均约7.4×, 但样本规模大, 且具有很好的代表性。这一项目发现了绝大多数人类基因组中等位基因频率大于1%的多态性位点, 对后续的人类群体结构和演化历史研究、表型的关联位点鉴定、疾病的个性化治疗都有重要的参考作用。同时, 1kGP将人类群体基因组研究的国际合作推向一个新的高峰, 该项目不但将基因组数据开放给公众, 还允许将样品的淋巴细胞系分享给研究者, 很大程度上促进1kGP资源的有效利用。2022年的一项研究中, 人类基因组结构变异联盟(Human Genome Structural Variation Consortium, HGSVC)联合美国多家单位利用这些细胞系资源将1kGP的数据扩展到3202个样本, 包括602个亲子三联体, 使测序深度达到约30×, 并用机器学习方法整合多个变异检测软件的结果提升罕见变异和结构变异的检测能力^[17]。除1kGP测序的26个较大的族群, 很多小群体中可能保留有丰富的遗传多样性。

2016年, 西蒙斯基因组多样性项目(Simons Genome Diversity Project)发表300个高质量基因组, 这些样本来自遗传、地理、语言和文化都比较独特的142个小族群, 补充现有人类基因组的多样性^[18]。

2 飞速增长的群体基因组数据

在精准医学和测序成本飞速下降的大背景下, 各国超大规模的人群基因组测序项目雨后春笋般地涌现出来, 这些项目的主要目的是揭示人群中的变异与疾病的关联, 并指导个性化医疗。已有近50个国家宣布或启动国家级群体基因组项目, 冰岛、英国、美国等国家已经在大人群众体的基因组研究方面建立起大规模的数据库和高效的数据共享、分析平台, 并取得显著的成果。多个国家的研究者还合作建立基因组聚合数据库(Genome Aggregation Database, gnomAD)用以整合、分享各种大型测序项目中的外显子组和基因组数据^[19]。

deCODE Genetics公司较早地对冰岛人开展大规模基因组测序, 并取得一系列重要成果。2015年, 他们发表2636个冰岛人的全基因组测序结果以及104220名冰岛人的Illumina芯片数据^[20], 分析天然存在的基因敲除变异对人类健康的影响^[21], 找到与阿尔茨海默病风险相关的罕见变异^[22], 估计Y染色体的点突变频率^[23]。根据亲子三联体, 他们估计后代的新生变异的数目以及亲本年龄的影响, 为Haldane提出的雄性驱动的演化理论(male-driven evolution)提供数据支撑^[24,25]。2020年, 他们报道27566个冰岛人的全基因组测序结果, 已覆盖冰岛约10%的人口^[26]。

英国生物银行(UK Biobank)^[27]是英国规模最大的有关致病或预防疾病的基因和环境因子的信息资源库。目的是探究特定基因、生活方式和健康状况之间的关系, 提高对致病基因的理解。2006年至2010年, UK Biobank共收集英国各地40~69岁之间50万名志愿者的数据信息, 包括基因信息和血液样本、生活方式及环境暴露数据。此外, 还收集志愿者超过10000项生物指标, 如血压、认知功能、饮食和骨密度等。UK Biobank在2023年11月公布近50万人的全基因组序列数据集, 并通过云平台向全世界的申请者开放。截至2023年末, 注册的研究者已经超过38000人, 其中80%以上来自英国以外, 用该数据库发表的学术论文

已超过10000项^[28]。

美国国立卫生研究院于2014年发起跨组学精准医学项目(Trans-Omics for Precision Medicine, TOPMed)旨在阐明心脏、肺、血液和睡眠障碍相关的遗传基础^[29]; 2018年又发起All of Us项目, 计划建立一个包含100万不同种族的美国人的研究队列, 与以往不同的是, 该项目提高非欧洲祖源的样本比例(占到40%以上), 以提高研究的代表性^[30]。到2024年, 该项目发表超过24.5万人的基因组测序结果^[31]。美国还开展百万退伍军人计划(Million Veteran Program, MVP)^[32], 在其最新的研究中完成超过63万个人的基因组测序和2068种性状的全基因组关联分析(genome-wide association study, GWAS)^[33]。

我国于2016年启动中国人群精准医学研究计划, 经过很多研究者的共同努力, 我国的大规模人群基因组项目也取得不错的进展。中国科学院生物物理研究所徐涛团队和何顺民团队牵头的“女娲”(NyuWa)中国人群基因组计划, 完成2999个不同中国样本的高测序深度全基因组数据, 并从SNV、插入缺失、转座子、微卫星、适应性进化信号等多方面解析中国人群基因组遗传变异^[34-36]。2020年, 中国科学院上海营养与健康研究所徐书华团队^[37]收集整合已有的汉族人测序数据, 发表10万人汉族基因组计划(The Han100K Initiative)的一期成果, 建立PGG.Han数据库, 并开发分析平台和工具。该数据库的2.0版本涵盖17546个高深度的全基因组测序、11878个低深度的全基因组测序、以及5002个全外显子组测序等信息。瑞金医院牵头的中国代谢解析计划(China Metabolic Analytics Project, ChinaMAP)于2020年发表来自8个民族的10588个个体的高深度全基因组测序数据及其系统分析^[38]。2021年, 西湖大学和中南大学研究团队联合发表西湖(中国)生物样本库(Westlake BioBank for Chinese, WBBC)的先导期研究成果, 涵盖4535个中国人的全基因组测序^[39]。2024年, 哈尔滨工业大学王亚东团队联合国内十余家单位发表当前最高精度的中国人群基因组变异图谱和参考基因组面板, 包括25734例样本^[40]。在疾病相关的群体基因组学研究中, 首都医科大学附属北京天坛医院王拥军团队^[41]报道10241例缺血性脑血管病患者的基因组特征, 华西医院也开展“十万例罕见病患者全基因组测序计划”, 目标是收集十万例罕见病患者和两万例自然人群的表型和全基因组测序数据, 研究罕见病

的遗传基础。

政府间或区域间合作的人群基因组计划也正蓬勃发展, 例如非洲人类遗传与健康计划(H3Africa)^[42]、亚洲基因组计划(GenomeAsia)^[43]等。2024年9月, 中国、美国、新加坡、英国等14个国家的科学家共同倡议启动人类基因组计划二期, 初步目标是完成全球1%人口(至少100个国家的8000万人)的基因组测序, 同时在至少10个国家构建多组学人群队列^[44]。目前, 测序一个人的全基因组的成本已经下降到约1000元人民币, 人类的全基因组数据还在迅速地增加, 超大规模的群体基因组数据能够为更全面地发掘人类的遗传多样性, 揭示精细的人群遗传结构和演化历史, 建立基因型与复杂表型、疾病的关联等提供数据支撑。

在其他物种中, 群体基因组研究的规模也越来越大, 代表性也越来越强。黑腹果蝇(*Drosophila melanogaster*)、拟南芥(*Arabidopsis thaliana*)等模式生物中的研究已经覆盖世界各地超过1000个基因组^[45,46], 在农作物、家养动物中样本量也迅速增加, 例如, 2018年, 中国农业科学院、国际水稻研究所和华大基因共同完成3010个水稻(*Oryza sativa*)基因组的测序和分析^[47]; 家犬基因组研究联盟2023年完成约2000只狗(*Canis familiaris*)的基因组测序和分析^[48]。

3 不断进步的方法技术

二十多年前, 可用于生物信息分析的软件很少, 加州大学圣克鲁兹分校(University of California, Santa Cruz)的James Kent开发GigAssembler将测序得到的大量序列片段拼接完成第一个基因组草图, 被称为人类基因组计划的英雄^[49]。相比人类基因组计划时期, 现在不论测序技术、数据分析方法、还是实验技术, 都取得了很大进步。

技术进步的一个重要的成果是参考基因组质量得到很大提高。在单分子实时测序(single molecule real-time sequencing)、纳米孔长读长测序(nanopore sequencing)、染色质构象捕捉(chromosome conformation capture)、全基因组光学图谱(optical genome mapping)等技术的辅助下, 多国科学家组成的端粒到端粒联盟(telomere-to-telomere, T2T)在2022年发表首个完整的人类基因组序列, 揭示此前由于技术原因无法组装的约8%的基因组区域, 主要包括端粒、着丝粒

