

基于高通量测序的植物群体基因组学研究进展

王云生

凯里学院环境与生命科学学院, 凯里 556011

摘要: 作为群体遗传学一种新的表现形式, 群体基因组学是将基因组概念和技术与群体遗传学理论体系相结合, 通过覆盖全基因组范围内的多态位点的分布式样推测位点特异性效应和全基因组效应, 从而提升人们对微进化的理解。近年来, 随着第二代高通量测序技术的出现和改进, 完成基因组测序的植物种类迅速增加, 大规模的重测序也随之开展。与此同时, 在一些尚未完成基因组测序的植物物种中, 也开展了一些平行测序。这些重测序和平行测序极大地促进了群体基因组学的发展, 加深了人们对相关植物种群在基因组水平上的遗传多样性、连锁不平衡水平、选择作用、群体历史及复杂性状的分子机理等群体基因组学方面的认识。本文简要介绍了群体基因组学的概念、研究方法等, 重点综述了基于高通量测序的植物群体基因组学的研究动态, 展望了植物群体基因组学的发展前景并讨论了存在的问题, 以期对相关研究提供借鉴和参考。

关键词: 基因组学; 重测序; 平行测序; 植物; 群体遗传学

Research progress of plant population genomics based on high-throughput sequencing

Yunsheng Wang

College of Environment and Life Science, Kaili University, Kaili 556011, China

Abstract: Population genomics, a new paradigm for population genetics, combine the concepts and techniques of genomics with the theoretical system of population genetics and improve our understanding of microevolution through identification of site-specific effect and genome-wide effects using genome-wide polymorphic sites genotyping. With the appearance and improvement of the next generation high-throughput sequencing technology, the numbers of plant species with complete genome sequences increased rapidly and large scale resequencing has also been carried out in recent years. Parallel sequencing has also been done in some plant species without complete genome sequences. These studies have greatly promoted the development of population genomics and deepened our understanding of the genetic diversity, level of linking disequilibrium, selection effect, demographical history and molecular mechanism of complex traits of relevant plant population at a genomic level. In this review, I briefly introduced the concept and research methods of population genomics and summarized the research progress of plant population genomics based on high-throughput sequencing. I also discussed

收稿日期: 2016-02-22; 修回日期: 2016-04-10

基金项目: 国家自然科学基金地区项目(编号: 31560091)和贵州省教育厅自然科学基金重点项目(编号: 黔教合 KY 字[2013]186)资助[Supported by the National Natural Science Foundation of China (No. 31560091) and Key Project of Natural Science Foundation of Education Department of Guizhou Province (No. KY[2013]186)]

作者简介: 王云生, 博士, 副教授, 研究方向: 群体遗传学, 基因组学。E-mail: wys3269@126.com

DOI: 10.16288/j.ycz.16-061

网络出版时间: 2016/6/30 10:40:47

URI: <http://www.cnki.net/kcms/detail/11.1913.R.20160630.1040.002.html>

the prospect as well as existing problems of plant population genomics in order to provide references for related studies.

Keywords: genomics; re-sequencing; parallel sequencing; plant; population genetic text

群体基因组学(Population genomics)是 Gulcher 等在 1998 年最先提出的^[1], 随即被广泛接受^[2~4]。21 世纪初, Black 等^[5]开始将群体基因组学视为一门学科, 并初步确定了它的狭义概念: 群体基因组学是群体遗传学一种新的表现形式, 是将基因组概念和技术与群体遗传学理论体系相结合, 通过覆盖全基因组范围内的多态位点的分布式样推测位点特异性效应(选择、突变、选型交配及重组等)和全基因组效应(遗传漂移、迁徙及近亲繁殖), 从而提升人们对微进化(Microevolution)的理解。不久, Luikart 等^[6]提出广义的群体基因组学概念: 通过对全基因组高覆盖度的多态位点同时进行研究, 从而更好地理解在进化过程中影响基因组和种群变异的因素, 如突变、遗传漂移、基因流及自然选择等所扮演的角色。群体基因组学不仅极大提升了推断群体进化历史动态的能力, 也提供了一种研究群体内及群体间的个体在基因组水平上的多样性概貌及等位基因多样性差异的手段^[7]。

随着第二代测序技术的诞生, 测序通量大幅提升, 单位数据量测序价格显著下降, 还允许使用组合的测序策略, 使多个样品可同时在一个反应池中进行测序^[8~10]。在此背景下, 越来越多的植物基因组完成全基因组测序和草图绘制, 同时以模式物种为起点的重测序陆续展开, 使得从真正基因组水平鉴定群体的遗传多样性变为现实, 群体基因组学作

为一门学科也真正发展起来^[7,11~13]。群体基因组学研究流程通常包括以下 4 个主要步骤^[6]: (1) 研究样本的选择应具有较强的代表性, 如果是物种群体, 则需要全地域采样, 同时应有较大的样本数量; (2) 产生具有较高基因组覆盖度的多态位点; (3) 对全部多态位点进行分析, 区分开中性进化的基因位点及受到选择作用的基因位点; (4) 利用中性进化位点推测种群的进行历程, 利用受到选择的基因位点推测种群的适应性进化及性状分化等(图 1)。

1 群体基因组研究高通量测序策略

在研究样本确定之后, 群体基因组学研究最基本的步骤是产生具有较高基因组覆盖度的多态位点, 目前主要有两种方法: 一种是基于分子杂交的芯片测序(Microarray), 这种方法在群体基因组学研究方面有一定的使用率^[14~18], 但芯片杂交测序假阳性率高, 在检测低频突变位点方面受到限制^[19]; 本文综述的主要是第二种方法, 即基于第二代高通量测序方法的植物群体基因组学研究动态。高通量测序主要包括以下 4 种测序策略: 全基因组重测序、外显子测序、限制酶切标签测序以及转录组测序^[20]。