和大片段重复等^[50]。其他很多物种, 如玉米(*Zea mays*)^[51]、水稻^[52]、西瓜(*Citrullus lanatus*)^[53]、家鸡(*Gallus gallus*)^[54]等的T2T基因组也被发表或被提上日程。完整的基因组使得变异的鉴定更加准确和全面, 新的计算方法被开发出来检测除SNV以外的其他类型的变异, 包括插入缺失、转座子序列、染色体倒位、拷贝数变异等复杂的结构变异^[55-57]。在自然群体变异检测的基础上, 泛基因组(pan-genome)的概念在2005年被提出, 并伴随着T2T基因组快速发展^[58]。泛基因组通过区分核心基因(core genome)和非核心基因(dispensable genome)整合多个个体或品种的基因组信息, 能够进一步增加基因组的代表性, 并减少分析过程中可能存在的偏倚, 为理解群体演化历史、解析表型和疾病的遗传基础提供准确的参考。人类的泛基因组草图在2023年发表, 包含47个来自全球不同遗传背景的人的基因组序列^[59]。同年, 中国人群泛基因组联盟也发表基于36个族群的中国人泛基因组参考图谱^[60]。在具有重要的经济和研究价值的农作物、家养动物和模式生物中, 复杂变异和泛基因组的相关研究在揭示适应、驯化等相关表型的遗传基础中开始发挥重要作用^[61-65]。

近年来, 基于群体多态性数据估计群体历史、推断基因流、以及检测自然选择的软件层出不穷, 已有文献详细地总结不同软件的基本原理、适用情形和不足^[66]。以检测正选择信号为例, 早期的方法主要基于不同物种间基因序列中突变的进化速度来推测选择压力的存在和强度。其中, dN/dS比率分析是一种经典方法, 通过比较编码区中非同义突变(会导致氨基酸变化)与同义突变(不改变氨基酸序列)的比率, 来推断基因是否受到选择^[67]。dN/dS比率大于1通常被认为是正选择的标志, 因为非同义突变的固定可能为生物提供适应性优势。此外, McDonald-Kreitman检验则结合种内多样性和种间分歧信息, 通过比较非同义突变和同义突变在种内和种间的比例差异, 提供一种更为灵敏的选择信号检测方法^[68]。当种间的非同义突变比例显著高于种内比例时, 可能反映出正选择的作用。这些方法为早期研究正选择提供重要的理论基础和分析工具, 但在数据复杂性和灵敏性上也存在一定的局限。过去二十年中, 大量正选择检测方法被开发出来。这些方法主要基于正选择产生的不同特征来检测选择, 主要包括三类: (i) 选择常造成局部区域的等位基因频谱

偏离突变-漂变平衡, 有利变异的快速固定使其紧密连锁的多态性位点因为搭便车(hitchhiking)效应频率升高。受选择区域周围的多态性迅速下降, 等位基因频谱中低频位点和后产生的等位基因(derived allele)的高频位点相对过剩, 这种信号可以被Tajima's *D*, Fay and Wu's *H*等统计量检测到(此类方法如DHtest^[69]); (ii) 发生在部分群体中的正选择会造成群体之间的遗传分化加快, 在受选择区域周围产生局部的高分化区域, 这种信号可以被PBS^[70], Ohana^[71]等方法检测到, 基于主成分分析的方法也属于此类的变种; (iii) 对于近期发生的正选择, 由于有利变异所在的单倍型没有足够的时间发生重组, 其单倍体长度会增加, 基于这类信息检测正选择的方法有iHS^[72], XP-EHH^[73], nSL^[74]等。

不同方法利用的信息不同, 对不同时空尺度的正选择检测能力存在差别(图2)。例如, 基于等位基因频谱统计量的方法有较为广泛的应用, 但当群体结构、种群大小变化等群体历史的影响较大时, 容易出现假阳性; 基于单倍型长度的方法对检测近期的选择比较有效, 但随着重组的发生, 单倍型的连锁很快被打破, 信息丢失; 基于群体分化的选择检测方法对检测局地适应(local adaptation)特别有效, 但无法检测发生在祖先群体中的正选择。为提高正选择检测的效力并缩小受选择变异的范围, 有方法整合上述多方面信息, 构建复合的统计量来检测正选择, 例如CMS^[75]。此外, 基于其他原理的正选择检测方法也被开发出来, 包括基于溯祖理论和基因树型的检测方法(如Relate^[76]), 或者利用计算机模拟和机器学习(如hierarchical boosting^[77], S/HIC^[78]等)检测正选择信号的方法。已有文章详细综述不同方法的检测原理和在不同选择情形下的检测能力^[79-81]。

新的群体遗传的分析方法还在被不断开发出来, 相比于已有的方法, 新的方法的改进通常在于利用更加综合的信息, 提高方法的敏感性和特异性; 优化计算方法, 速度更快并节省资源, 使大规模群体基因组的利用变得可行; 使用更加复杂的符合实际的模型, 提高结果的准确性; 部分软件还结合机器学习的方法, 把大规模的计算机模拟和训练应用到群体演化历史的估计和选择信号检测中, 提升软件的表现^[82]。例如, 在群体历史方面, 很多方法被开发出来, 这些方法能够揭示非常复杂的群体历史, 比如ADMIXTOOLS 2软件能

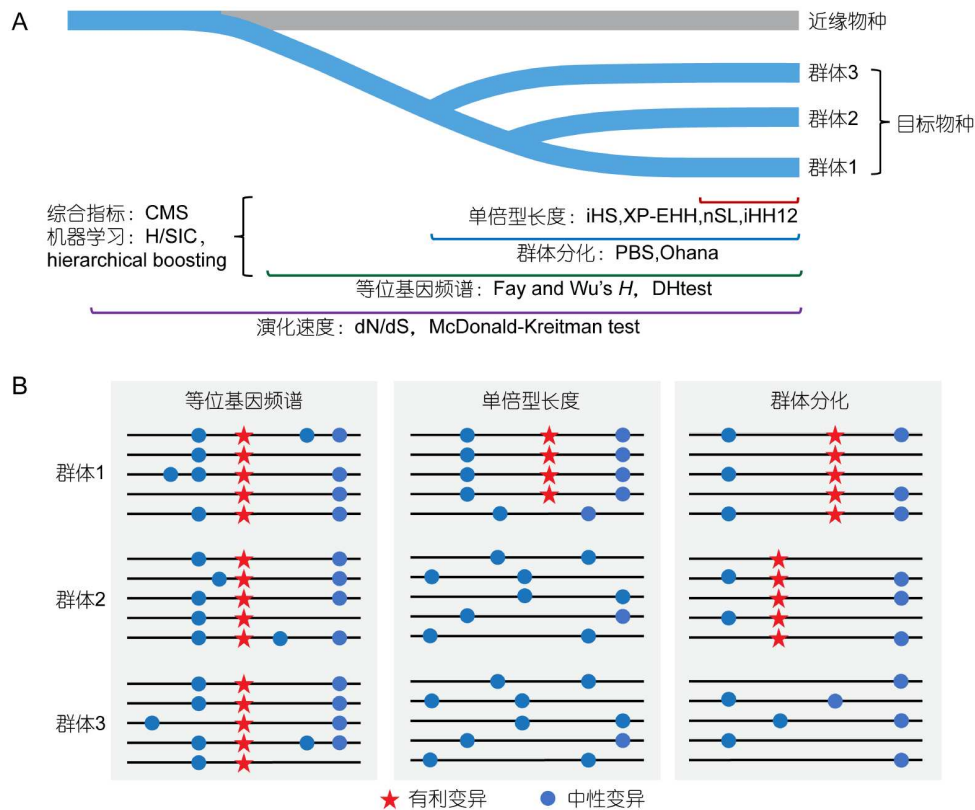


图 2 适合不同尺度和情形的正选择检测方法。A: 正选择在基因组中留下不同类型的特征, 图中展示利用不同类型信息来检测正选择信号的方法在不同时间尺度检测效力上存在的差异(图片基于文献[80]重新绘制)。基于单倍型长度的方法可以检测较近时期的正选择, 基于群体分化的方法可以检测群体分歧后的正选择, 正选择在等位基因频谱留下的信号可以持续较长时间。在不同物种间, 主要利用演化速度的信息检测正选择; B: 图中展示利用不同类型的信息检测正选择的3个示意图, 红色五角星表示有利变异, 蓝色圆圈表示中性变异。(左)有利变异受选择很快固定下来, 造成选择性清除, 该区域的等位基因频谱中低频和高频部分相对过剩; (中)有利变异在群体1中受正选择, 频率快速升高, 使单倍型长度增加; (右)群体1和群体2中不同的变异受到正选择, 造成群体间分化升高

Figure 2 Methods for detecting positive selection at different scales and scenarios. A: Positive selection leaves different types of features in the genome. The figure shows the difference in the power of methods in detecting positive selection signals on different time scales (the figure is adapted from ref. [80]). The method based on the length of haploid homozygosity can detect relatively recent positive selection. The method based on population differentiation can detect positive selection after population divergence. The positive selection signals left in the allele frequency spectrum can last long. The information of evolutionary rate is mainly used to detect positive selection between species; B: the figure shows three diagrams using different information to detect positive selection. The red pentagram represents beneficial mutation, and the blue circle represents neutral variation. (Left) Beneficial mutation is quickly fixed by selection, resulting in a selective sweep. Both rare alleles and high-frequency-derived alleles are in excess in the nearby region. (Middle) Beneficial mutation was under positive selection in population 1, and its frequency increased rapidly, resulting in extended haplotypes. (Right) Different mutations are positively selected in population 1 and population 2, resulting in increased population differentiation

够利用群体的等位基因频率计算多个 f 统计量, 并推测复杂的群体间分化和基因流事件。在选择信号检测中, 传统的方法主要基于硬选择性清除模型(hard sweep model)。该模型假设, 具有适应优势的突变在选择压力出现后通过突变产生, 并在选择作用下迅速扩张至较高频率, 最终在群体中固定。这一过程通常伴随着受选择位点周围遗传多样性的显著下降、低频等位基因的积累以及连锁不平衡区域的延长。然而, 近期研究表

明, 自然选择中软选择性清除模型(soft sweep model)也非常常见。该模型包括两种情形: 第一种是基于既有变异(standing variation)的选择, 即选择压力出现前, 有利变异已存在于群体中, 经历重组和突变积累后, 当选择发生时, 多个携带有利变异的单倍型能够同时响应选择; 第二种情形是重复产生有利变异(recurrent mutation), 即在群体规模较大或突变速率较高的情况下, 一个有利变异尚未固定时, 另一个有利变异已产生并共

同受到选择作用. 与硬选择性清除模型相比, 软选择性清除模型对目标区域周围的遗传多样性和连锁不平衡的影响较小, 从而更复杂地反映自然选择在群体中的作用机制. 基于软选择性清除模型的选择检测方法, 如H12^[83], iHH12^[84], HaploSweep^[85]等被陆续开发, 鉴定到很多以往研究漏掉的选择信号. 研究发现, 软选择性清除模型在人、黑腹果蝇、小鼠的自然群体中都更加普遍. 在计算机模拟方面, SLiM软件对于计算方法的优化使得现在可以根据任意复杂程度的群体模型产生基于个体的正向模拟(individual-based forward simulation)数据^[86].

古DNA记录着历史上某个时期的基因组信息, 结合考古定年, 能够极大地拓展研究的时间尺度, 给群体基因组学注入新鲜血液. 但由于古DNA经历长时间的降解, 片段很短, 并很容易受到微生物和人为环境的污染, 长期以来, 人们很难准确可靠地提取古DNA中的遗传信息. 古DNA技术在过去十几年中发展迅速, 研究建立严格的实验操作流程、防止污染的措施和探针富集等方法, 目前的技术已经能够让研究者从保存较好的约120万年前的猛犸象遗骸中提取古DNA^[87]. 2010年, 尼安德特人^[88]和丹尼索瓦人^[89]这两个古人类的基因组被发表, 他们是现代人的近亲, 在大约在55万年前与现代人分开. 古人类研究为理解现代人的演化历史提供非常重要的信息. 研究发现, 这两个古人类与走出非洲的现代人祖先发生多次基因交流, 这些遗传片段部分被保留在现代人的基因组中, 影响现代人的多种表型. 欧洲人的肤色、发色、身高等表型以及一些疾病相关的基因, 包括糖尿病等代谢病, 红斑狼疮等免疫疾病, 精神分裂症等神经系统疾病, 甚至包括新冠重症易感性等都受到尼安德特人的遗传渗透的影响^[90,91]. 研究表明, 藏族人群中发现的与高原适应相关的*EPAS1*的等位基因可能来自丹尼索瓦人^[92]. 古人类研究的相关进展详见^[93,94]. 该领域的先驱, 瑞典进化遗传学家Svante Pääbo因其在已灭绝人类基因组学研究方面的开创性贡献获得2022年诺贝尔生理学或医学奖. 他的研究不仅揭示现代人与已灭绝人类的遗传联系, 还推动古DNA技术的广泛应用, 为动植物驯化历史^[95]、古代疫病^[96]、发酵^[97]相关的微生物等研究提供全新视角.

全基因组关联分析在解析基因型表型关联中有非常重要的作用. 人类超大规模的基因组测序和比较丰

富的表型数据使得人类中的GWAS在建立基因型-表型关联上特别有效. 最近二十年中, GWAS在样本数目、模型复杂程度和计算效率等方面有明显进步. 已经有大量的研究揭示人类与各种表型相关的遗传变异, 在GWAS Atlas^[98]上已经有超过4500个人类表型的GWAS分析. 例如, 2022年的一项重要的工作根据约540万人的基因型-表型信息发现了12000多个与身高显著相关的变异位点, 这些变异能够解释绝大部分的人群中身高的差异, 说明已基本达到饱和^[99]. 此外, GWAS的数据类型也有了扩展. 除传统的宏观表型数据, 多种分子表型也被纳入关联分析中, 比如将基因表达水平、代谢物水平、蛋白质水平、RNA编辑水平、肠道微生物组成等作为表型的GWAS. 此外, 还出现全表型组关联分析(phenome-wide association studies, PheWAS). PheWAS寻找与某一个或一组多态性位点相关的表型, 可以发现表型间的关联和基因多效性等, 对揭示复杂的基因型表型关联有很好的补充作用^[100]. GWAS涉及数量遗传学, 超出本文的范畴, 已有文章详细综述GWAS相关的进展^[101,102], 这里不作深入展开.

4 适应的遗传基础和进化规律

适应是生物多样性产生和维持的重要动力, 其遗传学基础和演化规律一直是进化生物学的核心问题. 对人类、模式生物及非模式生物的群体基因组学研究已经阐明很多环境适应的遗传基础. 总的来看, 不同的适应相关表型具有差异的遗传架构(genetic architecture), 具体表现在与表型相关的遗传变异的位点数目和效应大小存在差异. 对于有的表型, 如杀虫剂抗性, 常常只涉及少数几个有主要效应的基因; 有的表型的影响基因有十数个到数十个, 比如已知的影响动物体色的基因; 而有的复杂表型涉及大量微效位点的共同作用, 例如身高、对温度的适应等. 基因编码和基因表达调控的变化都在适应中有重要作用. 适应的遗传架构中, 关于基因多效性、位点间相互作用的研究还比较少.

通过对不同地区人群多样性和选择信号的分析, 研究发现与局地适应(local adaptation)有关的遗传变异, 并据此绘制人类演化过程中的适应地图^[103]. 研究发现欧洲和北非人群中乳糖耐受与*LCT*基因的调控序

列变化相关^[104,105], 因纽特人群脂肪酸代谢相关基因变化使其能够适应高脂肪饮食^[106]. 研究还揭示在一些极端环境下的人群适应机制, 如藏族人的*EPAS1*, *EGLN1*, *HMOX2*, *TMEM247*等基因的变化参与高原适应^[70,107~109]; 印尼Bajau人群的基因*PDE10A*的突变与更大的脾脏相关, 利于储存携氧的红细胞, 为长期潜水供氧, 帮助其适应水下捕鱼的生活方式^[110]. 不同地区人群的肤色有明显的差异, 肤色的深浅对于人群的环境适应有重要意义. 深色皮肤中的黑色素能够抵挡紫外线带来的损伤, 而在日照不足的地区, 浅色皮肤能够接受足量的紫外线, 促进维生素D的合成和骨骼的发育. 近年来的研究发现数十个肤色相关的基因和遗传变异, 揭示不同人群肤色变化的遗传基础和演化轨迹^[111]. 研究发现大量的与肤色相关的遗传变异等均起源于非洲, 在人群迁移扩散过程中又发生新的变化^[112]. *MFSD12*, *SLC24A5*, *PDPK1*等影响非洲人群的肤色. 欧洲人和东亚人肤色变浅的机制不同, 基于欧洲人群的大量研究发现*SLC24A5*, *SLC45A2*, *TYR*, *TYRP1*, *ASIP*, *MC1R*, *IRF4*等基因的突变参与欧洲人的肤色变浅过程^[113], 在东亚人的肤色变浅机制中*OCA2*, *KITLG*, *SLC24A2*, *GNPAT*, *PAH*等基因有重要作用^[114,115]. 肤色相关的最新研究还深入到细胞层面, 结合GWAS、选择信号检测、大规模平行报告实验、基因编辑等方法, 鉴定并验证大量肤色相关的遗传变异和其功能^[116,117].