1.1 全基因组重测序

是对已完成基因组序列的同一物种或者近缘种

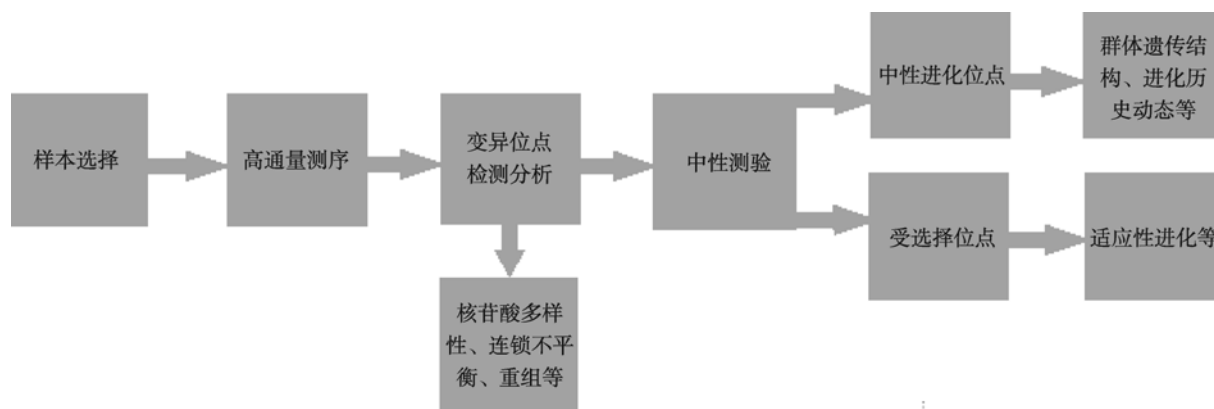


图 1 群体基因组学研究的流程

Fig. 1 The general workflow of population genomics research

的其他个体的进行全基因组测序,新个体基因组序列通过映射参考基因组获得同源数据集,然后检测基因组范围内的单核苷酸多态(Single nucleotide polymorphism, SNP)、插入缺失(Insertion/deletion, InDel)、结构变异(Structural variation, SV)位点及基因拷贝数变异(Copy number variation, CNV)等多态位点^[21]。全基因组重测序方法检测多态位点的质量高度依赖参考基因组的测序质量^[22]。

1.2 外显子重测序

仅针对基因组上编码蛋白质的基因及其相邻的上下游附近序列进行测序,利用参考基因组的外显子序列做探针,捕获待测样本的外显子,建立外显子测序文库,完成测序。目的是研究基因组功能序列的变异式样及其检测这些变异位点与特定表型如疾病等的关联作用^[23],也可以用作其他群体基因组方面的推测^[24,25]。外显子测序目标清晰具体,直接针对最有可能引起样本性状变异的基因组编码及其上下游邻接序列,测序范围压缩到全基因组的 1%~2%^[26,27]。

1.3 限制酶切标签测序(RAD-tag 测序)

通过一对位点特异性限制性酶切将基因组剪切成小片段,然后选择一定长度的酶切片段接上包括测序引物在内的接头,然后进行测序,其原理是根据同一物种或者近缘种不同个体间的基因组序列的高度同源性,其专一性酶切位点在理论上讲也高度一致^[28-30]。由于不同个体间的基因组测序位点同源性高, RAD-tag 测序大大提高了基因组变异的检测效率。

1.4 转录组平行测序

严格意义上的转录组通常是指特定细胞或者细胞类群在特定时间、特定功能状态下所转录出来的全部 mRNA,广义的转录组还包括非翻译的 rRNA、tRNA、ncRNA^[31,32]。目前转录组测序大都采用所谓的 RNA-seq 策略,即将 mRNA 反转录成双链 DNA,然后随机打断成,取一定长度范围的序列构建测序文库并进行高通量测序。如有参考基因组,则将测序数据映射到参考基因上以拼接更为完整的基因序列,对没有参考基因组的,则拼接成一致序列(Unigenes)^[33]。对于同种或者近缘种不同个体尤其是它们的相同组织分别进行 RNA-seq,可得到大量的同源基因序列,通

过同源序列比对,即可检测到大量的变异位点主要是 SNP,再利用 SNP 数据进行群体遗传学分析。RNA-seq 是后基因组时代应用最广泛的技术^[34]。

2 测序数据分析的基本流程

2.1 原始序列统计处理

测序完成后,首先是统计测序量如原始序列(Raw reads)、总碱基数量等,然后是对原始序列去掉测序接头、并进行质量分析、过滤掉低质量的 reads 以及外源污染 reads,得到高质量的纯净序列(Clean reads)进行进一步分析。

2.2 一致性序列组装

对于已有参考基因组的,通过与参考基因组序列的比对分析,利用贝叶斯统计模型检测出每个碱基位点的最大可能性基因型,组装出该个体基因组的 Unigenes,对于没有参考基因组的 RAD-tag 法及转录组测序数据进行同源性分析,去除冗余,获得 Unigenes。

2.3 多态位点的检测

SNP 的检测:提取同源序列中所有多态性位点,结合测序质量、测序深度、重复性等因素作进一步过滤筛选,最终得到可信度高的 SNP 数据集,并根据参考基因组序列对检测到的变异基因位点进行注释,对于无参考基因组的转录组数据则对照 GO、KEGG 等数据库进行注释,无参考基因组的 RAD-tag 法测序数据一般不注释。

InDel 的检测:对于重测序,在进行映射(Mapping)参考基因组时,进行容 Gap 的比对并检测可信的短 InDel,对于无参考基因组的 RAD-tag 法及转录组测序数据使用 BAM 等软件进行同源性分析,通过同源分析确定 Insertion/deletion。

SV 的检测:有参考基因组的重测序及外显子测序能够进行 SV 如插入、缺失、重复、倒位、易位的检测,并能对检测到的 SV 变异进行注释。

CNV 的检测:CNV 是指中等尺度的 DNA 片段的缺失或者扩增,长度介于 50 bp~1 Mb 之间,在重测序中通过合适算法及序列比对结果进行检测^[35-39]。

2.4 群体基因组学常用分析软件

群体基因组学从高通量测序到通过个体间基因

序列的差异来阐明群体的进化历史, 需要进行一系列的生物信息学分析, 而生物信息学分析都是借助各种软件来完成。到目前为止, 生物信息学分析软件大多是基于 linux 操作系统的开源软件, 使用者在网上可以自由下载。从测序原始数据的质量评估、低质量序列的过滤、测序量统计, 序列拼装、同源序列比对、变异位点检出等基本生物信息学分析, 到对变异位点进行中性测验、利用中性变异位点进行群体结构、遗传分化、系谱亲缘关系及复杂性状关联研究等群体遗传学分析的相关软件可谓层出不穷, 部分常用的软件见表 1。

3 植物群体基因组学研究动态

随着第二代测序技术及数据处理分析能力的迅

速提高, 群体基因组学研究已经成为生命科学的热点领域, 相关结果不仅阐明了研究目标群体在基因组的异质性程度及多态样式, 同时加深了人们对目标群体基因组连锁不平衡、重组、选择作用以及种群的遗传结构、栽培种群的驯化起源以及复杂性状的分子机制等的理解。