实验室演化是进化生物学中的重要研究方法. 在特定的条件下, 通过人工选择让演化在短时间内快速发生, 让研究者能够真实地看到演化的过程. 以大肠杆菌(*Escherichia coli*)、酿酒酵母(*Saccharomyces cerevisiae*)、黑腹果蝇等为对象的实验室演化和测序(evolve and resequence, E&R)为理解适应的遗传基础和进化规律提供非常重要的见解. 其中, 最著名的要数美国密西根州立大学的Richard Lenski研究组于1988年启动的大肠杆菌的长期演化实验, 至今已超过35年, 研究详细揭示适应性进化的轨迹和动力学特征^[118,119]. 实验室演化在其他领域的应用还为生物从单细胞到多细胞的转变^[120,121]、微生物和肿瘤抗性药的演化、工业生产中的菌株优化等提供了新的见解.

由于实验室演化中的实验条件比较单一, 群体的大小受限, 难以代替自然的演化过程, 对自然群体的适应性研究仍然具有无法替代的价值. 其中, 对模式

生物黑腹果蝇自然群体的研究揭示环境适应快速发生的重要机制和时空动态. 基因组分析发现, 软选择性清除模型是果蝇基因组中的自然选择的主要模式^[83]. 果蝇的有效群体很大, 遗传多态性较高, 在选择发生时能够利用既有变异或产生新的有利变异来快速适应环境. 对果蝇杀虫剂抗性相关的研究发现, 基因*Ace*的多个具有抗性的等位基因在不同果蝇群体中快速、重复地出现, 说明适应不受限于新生突变, 支持软选择性清除模型^[122]. 另一个研究发现, 果蝇基因组中与杀虫剂抗性有关的变异产生时间要早于人工合成杀虫剂的历史, 说明群体中既有变异对杀虫剂适应有很大贡献^[45].

对不同大陆的果蝇群体的研究都发现在果蝇表型和基因型上存在随纬度的渐变(clinal variation)^[123~126]和随季节振荡的周期性(seasonal oscillation)特征^[127,128], 并且这两组多态性存在相关^[127~129]. 由于环境变量(非生物的环境因素如温度、日照时长、紫外线强度等, 生物因素如物种多样性、物种间竞争强度)随着纬度和季节也存在渐变的特征, 果蝇的遗传多态性和表型在时空尺度上的变化反映果蝇群体对环境的快速响应. 这些变异影响的基因广泛地参与与生育^[130]、代谢^[131]、免疫^[132]、温度适应^[127,128]等相关的生命活动. 在最近的一个工作中, 研究者使用群体箱实验对半自然状态下的果蝇群体演化进行连续的采样, 通过混池测序获得不同时间点的群体遗传信息. 研究发现果蝇基因组中大量的遗传多态性的频率持续跟随着环境变化而改变, 这种跟随可在两周以内就被检测出来, 且在不同年份的实验中保持一致, 该研究揭示颗粒度极细的适应过程, 证明群体中既有变异可以在生态学时间尺度追踪环境变化, 刷新研究者对适应性进化发生的时间尺度认识^[133,134].

非模式生物常常具有独特的形态、生活史特征和环境选择压力, 对它们的研究揭示非常多样的适应机制. 令达尔文着迷的加拉帕格斯地雀的喙形和喙大小的多样性被发现与少数几个有主要效应的基因座位有关, 包括与颅面部发育相关的基因*ALX1*^[135], 还包括1个约525 kb的超基因(supergene)^[136]; 不同生境中的鹿鼠(*Peromyscus polionotus*)在毛色上存在深浅的变化以更好地隐藏在背景中, 分析发现*Mc1r*和*Agouti*等少数具有较大效应的基因的变异造成毛色表型的变化, 这些基因上的氨基酸编码变化和表达调控变化都能引起相关表型变化, 在*Mc1r*和*Agouti*基因之间还存在相互

作用^[137~139], 生活在海洋的三刺棘鱼(*Gasterosteus aculeatus*)通常具有腹鳍刺, 帮助它们抵御海洋中的捕食者, 而淡水环境中生活的三刺棘鱼不需要腹鳍刺保护, 它们的*Pitx1*的一个增强子发生突变, 改变*Pitx1*的表达, 使腹鳍刺消失^[140]; 桦尺蠖(*Biston betularia*)基因*cortex*的第一个内含子区域被转座子插入, 其中的调控元件促进*cortex*的转录, 造成工业黑化相关的深色表型^[141]; 有趣的是, *cortex*基因的不同单倍型还控制枯叶蝶(*Kallima* genus)的不同拟态表型^[142]。

在环境适应的相关研究中, 对极端环境的适应, 特别是高原适应有很多突出进展。研究表明, 在高原生活的动物在形态、生理、行为等方面都发生明显的改变, 常见的包括皮毛更浓密、颜色更深, 心室壁增厚、心肌纤维变粗, 肺泡增多、血管增加等, 这些改变有利于适应高原低氧、低温、强紫外照射等环境条件。对人和哺乳动物^[62,143~146]、鸟类^[147,148]、两爬^[149,150]、鱼类^[151]等多个类群的比较研究揭示很多与高原适应相关的关键基因和变异。不同物种适应高原的机制整体上趋同, 都以低氧诱导因子(hypoxia-inducible factor, HIF)通路为核心(包括*EPAS1*, *EGLN1*, *PPARA*等基因), 也涉及DNA损伤修复、能量代谢等多个生物学过程。不同物种的高原适应机制也存在差异, 例如, 高原鼯鼠(*Eospalax fontanierii*)等动物中红细胞数目和血红蛋白水平升高以提高供氧能力, 但藏族人群和部分鸟类中则是通过提高血红蛋白和氧气的亲和能力来适应低氧环境^[152]。研究还发现, 基因交流在高原适应中可能有重要作用, 藏族人的*EPAS1*可能来自丹尼索瓦人的遗传渗透^[92], 藏獒和藏灰狼(*Canis lupus laniger*)中与高原适应相关的*EPAS1*突变也可能来自某个目前未知或已经灭绝的犬科动物^[153]。另外, 肠道微生物对于宿主的高原适应也有贡献^[154]。最近的一些研究还揭示结构变异在高原适应中具有重要作用^[62,143,144,146]。

除野生动植物, 对家养动植物在驯化过程中的适应机制以及其野生群体的多样性也有很多不错的研究成果。在动物方面, 对家畜、家禽、宠物、饲养鱼类、昆虫(蚕)等的研究揭示其驯化历史以及一些受到关注的性状的遗传基础。例如, 对不同品种的狗的研究揭示与体型、毛色、性格、行为、遗传疾病风险等的遗传基础^[155]。研究还发现, 不同品种的狗的基因组中与淀粉消化相关的基因受到正选择, 包括编码淀粉酶的*AMY2B*基因, 编码麦芽糖酶的*MGAM*, 参与葡萄糖转

运的*SGLT1*等。这些基因通过拷贝数增加, 蛋白氨基酸序列变化, 表达量升高等多种途径, 增强了狗消化利用淀粉的能力^[156], 这揭示狗驯化过程中的关键步骤。在植物方面, 群体基因组学研究主要集中在具有经济价值的物种上面^[157], 研究揭示了作物的多样性和驯化历史^[158], 蔬菜和水果的风味^[159]、作物的产量^[160]、抗逆性^[161]等重要性状的遗传基础。家养动植物相关的研究进展对于充分开发动植物种质资源、改良和优化品种具有重要意义^[162]。

5 新时代的展望和挑战

目前, 研究者已经积累很多物种在群体水平的基因组信息, 迎来一个新的比较群体基因组学的时代。通过分析不同进化尺度下多个群体的多态性信息, 未来的研究可以从单个物种的多个群体进一步扩展到近缘物种的多个群体中, 很大程度上扩宽研究的进化尺度, 这为演化生物学研究带来了新的机遇和挑战^[163,164]。首先, 比较群体基因组学可以在很多方面提高现有方法的检测能力和精确程度, 例如通过近缘物种的多态性来推断祖先状态, 寻找演化上保守的功能元件, 检验近缘物种分歧后的平衡选择, 鉴定跨物种的基因流。在这方面很好的一个例子是, 国内已有团队开发新的比较群体基因组学计算方法, 例如HDMKPRF^[165]和CEGA^[166], 这些方法充分利用近缘物种的种内多态性和种间分歧的信息, 在检测正选择信号和平衡选择信号等方面相比以往方法有更好的表现, 在其他的物种系统中有广阔的应用前景。基于高质量长读长测序的比较群体基因组学还能重建基因和元件的演化历史提供更丰富的细节, 特别是对于大范围的结构变异、微卫星、多拷贝基因、转座子以及自私元件和宿主的军备竞赛等快速演化过程。结合群体分布的地理和生态等信息, 比较群体基因组学有望回答环境适应过程中的趋同演化, 以及群体分化和物种形成模式等重要的演化生物学问题。

在功能基因组学方面, 大规模的群体基因组结合上高质量的表型数据, 研究者可以解析更加复杂的基因型和表型的关联。虽然大规模高质量表型数据的收集依然是一个较大的挑战, 但目前已经有一些项目如“人类表型组”计划, 投入大量资源系统收集表型数据。结合转录组, 表观组和3D基因组、蛋白质组等多组学

数据, 群体基因组学可以探究基因表达调控在不同进化尺度下的演化规律. 利用基因编辑系统制备转基因细胞系或动植物品系, 可以解析近缘物种的等位基因在功能上的分化.

快速增长的数据量和新的研究需求也给研究者提出了新的挑战. 领域需要开发可靠的计算和分析方法, 更加高效地利用大规模甚至超大规模的数据, 例如需要更准确的软件检测不同类型的变异, 充分利用人工智能和计算机模拟等技术来帮助识别例如漂变、选择和基因流等造成的多态性模式, 需要更复杂的群体分化模型来重建演化历史等. 另外, 获得大规模高质量的表型数据、建立基因型-表型-环境之间的关联, 特别是涉及位点间相互作用和基因多效性的情况, 仍然非常具有挑战性.

在疾病研究方面, 将群体基因组学的理论框架应用于病原微生物, 解析病原微生物的变异和演化规律, 寻找跟致病性和免疫逃逸相关的关键位点, 能够给疾病防控、疫苗方案、药物设计等提供参考. 例如, 在新冠病毒疫情暴发初期, 本团队及合作者利用群体基因组学方法, 对新冠病毒的变异及演化规律开展一系列研究, 发现新冠病毒存在“L”和“S”两个主要谱系^[167]; 揭示L/S谱系与感染患者临床症状的相关性^[168]; 基于L/S分型探索新冠病毒传播中的奠基者效应^[169]; 揭示新冠病毒基因组变异的演进规律^[170]; 提出

“新冠病毒感染人数与进化速度呈现正反馈, 导致进化失控”的模型^[171]; 发现人群中新冠病毒的Spike蛋白正选择靶位点与其他冠状病毒存在显著差异^[172], Spike蛋白的变异影响新冠病毒的传播、免疫逃逸和致病性^[173]; 发现新冠病毒倾向于使用人类基因组中非偏好性密码子, 并基于此开发S蛋白在人类细胞中表达的密码子优化算法^[174]; 结合经典传染病学理论, 融合病毒变异、免疫衰退及免疫逃逸因素, 开发SIRSVIDE模型, 用于预测新冠病毒传播性增强、致病性变弱及免疫逃逸能力增强的长期演化趋势^[175]. 针对肿瘤等微观进化过程开发新的理论和计算方法^[176-178], 结合越来越多的测序数据, 有望理解肿瘤进化过程的随机性和驱动力, 为肿瘤治疗提供新观点.

展望未来, 随着基因编辑、人工智能和合成生物学等领域的快速发展, 群体基因组学将迎来更多新的机遇和挑战, 例如通过深度学习建立模型, 预测变异的功能; 改良家养动植物或微生物的基因, 获得更好的性状; 通过基因编辑影响突变率和重组率等基本的群体遗传参数, 改变物种进化速度和方向; 设计人造生命、重现已灭绝的生物等. 在这一过程中, 跨学科的合作将变得更加重要. 生物学、数学、计算机等不同领域研究者的共同努力将进一步推动群体基因组学的发展和在演化生物学、生态学、农业和医学等领域的广泛应用.

致谢 感谢北京大学生命科学学院博士研究生刘晨露、李伟轩、徐启姮为文章修改提出的宝贵意见.

参考文献

- 1 Hubby J L, Lewontin R C. A molecular approach to the study of genic heterozygosity in natural populations. I. the number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics*, 1966, 54: 577-594
- 2 Lewontin R C, Hubby J L. A molecular approach to the study of genic heterozygosity in natural populations. II. amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 1966, 54: 595-609
- 3 Kimura M. Evolutionary rate at the molecular level. *Nature*, 1968, 217: 624-626
- 4 Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature*, 1973, 246: 96-98
- 5 Kingman J F C. The coalescent. *Stoch Proc Appl*, 1982, 13: 235-248
- 6 Kreitman M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*, 1983, 304: 412-417
- 7 Luikart G, Kardos M, Hand B K, et al. Population genomics: advancing understanding of nature. In: Rajora O P, ed. *Population Genomics: Concepts, Approaches and Applications*. Cham: Springer, 2019. 3-79
- 8 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860-921
- 9 Venter J C, Adams M D, Myers E W, et al. The sequence of the human genome. *Science*, 2001, 291: 1304-1351
- 10 Gibbs R A. The Human Genome Project changed everything. *Nat Rev Genet*, 2020, 21: 575-576

- 11 The International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005, 437: 1299–1320
- 12 Frazer K A, Ballinger D G, Cox D R, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 2007, 449: 851–861
- 13 Altshuler D M, Gibbs R A, Altshuler D M, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*, 2010, 467: 52–58
- 14 Durbin R M, Altshuler D, Durbin R M, et al. A map of human genome variation from population-scale sequencing. *Nature*, 2010, 467: 1061–1073
- 15 Auton A, Abecasis G R, Altshuler (Co-Chair) D M, et al. A global reference for human genetic variation. *Nature*, 2015, 526: 68–74
- 16 Sudmant P H, Rausch T, Gardner E J, et al. An integrated map of structural variation in 2504 human genomes. *Nature*, 2015, 526: 75–81
- 17 Hanchard N A, Choudhury A. 1000 Genomes Project phase 4: the gift that keeps on giving. *Cell*, 2022, 185: 3286–3289
- 18 Mallick S, Li H, Lipson M, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 2016, 538: 201–206
- 19 Chen S, Francioli L C, Goodrich J K, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*, 2024, 625: 92–100
- 20 Gudbjartsson D F, Helgason H, Gudjonsson S A, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*, 2015, 47: 435–444
- 21 Sulem P, Helgason H, Oddson A, et al. Identification of a large set of rare complete human knockouts. *Nat Genet*, 2015, 47: 448–452
- 22 Steinberg S, Stefansson H, Jonsson T, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer’s disease. *Nat Genet*, 2015, 47: 445–447
- 23 Helgason A, Einarsson A W, Guðmundsdóttir V B, et al. The Y-chromosome point mutation rate in humans. *Nat Genet*, 2015, 47: 453–457
- 24 Jónsson H, Sulem P, Kehr B, et al. Parental influence on human germline *de novo* mutations in 1548 trios from Iceland. *Nature*, 2017, 549: 519–522
- 25 Kong A, Frigge M L, Masson G, et al. Rate of *de novo* mutations and the importance of father’s age to disease risk. *Nature*, 2012, 488: 471–475
- 26 Skov L, Coll Macià M, Sveinbjörnsson G, et al. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature*, 2020, 582: 78–83
- 27 Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 2018, 562: 203–209
- 28 Allen N E, Lacey B, Lawlor D A, et al. Prospective study design and data analysis in UK Biobank. *Sci Transl Med*, 2024, 16: eadf4428
- 29 Taliun D, Harris D N, Kessler M D, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, 2021, 590: 290–299
- 30 Denny J C, Rutter J L, Goldstein D B, et al. The “All of Us” Research Program. *N Engl J Med*, 2019, 381: 668–676
- 31 Bick A G, Metcalf G A, Mayo K R, et al. Genomic data in the All of Us Research Program. *Nature*, 2024, 627: 340–346
- 32 Gaziano J M, Concato J, Brophy M, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*, 2016, 70: 214–223
- 33 Verma A, Huffman J E, Rodriguez A, et al. Diversity and scale: genetic architecture of 2068 traits in the VA Million Veteran Program. *Science*, 2024, 385: eadj1182
- 34 Zhang P, Luo H, Li Y, et al. NyuWa Genome resource: a deep whole-genome sequencing-based variation profile and reference panel for the Chinese population. *Cell Rep*, 2021, 37: 110017
- 35 Niu Y, Teng X, Zhou H, et al. Characterizing mobile element insertions in 5675 genomes. *Nucleic Acids Res*, 2022, 50: 2493–2508
- 36 Liu S, Luo H, Zhang P, et al. Adaptive selection of *Cis*-regulatory elements in the Han Chinese. *Mol Biol Evol*, 2024, 41: msae034
- 37 Gao Y, Zhang C, Yuan L, et al. PGG.Han: the Han Chinese genome database and analysis platform. *Nucleic Acids Res*, 2020, 48: D971–D976
- 38 Cao Y, Li L, Xu M, et al. The ChinaMAP analytics of deep whole genome sequences in 10588 individuals. *Cell Res*, 2020, 30: 717–731
- 39 Cong P K, Bai W Y, Li J C, et al. Genomic analyses of 10376 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. *Nat Commun*, 2022, 13: 2939
- 40 Jiang T, Guo H, Liu Y, et al. A comprehensive genetic variant reference for the Chinese population. *Sci Bull*, 2024, 69: 3820–3825
- 41 Cheng S, Xu Z, Bian S, et al. The STROMICS genome study: deep whole-genome sequencing and analysis of 10K Chinese patients with ischemic stroke reveal complex genetic and phenotypic interplay. *Cell Discov*, 2023, 9: 75
- 42 Choudhury A, Aron S, Botigué L R, et al. High-depth African genomes inform human migration and health. *Nature*, 2020, 586: 741–748