3.1 基因组水平的核苷酸多样性及变异式样

基因组水平的核苷酸多样性是由 SNPs、InDel、SV 及 CNV 等因素形成, 其高低是一个种群最核心和最基本的遗传性特性, 是种群遗传多样性、有效群体大小(N_e)和进化潜力的重要指标。核苷酸多样性分布式样则是推测种群遗传结构、群体历史、适应性进行及形态、理化性状遗传关联分析等的依据。

表 1 群体基因组学常用分析软件

Table 1 General software for population genomics analysis

| 软件名称 | 用途 | 来源 |
|----------------|--------------|---|
| SolexaQA | 测序质量评估及分析 | http://solexaqa.sourceforge.net/ |
| NGS QC Toolkit | 质量评估过滤 | http://www.nipgr.res.in/ngsqttoolkit.html |
| Phred | 质量评估分析 | http://www.phrap.org/phredphrapconsed.html |
| FASTX-Toolkit | 原始序列预处理 | http://hannonlab.cshl.edu/fastx_toolkit/download.html |
| Velvet | 组装 | http://www.ebi.ac.uk/~zerbino/velvet/ |
| SOAPdenovo | 序列拼装, 过滤 | http://soap.genomics.org.cn/soapdenovo.html |
| Trinity | 序列拼装 | http://trinitymaseq.sourceforge.net |
| Phrap | 序列组装 | http://www.phrap.org/ |
| ABYSS | 序列组装 | http://www.bcgsc.ca/platform/bioinfo/software/abyss |
| Bowtie | 短序列比对 | http://bowtie-bio.sourceforge.net/index.shtml |
| Clustalw | 基因组序列比对 | http://www.ebi.ac.uk/Tools/msa/clustalw2/ |
| MUSCLE | 基因组序列比对 | http://www.ebi.ac.uk/Tools/msa/muscle/ |
| SpeedSeq | 基因组结构变异分析 | http://genome.wustl.edu/people/groups/detail/hall-lab/ |
| Samtools | 基因分型及 SNP 检测 | http://samtools.sourceforge.net/ |
| GATK | 基因分型及 SNP 检测 | https://www.broadinstitute.org/gatk/ |
| FreeBayes | 基因分型 | https://www.msi.umn.edu/sw/freebayes |
| SOAPSnp | SNP 检测 | http://soap.genomics.org.cn/soapsnp.html |
| BayeScan | 中性测验 | http://cmpg.unibe.ch/software/bayescan/ |
| Mega | 系统树构建 | http://megasoftware.net/ |
| frappe | 群体结构和系谱 | http://med.stanford.edu/tanglab/software/frappe.html |
| EIGENSOFT | 主成分分析 | http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm |
| IMPUTE | 全基因组关联分析 | https://mathgen.stats.ox.ac.uk/impute/impute.html |
| PLINK | 全基因组关联分析 | http://pngu.mgh.harvard.edu/~purcell/plink/ |
| R | 数据统计 | https://www.r-project.org/ |

随着大规模重测序及平行测序的开展,一些植物物种主要是作物及其近缘野生种在基因组水平上的变异程度及式样已经被阐明,部分结果见表 2。一般而言,野生种群的基因组核苷酸多态性高于栽培种群,栽培作物的地方品种群体基因组多态性又高于商业品种群体。这反应了作物驯化及育种的过程造成了群体遗传多样性的缩减,同时也从一个侧面说明野生资源具有重要的潜在价值,是栽培物种遗传多样性的重要补充源,应该积极加以保护和利用。

研究还表明,核苷酸多样性在基因组不同区段或者功能域并非均匀分布。例如:蒺藜苜蓿(*Medicago truncatula*)核苷酸多样性从着丝粒到端粒有明显的下降趋势^[50],而拟南芥(*Arabidopsis thaliana*)^[51]、玉米(*Zea mays*)^[47]正相反,越靠近端粒,基因组的核苷酸多样性越高。水稻(*Oryza sativa*)基因组的 SNPs 中有 82% 分布于基因间隔区, 14.4% 分布于内含子区,只有 3.6% 分布于编码区^[44]。芝麻(*Sesamum indicum*)基因组的 SNPs 有 25.1% 位于编码区,其中 5'端非翻译区、编码区和 3'端非翻译区各占基因组总 SNPs

的 1.5%、7.5% 和 1.7%,其余出现在内含子区^[41]。毛果杨(*Populus trichocarpa*)的木质部转录组中,超过 2/3 的 SNPs 位于基因的外显子区,其余 18% 和 14% 分别位于内含子区和非翻译序列区^[52]。水稻基因编码区的无义突变/有义突变的比值为 1.29^[44],而在拟南芥中这一比值是 0.83^[51],在大豆(*Glycine max*)中是 1.36^[45]。番茄(*Solanum section lycopersicon*)果实生长发育基因上的无义突变/有义突变的比值高于其他类型的基因^[53]。

除了 SNPs,基因组上还具有大量的 InDel、SV 及 CNV 变异,这些变异方式在种群间也有较大的差异,在同一种群不同基因组区的分布也是不均匀的。例如:在 50 份水稻的基因组上检测到 94 700 个结构变异位点^[44]。302 份大豆群体中检测出 876 799 个 InDels (6 bp)、1614 个 CNVs 和 6388 个 SVs^[46]。玉米基因组的结构变异也普遍存在,而且在重要的农艺性状关联位点富集^[54]。芝麻基因组中 InDels 中,71.7% 位于基因间隔区,1.5% 位于编码区,5.0% 位于非翻译区,其多样性分布态势与 SNP 分布态势有强烈的相关性^[41]。对 115 份黄瓜(*Cucumis sativas*)进行深度重测序,结果表明黄瓜基因组的 SVs 大部分产生于非同源末端连接重排,通常出现高核苷酸多态区域^[55]。

3.2 连锁不平衡与重组

连锁不平衡(Linkage disequilibrium, LD)指的是一个群体内不同座位等位基因之间的非随机关联,LD 水平对自然群体复杂性状的关联分析具有重要的参考价值,是群体基因组学研究的重要方面^[56]。种群 LD 水平通常用基因组全部成对多态位点的平均关联值 r^2 (Squared allele-frequency correlations) 从最大值降低到的一半或者某一阈值时的物理距离来衡量。需要指出的是,种群 LD 水平与研究样本的数量、地理来源及样本间亲缘关系等有密切的关系。通常来说,研究样本数量越少,地理来源、亲缘系谱越近,LD 水平越高,反之越低。重测序研究表明:芝麻(29 份)LD 水平($r^2 = 0.15$)约为 150 kb^[41],谷子(*Setaria italica*) (916 份) LD 水平($r^2 = 0.25$)约为 100 kb^[42]。玉米自交系(27 份) LD 水平($r^2 = 0.1$)一般在 2 kb 以内^[47]。栽培玉米(83 份) LD 水平($r^2 = 0.2$)约为 5.5 kb,野生玉米(*Zea mays* subsp. *parviglumis*)

表 2 部分植物基因组的核苷酸多样性

Table 2 Nucleotide diversity in some plant genomes

| 物种 | 核苷酸多样性 | | 参考 文献 |
|--|-----------------------------|----------------------------|----------|
| | $\theta\pi(\times 10^{-3})$ | $\theta w(\times 10^{-3})$ | |
| 欧洲山杨(<i>Populus tremula</i>) | 13.3 | | [40] |
| 美洲山杨(<i>Populus tremuloides</i>) | 14.4 | | [40] |
| 毛果杨(<i>Populus trichocarpa</i>) | 5.9 | | [40] |
| 芝麻(<i>Sesamum indicum</i>) | 2.5 | 3.0 | [41] |
| 谷子(<i>Foxtail millet</i>) | 1.0 | | [42] |
| 鹰嘴豆(<i>Cicer arietinum</i>) | 2.0 | 1.8 | [43] |
| 多年生普通野生稻(<i>Oryza rufipogo</i>) | 7.2 | 10.6 | [44] |
| 一年生普通野生稻(<i>Oryza nivara</i>) | 6.3 | 9.4 | [44] |
| 粳稻(<i>Oryza sativa</i> subsp. <i>japonica</i>) | 3.7 | 5.5 | [44] |
| 籼稻(<i>Oryza sativa</i> subsp. <i>indica</i>) | 5.7 | 6.8 | [44] |
| 野生大豆(<i>Glycine soja</i>) | 2.9 | | [45] |
| 大豆地方品种(<i>Glycine max</i>) | 1.4 | | [46] |
| 大豆改良品种(<i>Glycine max</i>) | 1.1 | | [46] |
| 玉米地方品种(<i>Zea mays</i>) | 6.6 | | [47] |
| 玉米改良品种(<i>Zea mays</i>) | 3.0 | | [48] |
| 野生西瓜(<i>Citrullus vulgaris</i>) | 7.6 | 6.4 | [49] |
| 栽培西瓜(<i>Citrullus lanatus</i>) | 1.4 | 1.5 | [49] |
| 蒺藜苜蓿(<i>Medicago truncatula</i>) | 4.3 | 6.3 | [50] |

(19 份)仅仅达到 1 kb^[54]。不同水稻种群 LD(r^2 衰减到最大值的一半)差异很大: 多年生野生稻(5 份)及一年生野生稻(5 份)在 10 kb 以内, 粘稻(9 份)大约为 65 kb, 粳稻(25 份)达到 200 kb^[44]。但对地理来源广泛的 446 份普通野生稻(*Oryza rufipogon*)和 1083 份粳稻和籼稻的重测序分析结果表明: LD 水平(r^2 衰减到最大值的一半)在普通野生稻(*O. rufipogon*)中为 20 kb, 籼稻为 123 kb, 粳稻为 167 kb^[57]。深度重测序分析表明: 野生大豆 LD(14 份)水平(r^2 衰减到最大值的一半)超过 75 kb, 栽培大豆(17 份)超过 150 kb^[45]。但更多的样本重测序研究显示: 野生大豆(62 份)的 LD 水平(r^2 衰减到最大值的一半)大约 27 kb, 栽培大豆地方品种(130 份)和改良品种群体(110 份)分别为 83 kb 和 133 kb^[46]。对黄瓜核心资源的重测序分析显示: 印度黄瓜(30 份)LD 水平($r^2 = 0.2$)仅为 3.2 kb, 东亚黄瓜(37 份)、欧亚黄瓜(29 份)、西双版纳黄瓜(19 份)均超过 55 kb, 最高达到 140 kb^[58]。

重组是指真核生物细胞减数分裂期姊妹染色体单体上的同源基因发生交换, 从而产生重组型配子的遗传学现象, 是物种繁衍过程增加遗传多样性的一个重要途径。重组打断了连锁, 因而对 LD 水平高低的影响大。研究表明: 基因组上存在所谓重组热点及冷点。例如: 对 26 份蒺藜苜蓿的全基因组重测序分析发现: 高重组区域位于靠近着丝粒区域, 并且具有更高的基因密度和更低的 GC 含量^[59]。猴面花(*Mimulus guttatus*)基因组存在约 13 000 重组热点, 这些热点与重组冷点(非重组区)交织分布。在基因区, 重组率表现出一定的极性, 在起始位点出现峰值, 然后迅速衰退。基因组各区域的重重组率差异显著, 即使在热点之间也可能相差几百倍^[60]。

3.3 选择作用

群体基因组学研究关键的目标之一是鉴定选择作用位点^[61]。从施加选择作用的外部力量来看, 选择作用包括两种类型: 一是自然选择作用, 主要是各种环境、生态因子对野生群体和栽培群体的选择作用, 自然选择作用主要发现于响应环境或者抗逆基因位点, 例如: 重测序研究鉴定到拟南芥存在多个受到地形-气候等环境因子如湿度、光照强度及土壤类型的选择作用的基因位点, 功能注释表明: 这些基因与水分代谢、感光生理及重金属解毒、钙镁