- 43 Wall J D, Stawiski E W, Ratan A, et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*, 2019, 576: 106–111
- 44 Liu W, Li Y, Patrinos G P, et al. The 1% gift to humanity: the Human Genome Project II. *Cell Res*, 2024, 34: 747–750
- 45 Chen J, Liu C, Li W, et al. From sub-Saharan Africa to China: evolutionary history and adaptation of *Drosophila melanogaster* revealed by population genomics. *Sci Adv*, 2024, 10: eadh3425
- 46 Alonso-Blanco C, Andrade J, Becker C, et al. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, 2016, 166: 481–491
- 47 Wang W, Mauleon R, Hu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, 2018, 557: 43–49
- 48 Meadows J R S, Kidd J M, Wang G D, et al. Genome sequencing of 2000 canids by the Dog10K consortium advances the understanding of demography, genome function and architecture. *Genome Biol*, 2023, 24: 187
- 49 Kent W J, Haussler D. Assembly of the working draft of the human genome with GigAssembler. *Genome Res*, 2001, 11: 1541–1548
- 50 Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science*, 2022, 376: 44–53
- 51 Chen J, Wang Z, Tan K, et al. A complete telomere-to-telomere assembly of the maize genome. *Nat Genet*, 2023, 55: 1221–1231
- 52 Shang L, He W, Wang T, et al. A complete assembly of the rice Nipponbare reference genome. *Mol Plant*, 2023, 16: 1232–1236
- 53 Zhang Y, Zhao M, Tan J, et al. Telomere-to-telomere Citrullus super-pangenome provides direction for watermelon breeding. *Nat Genet*, 2024, 56: 1750–1761
- 54 Huang Z, Xu Z, Bai H, et al. Evolutionary analysis of a complete chicken genome. *Proc Natl Acad Sci USA*, 2023, 120: e2216641120
- 55 Zheng Z, Zhu M, Zhang J, et al. A sequence-aware merger of genomic structural variations at population scale. *Nat Commun*, 2024, 15: 960
- 56 Ho S S, Urban A E, Mills R E. Structural variation in the sequencing era. *Nat Rev Genet*, 2020, 21: 171–189
- 57 Ahsan M U, Liu Q, Perdomo J E, et al. A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. *Nat Methods*, 2023, 20: 1143–1158
- 58 Sherman R M, Salzberg S L. Pan-genomics in the human genome era. *Nat Rev Genet*, 2020, 21: 243–254
- 59 Liao W W, Asri M, Ebler J, et al. A draft human pangenome reference. *Nature*, 2023, 617: 312–324
- 60 Gao Y, Yang X, Chen H, et al. A pangenome reference of 36 Chinese populations. *Nature*, 2023, 619: 112–121
- 61 Kang M, Wu H, Liu H, et al. The pan-genome and local adaptation of *Arabidopsis thaliana*. *Nat Commun*, 2023, 14: 6259
- 62 An X, Mao L, Wang Y, et al. Genomic structural variation is associated with hypoxia adaptation in high-altitude zokors. *Nat Ecol Evol*, 2024, 8: 339–351
- 63 Tong X, Han M J, Lu K, et al. High-resolution silkworm pan-genome provides genetic insights into artificial selection and ecological adaptation. *Nat Commun*, 2022, 13: 5619
- 64 Yan H, Sun M, Zhang Z, et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat Genet*, 2023, 55: 507–518
- 65 Li N, He Q, Wang J, et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat Genet*, 2023, 55: 852–860
- 66 Bourgeois Y X C, Warren B H. An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Mol Ecol*, 2021, 30: 6036–6071
- 67 Yang Z, Bielawski J P. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*, 2000, 15: 496–503
- 68 McDonald J H, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, 1991, 351: 652–654
- 69 Zeng K, Fu Y X, Shi S, et al. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*, 2006, 174: 1431–1439
- 70 Yi X, Liang Y, Huerta-Sanchez E, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 2010, 329: 75–78
- 71 Cheng J Y, Stern A J, Racimo F, et al. Detecting selection in multiple populations by modeling ancestral admixture components. *Mol Biol Evol*, 2021, 39: msab294
- 72 Voight B F, Kudaravalli S, Wen X, et al. A map of recent positive selection in the human genome. *PLoS Biol*, 2006, 4: e72
- 73 Sabeti P C, Varilly P, Fry B, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 2007, 449: 913–918
- 74 Ferrer-Admetlla A, Liang M, Korneliussen T, et al. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol*, 2014, 31: 1275–1291

- 75 Grossman S R, Shylakhter I, Karlsson E K, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. [Science](#), 2010, 327: 883–886
- 76 Speidel L, Forest M, Shi S, et al. A method for genome-wide genealogy estimation for thousands of samples. [Nat Genet](#), 2019, 51: 1321–1329
- 77 Pybus M, Luisi P, Dall’Olio G M, et al. Hierarchical boosting: a machine-learning framework to detect and classify hard selective sweeps in human populations. [Bioinformatics](#), 2015, 31: 3946–3952
- 78 Schrider D R, Kern A D, Payseur B A. S/HIC: robust identification of soft and hard sweeps using machine learning. [PLoS Genet](#), 2016, 12: e1005928
- 79 Vitti J J, Grossman S R, Sabeti P C. Detecting natural selection in genomic data. [Annu Rev Genet](#), 2013, 47: 97–120
- 80 Sabeti P C, Schaffner S F, Fry B, et al. Positive natural selection in the human lineage. [Science](#), 2006, 312: 1614–1620
- 81 Hejase H A, Dukler N, Siepel A. From summary statistics to gene trees: methods for inferring positive selection. [Trends Genet](#), 2020, 36: 243–258
- 82 Huang X, Rymbekova A, Dolgova O, et al. Harnessing deep learning for population genetic inference. [Nat Rev Genet](#), 2024, 25: 61–78
- 83 Garud N R, Messer P W, Buzbas E O, et al. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. [PLoS Genet](#), 2015, 11: e1005004
- 84 Torres R, Szpiech Z A, Hernandez R D, et al. Human demographic history has amplified the effects of background selection across the genome. [PLoS Genet](#), 2018, 14: e1007387
- 85 Zhao S, Chi L, Fu M, et al. HaploSweep: detecting and distinguishing recent soft and hard selective sweeps through haplotype structure. [Mol Biol Evol](#), 2024, 41: msae192
- 86 Haller B C, Messer P W. SLiM 4: multispecies eco-evolutionary modeling. [Am Nat](#), 2023, 201: E127–E139
- 87 van der Valk T, Pečnerová P, Díez-del-Molino D, et al. Million-year-old DNA sheds light on the genomic history of mammoths. [Nature](#), 2021, 591: 265–269
- 88 Green R E, Krause J, Briggs A W, et al. A draft sequence of the neandertal genome. [Science](#), 2010, 328: 710–722
- 89 Reich D, Green R E, Kircher M, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. [Nature](#), 2010, 468: 1053–1060
- 90 Sankararaman S, Mallick S, Dannemann M, et al. The genomic landscape of Neanderthal ancestry in present-day humans. [Nature](#), 2014, 507: 354–357
- 91 Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. [Nature](#), 2020, 587: 610–612
- 92 Huerta-Sánchez E, Jin X, Asan X, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. [Nature](#), 2014, 512: 194–197
- 93 Liu Y, Mao X, Krause J, et al. Insights into human history from the first decade of ancient human genomics. [Science](#), 2021, 373: 1479–1484
- 94 Bennett E A, Fu Q. Ancient genomes and the evolutionary path of modern humans. [Cell](#), 2024, 187: 1042–1046
- 95 Frantz L A F, Bradley D G, Larson G, et al. Animal domestication in the era of ancient genomics. [Nat Rev Genet](#), 2020, 21: 449–460
- 96 Spyrou M A, Bos K I, Herbig A, et al. Ancient pathogen genomics as an emerging tool for infectious disease research. [Nat Rev Genet](#), 2019, 20: 323–340
- 97 Liu Y, Miao B, Li W, et al. Bronze Age cheese reveals human-Lactobacillus interactions over evolutionary history. [Cell](#), 2024, 187: 5891–5900. e8
- 98 Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. [Nat Genet](#), 2019, 51: 1339–1348
- 99 Yengo L, Vedantam S, Marouli E, et al. A saturated map of common genetic variants associated with human height. [Nature](#), 2022, 610: 704–712
- 100 Bush W S, Oetjens M T, Crawford D C. Unravelling the human genome-phenome relationship using phenome-wide association studies. [Nat Rev Genet](#), 2016, 17: 129–145
- 101 Visscher P M, Wray N R, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. [Am J Hum Genet](#), 2017, 101: 5–22
- 102 Abdellaoui A, Yengo L, Verweij K J H, et al. 15 years of GWAS discovery: realizing the promise. [Am J Hum Genet](#), 2023, 110: 179–194
- 103 Fan S, Hansen M E B, Lo Y, et al. Going global by adapting local: a review of recent human adaptation. [Science](#), 2016, 354: 54–59
- 104 Tishkoff S A, Reed F A, Ranciaro A, et al. Convergent adaptation of human lactase persistence in Africa and Europe. [Nat Genet](#), 2007, 39: 31–40

- 105 Bersaglieri T, Sabeti P C, Patterson N, et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet*, 2004, 74: 1111–1120
- 106 Fumagalli M, Moltke I, Grarup N, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*, 2015, 349: 1343–1347
- 107 Deng L, Zhang C, Yuan K, et al. Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders. *Natl Sci Rev*, 2019, 6: 1201–1222
- 108 Xiang K, Ouzhuluobu K, Peng Y, et al. Identification of a Tibetan-specific mutation in the hypoxic gene *EGLN1* and its contribution to high-altitude adaptation. *Mol Biol Evol*, 2013, 30: 1889–1898
- 109 Yang D, Peng Y, Ouzhuluobu Y, et al. *HMOX2* functions as a modifier gene for high-altitude adaptation in Tibetans. *Hum Mutat*, 2016, 37: 216–223
- 110 Ilardo M A, Moltke I, Korneliussen T S, et al. Physiological and genetic adaptations to diving in sea nomads. *Cell*, 2018, 173: 569–580.e15
- 111 Liu J, Bitsue H K, Yang Z. Skin colour: a window into human phenotypic evolution and environmental adaptation. *Mol Ecol*, 2024, 33: e17369
- 112 Crawford N G, Kelly D E, Hansen M E B, et al. Loci associated with skin pigmentation identified in African populations. *Science*, 2017, 358: eaan8433
- 113 Lamason R L, Mohideen M A P K, Mest J R, et al. *SLC24A5*, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, 2005, 310: 1782–1786
- 114 Yang Z, Zhong H, Chen J, et al. A genetic mechanism for convergent skin lightening during recent human evolution. *Mol Biol Evol*, 2016, 33: 1177–1187
- 115 Adhikari K, Mendoza-Revilla J, Sohail A, et al. A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nat Commun*, 2019, 10: 358
- 116 Feng Y, Xie N, Inoue F, et al. Integrative functional genomic analyses identify genetic variants influencing skin pigmentation in Africans. *Nat Genet*, 2024, 56: 258–272
- 117 Bajpai V K, Swigut T, Mohammed J, et al. A genome-wide genetic screen uncovers determinants of human pigmentation. *Science*, 2023, 381: eade6289
- 118 Barrick J E, Yu D S, Yoon S H, et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 2009, 461: 1243–1247
- 119 Couce A, Limdi A, Magnan M, et al. Changing fitness effects of mutations through long-term bacterial evolution. *Science*, 2024, 383: eadd1417
- 120 Ratcliff W C, Fankhauser J D, Rogers D W, et al. Origins of multicellular evolvability in snowflake yeast. *Nat Commun*, 2015, 6: 6102
- 121 Bozdag G O, Zamani-Dahaj S A, Day T C, et al. *De novo* evolution of macroscopic multicellularity. *Nature*, 2023, 617: 747–754
- 122 Karasov T, Messer P W, Petrov D A, et al. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet*, 2010, 6: e1000924
- 123 Bergland A O, Tobler R, González J, et al. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Mol Ecol*, 2016, 25: 1157–1174
- 124 Hoffmann A A, Anderson A, Hallas R. Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. *Ecol Lett*, 2002, 5: 614–618
- 125 Pitchers W, Pool J E, Dworkin I. Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many? *Evolution*, 2013, 67: 438–452
- 126 Schmidt P S, Paaby A B. Reproductive diapause and life-history clines in North American populations of *Drosophila melanogaster*. *Evolution*, 2008, 62: 1204–1215
- 127 Bergland A O, Behrman E L, O'Brien K R, et al. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet*, 2014, 10: e1004775
- 128 Machado H E, Bergland A O, Taylor R, et al. Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila*. *eLife*, 2021, 10: e67577
- 129 Rodrigues M F, Vrbáňovskí M D, Cogni R. Clinal and seasonal changes are correlated in *Drosophila melanogaster* natural populations. *Evolution*, 2021, 75: 2042–2054
- 130 Cogni R, Kuczyński C, Koury S, et al. The intensity of selection acting on the *couch potato* gene-spatial-temporal variation in a diapause cline.