运输等功能密切相关^[62,63]。对猴面花的研究表明: 新的环境对引进群体的选择作用主要作用于开花时间、生态压力反应、营养转运和抗冻等基因^[64]。另一种是人工选择作用, 主要是人类早期对作物的驯化及现代育种改良活动对特定性状的持续强化, 主要作用于农艺、形态性状, 通常会造成栽培群体的相应位点及其邻接区遗传多样性急剧减少, 即形成一段所谓选择清除(Selective sweep)基因区, 同时该基因区与野生群体的遗传分化程度增加。重测序研究表明: 栽培水稻^[49,57]、玉米^[65]、大豆^[45,46]等基因组都存在大量的选择性清除基因区, 几乎所有与驯化性状相关的基因位点都位于这些区域, 如水稻直立生长(*prog1*)、落粒(*sh4*)、外壳颜色(*Bh4*)、谷粒宽度(*qSW5*)及叶鞘、叶尖色(*OsCI*)等基因^[44,57]; 大豆的花期、千粒重及种子硬度、茎的缠绕性及豆荚开裂性等 QTL 位点^[46]; 玉米的影响叶序的基因(*abph1*)、影响多个农艺性状的编码谷氨酰胺合成酶基因(*GRMZM2G036464*)及硝酸还原酶基因(*GRMZM2G428027*)等^[65]; 菜豆(*Phaseolus vulgaris*)的春化途径基因(*VRN1* 和 *VRN2*)、开花基因(*FRL1* 和 *TFL2*)等^[66]; 桃的食用品质(147 个)和风味品质(262 个)相关的基因位点^[67]等。除了古老的驯化作用, 现代人工育种改良也在部分商业品种上留下了明显的选择痕迹, 如玉米的淀粉合成关键基因(*bt2*)、甜味基因(*su1*)、赤霉素氧化酶基因(*GRMZM2G152354*、*GRMZM2G036340*)等^[68]; 菜豆粒重基因^[66]等。

从植物种群应对自然选择的适应性方式来看, 选择作用又包括正选择(Positive selection)、歧化选择(Diversifying selection)、负选择(Negative selection)。重测序研究分析表明: 调控基因表达相关的歧化选择对毛果杨的适应性十分重要, 在其基因组上测验的 SNPs 中, 有 2.1% 显示出歧化选择的分布式样^[69]。水稻地方品种中有 11 个基因家族显示出正选择迹象, 其中包括 NB-LRR 抗性蛋白基因家族^[70]。黄瓜核心种质的基因组中存在大量的歧化选择作用位点, 包括一些抗病基因^[58]。番茄水果生长分化基因表现出明显的正选择迹象^[53]。蒺藜苜蓿的基因组整体上表现出明显的负选择迹象^[50]。

3.4 种群遗传结构

群体遗传结构是指一个大群体中的地理、生态

来源不同的亚群间在基因频率和基因型频率发生显著差异的现象,反映了种群在漫长的进化过程或者人为选择的过程中产生了地理、生态适应性分化,群体遗传结构是推断种群历史动态如群体大小、地理分布变迁、基因流、遗传分化等的重要依据。与传统的分子群体遗传学相比较,高通量重测序或者平行测序研究样本通常有限,但也有部分研究进行了大样本分析。群体基因组研究表明,北美毛果杨在多个空间尺度呈现明显的地理分化,沿纬度梯度形成了明显的群体亚结构,推测遗传漂变在毛果杨较近的进化历史中具有重要的作用^[69,71]。不同地域的拟南芥群体存在显著的遗传分化,分化指数 F_{ST} 为 0.06~0.21^[72]。野生黄瓜(印度组)则显示出明显的亚结构和遗传分化^[58]。野生水稻(*O. rufipogon*)可分为 3 个遗传分支,显示强烈的地理分化^[57]。粳稻和籼稻可进一步分为 3 个遗传亚组,分析认为光、温是驱动中国栽培水稻分化的主要生态因子^[70]。聚类分析显示栽培大豆分为两大遗传支系,各自又可细分为 3~4 个遗传分支,呈现出强烈的地理生态相近性^[46]。谷子按照花期可分为 2 个遗传分支:春播型和夏播型,中国北方和西北部的高纬度地区的大部分地方品种属于春播型,而中南部地方品种属于夏播型^[42]。

3.5 栽培作物的驯化

自然物种的形成和进化是一个漫长的过程,大多数在百万年以上。然而,栽培物种进化历史不过万年,短的只有几十年,其直接野生祖先种易被确定。对于物种起源进化的历史动态研究而言,祖先野生种-栽培种复合群体是一个十分有价值的研究系统,因而作物的驯化也就成了群体基因组学研究的重点。同时,利用基因组水平的变异对作物驯化事件进行推断,克服了分子群体遗传学(分子标记或者基因位点数目有限)研究中在推断种群内个体间的系谱亲缘关系上可信度不高的弊端。群体基因组研究表明,粳稻首先驯化于中国南方珠江流域中部(广西)某一个群体,籼稻来自于随后粳稻与当地野生稻的杂交,然后向东南亚和南亚扩散^[57]。这种先形成一个栽培种,然后栽培种与野生种回交形成另一个栽培种的现象,还发生在柚、甜橙的起源上^[73]。栽培黄瓜^[58]、大豆^[46]来源于单个驯化事件,栽培菜

豆则分别独立驯化于中美洲和安第斯山脉^[66]。同时研究表明,作物大多在驯化过程中经历了遗传瓶颈,即比较野生祖先种群、栽培种群发生了基因组范围内的遗传多样性降低,除玉米^[65]等极少数作物种类外,多数作物种类表现出较为严重的遗传瓶颈^[44,46,53,58,67,74]。除了普遍性降低的遗传瓶颈,在受到驯化选择的基因组区特别是驯化性状(产量、品质、风味、形态等)相关的基因所处的位置及其邻接序列还出现了所谓的的选择性清除区^[44,46,53,65,58,67]。

3.6 复杂性状的全基因组关联分析

全基因组关联分析(Genome wide association study, GWAS)是 Risch 等^[75]最先在 1996 年提出的一种研究思路。它以 LD 为基础,通过识别成百上千个的定位群体(主要是自然群体)中基因组范围内的高密度分子标记,筛选出与复杂性状表现型变异相关联的分子标记,进而分析这些分子标记对表现型的遗传效应^[76]。GWAS 研究能够充分利用群体进化过程中数千世代发生的重组事件,通过打破 LD 或者一段染色体区域内多态性位点的相关性来增加复杂性状的定位效率^[77]。GWAS 是近年来群体基因组学研究的热点,GWAS 已经在人类身高、体重及癌症、糖尿病、肥胖等超过 500 个复杂的遗传性疾病中检测到关联的变异位点 6000 个以上^[78],在动物复杂性状的研究也取到巨大的成功^[79,80]。全基因组关联分析在植物研究中目前主要应用于模式植物和重要经济作物中,关联研究的性状包括形态、农艺及生理等性状。例如大豆的茎干直立程度、粒重、花色、种皮颜色、绒毛形态、抗胞囊线虫病及脐色等^[46],菜豆的种子大小及粒重^[66],谷子的花期、分蘖数、单株总穗数及果壳颜色^[42],水稻的形态性状(分蘖数及叶片着生角度)、产量成分性状(粒宽、粒长、粒重及小穗数)、谷物品质性状(糊化温度及直链淀粉含量)、颜色性状(稃尖颜色、果皮颜色及颖壳颜色)和生理性状(抽穗期、耐旱性及精米率)^[70],玉米的棒色、花丝颜色、花期及抗旱生理^[68,81],高粱的株高及花序形态^[82],拟南芥的蛇纹岩土壤的适应生理^[63]等。

4 展 望

群体基因组学是伴随着第一代基因组测序技术

的兴起而诞生,随着第二代高通量测序技术的出现而兴盛,也必将随着更先进的测序如第三代单分子测序技术得到进一步发展。而其中的关键就是测序的通量越来越高,获得单位数据量的费用越来越低,正在向普通实验室和科技工作者所能承受的范围接近。可以预见,过去基于分子标记的群体遗传学研究将越来越多被基因组重测序或者平行测序的策略所取代,群体基因组学研究重点今后将逐渐从模式植物种转向非模式植物种。同时,随着群体基因组学研究的展开,人们对植物种群的起源扩散、群体动态、适应性分化等进化历史将会有更加详细深入的了解,对物种性状表现的遗传和分子机理将有更为深刻和精细的理解。然而,对大多数普通研究者而言,要解决大数据的处理和分析仍然具有较大的困难,从而不得不求助于专门的生物信息分析员,这样不仅大幅度增加了研究费用,也造成数据处理和分析不彻底。要解决这个问题,需要多方努力,对群体遗传学研究人员而言,需要努力学习基于大中型服务器的操作系统如 linux 及相应应用软件安装、测试及操作能力,需要学会基础的编程知识,需要学会应用大型统计软件如 R、SAS 等。而对其他相关的科技工作者,开发界面更加友好、功能更加强大、能进行程式化操作的大型分析软件将对群体基因组学研究有极大促进作用。

参考文献(References):

- [1] Gulcher J, Stefansson K. Population genomics: Laying the groundwork for genetic disease modelling and targeting. *Clin Chem Lab Med*, 1998, 36(8): 523–527. [DOI]
- [2] Goldstein DB, Weale ME. Population genomics: Linkage disequilibrium holds the key. *Curr Biol*, 2001, 11(14): R576–R579. [DOI]
- [3] Jorde LB, Watkins WS, Bamshad MJ. Population genomics: A bridge from evolutionary history to genetic medicine. *Hum Mol Genet*, 2001, 10(20): 2199–2207. [DOI]
- [4] Gibson G, Mackay TF. C. Enabling population and quantitative genomics. *Genet Res*, 2002, 80(1): 1–6. [DOI]
- [5] Black WC, Baer CF, Antolin MF, DuTeau NM. Population genomics: Genome-wide sampling of insect populations. *Annu Rev Entomol*, 2001, 46(1): 441–469. [DOI]
- [6] Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: From genotyping to genome typing. *Nat Rev Genet*, 2003, 4(12): 981–994. [DOI]
- [7] Ellegren H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol*, 2014, 29(1): 51–63. [DOI]
- [8] Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet*, 2007, 24(3): 133–141. [DOI]
- [9] Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods*, 2008, 5(10): 887–893. [DOI]
- [10] Cronn R, Liston A, Parks M, Gernandt DS, Shen RK, Mockler T. Multiplex sequencing of plant chloroplast genomes using solexa sequencing-by-synthesis technology. *Nucleic Acids Res*, 2008, 36(19): e122. [DOI]
- [11] Stratton M. Genome resequencing and genetic variation. *Nat Biotechnol*, 2008, 26(1): 65–66. [DOI]
- [12] Xu XY, Bai GH. Whole-genome resequencing: Changing the paradigms of SNP detection, molecular mapping and gene discovery. *Mol Breeding*, 2015, 35: 33. [DOI]
- [13] Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, Pachter L, Myers E, Langley CH. Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol*, 2007, 5(11): e310. [DOI]
- [14] Yu P, Wang CH, Xu Q, Feng Y, Yuan XP, Yu HY, Wang YP, Tang SX, Wei XH. Detection of copy number variations in rice using array-based comparative genomic hybridization. *BMC Genomics*, 2011, 12: 372. [DOI]
- [15] McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddalo JA, Stupar RM. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol*, 2012, 159(4): 1295–1308. [DOI]
- [16] Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A. Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol*, 2012, 12: 16. [DOI]
- [17] Nimmakayala P, Levi A, Abburi L, Abburi VL, Tomason YR, Saminathan T, Vajja VG, Malkaram S, Reddy R, Wehner TC, Mitchell SE, Reddy UK. Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC Genomics*, 2014, 15: 767. [DOI]
- [18] Wen ZX, Boyse JF, Song QJ, Cregan PB, Wang DC. Ge-