- [Evolution](#), 2014, 68: 538–548
- 131 Cogni R, Kuczynski K, Lavington E, et al. Variation in *Drosophila melanogaster* central metabolic genes appears driven by natural selection both within and between populations. [Proc R Soc B](#), 2015, 282: 20142688
- 132 Behrman E L, Howick V M, Kapun M, et al. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. [Proc R Soc B](#), 2018, 285: 20172599
- 133 Bitter M C, Berardi S, Oken H, et al. Continuously fluctuating selection reveals fine granularity of adaptation. [Nature](#), 2024, 634: 389–396
- 134 Rudman S M, Greenblum S I, Rajpurohit S, et al. Direct observation of adaptive tracking on ecological time scales in *Drosophila*. [Science](#), 2022, 375: eabj7484
- 135 Lamichhaney S, Berglund J, Almén M S, et al. Evolution of Darwin’s finches and their beaks revealed by genome sequencing. [Nature](#), 2015, 518: 371–375
- 136 Enbody E D, Sendell-Price A T, Sprehn C G, et al. Community-wide genome sequencing reveals 30 years of Darwin’s finch evolution. [Science](#), 2023, 381: eadf6218
- 137 Hoekstra H E, Hirschmann R J, Bunday R A, et al. A single amino acid mutation contributes to adaptive beach mouse color pattern. [Science](#), 2006, 313: 101–104
- 138 Steiner C C, Weber J N, Hoekstra H E, et al. Adaptive variation in beach mice produced by two interacting pigmentation genes. [PLoS Biol](#), 2007, 5: e219
- 139 Vignieri S N, Larson J G, Hoekstra H E. The selective advantage of crypsis in mice. [Evolution](#), 2010, 64: 2153
- 140 Chan Y F, Marks M E, Jones F C, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. [Science](#), 2010, 327: 302–305
- 141 Hof A E V T, Campagne P, Rigden D J, et al. The industrial melanism mutation in British peppered moths is a transposable element. [Nature](#), 2016, 534: 102–105
- 142 Wang S, Teng D, Li X, et al. The evolution and diversification of oakleaf butterflies. [Cell](#), 2022, 185: 3138–3152.e20
- 143 Shi J, Jia Z, Sun J, et al. Structural variants involved in high-altitude adaptation detected using single-molecule long-read sequencing. [Nat Commun](#), 2023, 14: 8282
- 144 Gao X, Wang S, Wang Y F, et al. Long read genome assemblies complemented by single cell RNA-sequencing reveal genetic and cellular mechanisms underlying the adaptive evolution of yak. [Nat Commun](#), 2022, 13: 4887
- 145 Yu L, Wang G D, Ruan J, et al. Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. [Nat Genet](#), 2016, 48: 947–952
- 146 Jiang Y F, Wang S, Wang C L, et al. Pangenome obtained by long-read sequencing of 11 genomes reveal hidden functional structural variants in pigs. [iScience](#), 2023, 26: 106119
- 147 Qu Y, Chen C, Chen X, et al. The evolution of ancestral and species-specific adaptations in snowfinches at the Qinghai-Tibet Plateau. [Proc Natl Acad Sci USA](#), 2021, 118: e2012398118
- 148 Qu Y, Chen C, Xiong Y, et al. Rapid phenotypic evolution with shallow genomic differentiation during early stages of high elevation adaptation in Eurasian Tree Sparrows. [Natl Sci Rev](#), 2020, 7: 113–127
- 149 Sun Y B, Fu T T, Jin J Q, et al. Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. [Proc Natl Acad Sci USA](#), 2018, 115: E10634–E10641
- 150 Li J T, Gao Y D, Xie L, et al. Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. [Proc Natl Acad Sci USA](#), 2018, 115: 8406–8411
- 151 Wang Y, Yang L, Zhou K, et al. Evidence for adaptation to the Tibetan Plateau inferred from Tibetan loach transcriptomes. [Genome Biol Evol](#), 2015, 7: 2970–2982
- 152 Zhu X, Guan Y, Signore A V, et al. Divergent and parallel routes of biochemical adaptation in high-altitude passerine birds from the Qinghai-Tibet Plateau. [Proc Natl Acad Sci USA](#), 2018, 115: 1865–1870
- 153 Wang M S, Wang S, Li Y, et al. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. [Mol Biol Evol](#), 2020, 37: 2616–2629
- 154 Zhang Z, Xu D, Wang L, et al. Convergent evolution of rumen microbiomes in high-altitude mammals. [Curr Biol](#), 2016, 26: 1873–1879
- 155 Buckley R M, Ostrander E A. Large-scale genomic analysis of the domestic dog informs biological discovery. [Genome Res](#), 2024, 34: 811–821

- 156 Axelsson E, Ratnakumar A, Arendt M L, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*, 2013, 495: 360–364
- 157 Song B, Ning W, Wei D, et al. Plant genome resequencing and population genomics: current status and future prospects. *Mol Plant*, 2023, 16: 1252–1268
- 158 Liang Z, Duan S, Sheng J, et al. Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat Commun*, 2019, 10: 1190
- 159 Tieman D, Zhu G, Resende Jr. M F R, et al. A chemical genetic roadmap to improved tomato flavor. *Science*, 2017, 355: 391–394
- 160 Si L, Chen J, Huang X, et al. *OsSPL13* controls grain size in cultivated rice. *Nat Genet*, 2016, 48: 447–456
- 161 Todesco M, Owens G L, Bercovich N, et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 2020, 584: 602–607
- 162 Andersson L, Purugganan M. Molecular genetic variation of animals and plants under domestication. *Proc Natl Acad Sci USA*, 2022, 119: e2122150119
- 163 Lawrie D S, Petrov D A. Comparative population genomics: power and principles for the inference of functionality. *Trends Genet*, 2014, 30: 133–139
- 164 McGrath C, Zhang G. Highlight: comparative population genomics—answering old questions with new Data. *Genome Biol Evol*, 2022, 14: evab278
- 165 Zhao S, Zhang T, Liu Q, et al. Identifying lineage-specific targets of natural selection by a Bayesian analysis of genomic polymorphisms and divergence from multiple species. *Mol Biol Evol*, 2019, 36: 1302–1315
- 166 Zhao S, Chi L, Chen H. CEGA: a method for inferring natural selection by comparative population genomic analysis across species. *Genome Biol*, 2023, 24: 219
- 167 Tang X, Wu C, Li X, et al. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev*, 2020, 7: 1012–1023
- 168 Hu B, Liu R, Tang X, et al. The concordance between the evolutionary trend and the clinical manifestation of the two SARS-CoV-2 variants. *Natl Sci Rev*, 2021, 8: nwab073
- 169 Ruan Y, Luo Z, Tang X, et al. On the founder effect in COVID-19 outbreaks: how many infected travelers may have started them all? *Natl Sci Rev*, 2021, 8: nwaa246
- 170 Tang X, Ying R, Yao X, et al. Evolutionary analysis and lineage designation of SARS-CoV-2 genomes. *Sci Bull*, 2021, 66: 2297–2311
- 171 Ruan Y, Hou M, Tang X, et al. The runaway evolution of SARS-CoV-2 leading to the highly evolved delta strain. *Mol Biol Evol*, 2022, 39: msac046
- 172 Tang X, Qian Z, Lu X, et al. Adaptive evolution of the Spike protein in coronaviruses. *Mol Biol Evol*, 2023, 40: msad089
- 173 Yao Z, Zhang L, Duan Y, et al. Molecular insights into the adaptive evolution of SARS-CoV-2 spike protein. *J Infect*, 2024, 88: 106121
- 174 Wu X, Shan K, Zan F, et al. Optimization and deoptimization of codons in SARS-CoV-2 and related implications for vaccine development. *Adv Sci*, 2023, 10: e2205445
- 175 Jin K, Tang X, Qian Z, et al. Modeling viral evolution: a novel SIRSVIDE framework with application to SARS-CoV-2 dynamics. *hLife*, 2024, 2: 227–245
- 176 Wu C I, Wang H Y, Ling S, et al. The ecology and evolution of cancer: the ultra-microevolutionary process. *Annu Rev Genet*, 2016, 50: 347–369
- 177 Zhang L, Deng T, Liufu Z, et al. The theory of massively repeated evolution and full identifications of cancer-driving nucleotides (CDNs). *eLife*, 2024, 13: RP99340
- 178 Zhang L, Deng T, Liufu Z, et al. Characterization of cancer-driving nucleotides (CDNs) across genes, cancer types, and patients. *eLife*, 2024, 13: RP99341

Frontiers in population genomics: development, technological innovation, and interdisciplinary integration

CHEN JunHao, ZHANG Lin & LU Jian^{*}

State Key Laboratory of Gene Function and Modulation Research, School of Life Sciences, Peking University, Beijing 100871, China

** Correspondence author; Email: luj@pku.edu.cn*

Population genomics combines high-throughput sequencing technologies and statistical methods to investigate genome-wide variations between and within populations. Population genomics elucidates the origins of genetic variations and their roles in evolution and adaptation by analyzing processes such as mutation, recombination, genetic drift, natural selection, and migration. In the past two decades, the field has undergone rapid advancement with an explosion of the data scale and constant iterations of analytical approaches. Significant breakthroughs have been made in many aspects, including histories of human migration, adaptive evolution of natural populations, domestication, and evolutionary trajectory of pathogens, which enhanced our understanding of evolutionary histories and the genetic mechanism of environmental adaptation. This review briefly summarizes the development history and main achievements of population genomics and discusses the opportunities and challenges faced by population genomics in the new era.

population genomics, genetic diversity, evolutionary history, adaptation

doi: [10.1360/SSV-2024-0311](https://doi.org/10.1360/SSV-2024-0311)