- onomic consequences of selection and genome-wide association mapping in soybean. *BMC Genomics*, 2015, 16: 671. [DOI]
- [19] Wu J, Lenchik NI, Gerling IC. Approaches to reduce false positives and false negatives in the analysis of microarray data: applications in type 1 diabetes research. *BMC Genomics*, 2008, 9(Suppl. 2): S12. [DOI]
- [20] Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*, 2011, 12(7): 499–510. [DOI]
- [21] Huang XH, Feng Q, Qian Q, Zhao Q, Wang L, Wang AH, Guan JP, Fan DL, Weng QJ, Huang T, Dong GJ, Sang T, Han B. High-throughput genotyping by whole-genome resequencing. *Genome Res*, 2009, 19(6): 1068–1076. [DOI]
- [22] Imelfort M, Edwards D. De novo sequencing of plant genomes using second-generation technologies. *Brief Bioinform*, 2009, 10(6): 609–618. [DOI]
- [23] Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*, 2011, 12(11): 745–755. [DOI]
- [24] Li YR, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao HZ, Korneliussen T, Grarup N, Guo YR, Hellman I, Jin X, Li QB, Liu JT, Liu X, Sparsø T, Tang MF, Wu HL, Wu RH, Yu C, Zheng HC, Astrup A, Bolund L, Holmkvist J, Jørgensen T, Kristiansen K, Schmitz O, Schwartz TW, Zhang XQ, Li RQ, Yang HM, Wang J, Hansen T, Pedersen O, Nielsen R, Wang J. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet*, 2010, 42(11): 969–972. [DOI]
- [25] Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I. Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat Biotechnol*, 2012, 30(8): 798–802. [DOI]
- [26] Hodges E, Xuan ZY, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective. *Nat Genet*, 2007, 39(12): 1522–1527. [DOI]
- [27] Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome sequencing: Current and future perspectives. *G3 (Bethesda)*, 2015, 5(8): 1543–1550. [DOI]
- [28] Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*, 2007, 17(2): 240–248. [DOI]
- [29] Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 2008, 3(10): e3376. [DOI]
- [30] Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: An inexpensive method for De Novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 2012, 7(5): e37135. [DOI]
- [31] Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10(1): 57–63. [DOI]
- [32] Costa V, Angelini C, De Feis I, Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol*, 2010, 2010: 853916. [DOI]
- [33] Mutz KO, Heilkenbrinker A, Lönne M, Walter JG, Stahl F. Transcriptome analysis using next-generation sequencing. *Cur Opin Biotech*, 2013, 24(1): 22–30. [DOI]
- [34] Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature*, 2000, 405(6788): 827–836. [DOI]
- [35] Shen YF, Wan ZZ, Coarfa C, Drabek R, Chen L, Ostrowski EA, Liu Y, Weinstock GM, Wheeler DA, Gibbs RA, Yu FL. next-generation resequencing data A SNP discovery method to assess variant allele probability from next-generation resequencing data. *Genome Res*, 2010, 20(2): 273–280. [DOI]
- [36] O'Neil ST, Dzurisin JDK, Carmichael RD, Lobo NF, Emrich SJ, Hellmann JJ. RPeoaseprcu h alratictleion-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. *BMC Genomics*, 2010, 11: 310. [DOI]
- [37] Davey JW, Blaxter ML. RADSeq: Next-generation population genetics. *Brief Funct Genomics*, 2010, 9(5–6): 416–423. [DOI]
- [38] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 2011, 43(5): 491–498. [DOI]
- [39] Hohenlohe PA, Catchen J, Cresko WA. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. In: Pompanon F, Bonin A, eds. *Data Production and Analysis in Population Genomics: Methods in Molecular Biology*. New York: Humana Press, 2012, 888: 235–260. [DOI]

- [40] Wang J, Street NR, Scofield DG, Ingvarsson PK. Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related *Populus* Species. *Genetics*, 2015, 202(3): 1185–1200. [DOI]
- [41] Wang LH, Han XL, Zhang YX, Li DH, Wei X, Ding X, Zhang XR. Deep resequencing reveals allelic variation in *Sesamum indicum*. *BMC Plant Biol*, 2014, 14: 225. [DOI]
- [42] Jia GQ, Huang XH, Zhi H, Zhao Y, Zhao Q, Li WJ, Chai Y, Yang LF, Liu KY, Lu HY, Zhu CR, Lu YQ, Zhou CC, Fan DL, Wg QJ, Guo YL, Huang T, Zhang L, Lu TT, Feng Q, Hao HF, Liu HK, Lu P, Zhang N, Li YH, Guo EH, Wang SJ, Wang SY, Liu JR, Zhang WF, Chen GQ, Zhang BJ, Li W, Wang YF, Li HQ, Zhao BH, Li JY, Diao XM, Han B. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat Genet*, 2013, 45(8): 957–964. [DOI]
- [43] Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol*, 2013, 31(3): 240–246. [DOI]
- [44] Xu X, Liu X, Ge S, Jensen JD, Hu FY, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li JX, He WM, Zhang GJ, Zheng XM, Zhang FM, Li YR, Yu C, Kristiansen K, Zhang XQ, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying genomically important genes. *Nat Biotechnol*, 2012, 30(1): 105–111. [DOI]
- [45] Lam HM, Xu X, Liu X, Chen WB, Yang GH, Wong FL, Li MW, He WM, Qin N, Wang B, Li J, Jian M, Wang J, Shao GH, Wang J, Sun SSM, Zhang GY. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet*, 2010, 42(12): 1053–1059. [DOI]
- [46] Zhou ZK, Jiang Y, Wang Z, Gou ZH, Lyu J, Li WY, Yu YJ, Shu LP, Zhao YJ, Ma YM, Fang C, Shen YT, Liu TF, Li CC, Li Q, Wu M, Wang M, Wu YS, Dong Y, Wan WT, Wang X, Ding ZL, Gao YD, Xiang H, Zhu BG, Lee SH, Wang W, Tian ZX. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol*, 2015, 33(4): 408–416. [DOI]
- [47] Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES. A first-generation haplotype map of maize. *Science*, 2009, 326(5956): 1115–1117. [DOI]
- [48] Lai JS, Li RQ, Xu X, Jin WW, Xu ML, Zhao HN, Xiang ZK, Song WB, Ying K, Zhang M, Jiao YP, Ni PX, Zhang JG, Li D, Guo XS, Ye KX, Jian M, Wang B, Zheng HS, Liang HQ, Zhang XQ, Wang SC, Chen SJ, Li JS, Fu Y, Springer NM, Yang HM, Wang J, Dai JR, Schnable PS, Wang J. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet*, 2010, 42(11): 1027–1030. [DOI]
- [49] Guo SG, Zhang JG, Sun HH, Salse J, Lucas WJ, Zhang HY, Zheng Y, Mao LY, Ren Y, Wang ZW, Min JM, Guo XS, Murat F, Ham BK, Zhang ZL, Gao S, Huang MY, Xu YM, Zhong SL, Bombarely A, Mueller LA, Zhao H, He HJ, Zhang Y, Zhang ZH, Huang SW, Tan T, Pang EL, Lin K, Hu Q, Kuang HH, Ni PX, Wang B, Liu J, Kou QH, Hou WJ, Zou XH, Jiang J, Gong GY, Klee K, Schoof H, Huang Y, Hu XS, Dong SS, Liang DQ, Wang J, Wu K, Xia Y, Zhao X, Zheng ZQ, Xing M, Liang XM, Huang BQ, Lv T, Wang JY, Yin Y, Yi HP, Li RQ, Wu MZ, Levi A, Zhang XP, Giovannoni JJ, Wang J, Li YF, Fei ZJ, Xu Y. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet*, 2013, 45(1): 51–58. [DOI]
- [50] Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Bharti AK, Woodward JE, May GD, Gentzbittel L, Ben C, Denny R, Sadowsky MJ, Ronfort J, Bataillon T, Young ND, Tiffin P. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc Natl Acad Sci USA*, 2011, 108(42): E864–E870. [DOI]
- [51] Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen HM, Frazer KA, Huson DH, Schölkopf B, Nordborg M, Rätsch G, Ecker JR, Weigel D. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, 2007, 317(5836): 338–342. [DOI]
- [52] Geraldes A, Pang J, Thiessen N, Cezard T, Moore R, Zhao YJ, Tam A, Wang SC, Friedmann M, Birol I, Jones SJM, Cronk QCB, Douglas CJ. SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. *Mol Ecol Resour*, 2011, 11(Suppl. 1): 81–92. [DOI]
- [53] The 100 Tomato Genome Sequencing Consortium, Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, Wang J, Zhang GY, Li N, Mao LK, Bakker F, Dirks R, Breit T, Gravendeel B, Huits H, Struss D, Swason-Wagner R, van Leeuwen H, van Ham RCHJ, Fito L, Guigner L, Sevilla M, Ellul P, Ganko EW, Kapur A, Reclus E, de Geus B, van de Geest H, te Lintel Hekkert B, van Haarst

- JC, Smits L, Koops A, Sanchez-Perez GS, van Heusden AW, Visser R, Quan ZW, Min JM, Liao L, Wang XL, Wang GB, Yue Z, Yang XH, Xu N, Schranz E, Smets EF, Vos RA, Rauwerda J, Ursem R, Schuit C, Kerns M, van den Berg J, Vriezen WH, Janssen A, Datema E, Jahrman T, Moquet F, Bonnet J, Peters SA. Exploring genetic variation in the tomato (*Solanum section Lycopersicon*) clade by whole-genome sequencing. *Plant J*, 2014, 80(1): 136–148. [DOI]
- [54] Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai JS, Li M, Liu X, Lu YL, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong TZ, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang ZW, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang GY, Xu YB, Ware D. Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet*, 2012, 44(7): 803–807. [DOI]
- [55] Zhang ZH, Mao LY, Chen HM, Bu FJ, Li GC, Sun JJ, Li S, Sun HH, Jiao C, Blakely R, Pan JS, Cai R, Luo RB, van de Peer Y, Jacobsen E, Fei ZJ, Huang SW. Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell*, 2015, 27(6): 1595–1604. [DOI]
- [56] Gupta PK, Rustgi S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol Biol*, 2005, 57(4): 461–485. [DOI]
- [57] Huang XH, Kurata N, Wei XH, Wang ZX, Wang AH, Zhao Q, Zhao Y, Liu KY, Lu HY, Li WJ, Guo YL, Lu YQ, Zhou CC, Fan DL, Weng QJ, Zhu CR, Huang T, Zhang L, Wang YC, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan XP, Xu Q, Dong GJ, Zhan QL, Li CY, Fujiyama A, Toyoda A, Lu TT, Feng Q, Qian Q, Li JY, Han B. A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 2012, 490(7421): 497–501. [DOI]
- [58] Qi JJ, Liu X, Shen D, Miao H, Xie BY, Li XX, Zeng P, Wang SH, Shang Y, Gu XF, Du YC, Li Y, Lin T, Yuan JH, Yang XY, Chen JF, Chen HM, Xiong XY, Huang K, Fei ZJ, Mao LY, Tian L, Städler T, Renner SS, Kamoun S, Lucas WJ, Zhang ZH, Huang SW. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat Genet*, 2013, 45(12): 1510–1518. [DOI]
- [59] Paape T, Zhou P, Branca A, Briskine R, Young N, Tiffin P. Fine-scale population recombination rates, hotspots, and correlates of recombination in the *Medicago truncatula* Genome. *Genome Biol Evol*, 2012, 4(5): 726–737. [DOI]
- [60] Hellsten U, Wright KM, Jenkins J, Shu SQ, Yuan YW, Wessler SR, Schmutz J, Willis JH, Rokhsar DS. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc Natl Acad Sci USA*, 2013, 110(48): 19478–19482. [DOI]
- [61] Buerkle CA, Gompert Z, Parchman TL. The $n = 1$ constraint in population genomics. *Mol Ecol*, 2011, 20(8): 1575–1581. [DOI]
- [62] Fischer MC, Rellstab C, Tedder A, Zoller S, Gugerli F, Shimizu KK, Holderegger R, Widmer A. Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol Ecol*, 2013, 22(22): 5594–5607. [DOI]
- [63] Turner TL, Bourne EC, von Wettberg EJ, Hu TT, Nuzhdin SV. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet*, 2010, 42(3): 260–263. [DOI]
- [64] Puzey J, Vallejo-Marín M. Genomics of invasion: Diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Mol Ecol*, 2014, 23(18): 4472–4485. [DOI]
- [65] Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai JS, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang GY, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J. Comparative population genomics of maize domestication and improvement. *Nat Genet*, 2012, 44(7): 808–811. [DOI]
- [66] Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu SQ, Song QJ, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao DY, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia GF, Kelly JD, Kudrna D, Lee R, Richard MMS, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, Jackson SA. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet*, 2014, 46(7): 707–713. [DOI]
- [67] Cao K, Zheng ZJ, Wang LR, Liu X, Zhu GR, Fang WC, Cheng SF, Zeng P, Chen CW, Wang XW, M, Zhong X, Wang XL, Zhao P, Bian C, Zhu YL, Zhang JH, Ma GS, Chen CX, Li YJ, Hao FG, Li Y, Huang GD, Li YX, Li HY, Guo J, Xu X, Wang J. Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops.

- Genome Biol*, 2014, 15(7): 415. [DOI]
- [68] Jiao YP, Zhao HN, Ren LH, Song WB, Zeng B, Guo JJ, Wang BB, Liu ZP, Chen J, Li W, Zhang M, Xie SJ, Lai JS. Genome-wide genetic changes during modern breeding of maize. *Nat Genet*, 2012, 44(7): 812–815. [DOI]
- [69] Zhou L, Bawa R, Holliday JA. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (*Populus trichocarpa*). *Mol Ecol*, 2014, 23(10): 2486–2499. [DOI]
- [70] Huang XH, Wei XH, Sang T, Zhao Q, Feng Q, Zhao Y, Li CY, Zhu CR, Lu TT, Zhang ZW, Li M, Fan DL, Guo YL, Wang AH, Wang L, Deng LW, Li WJ, Lu YQ, Weng QJ, Liu KY, Huang T, Zhou TY, Jing YF, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li JY, Han B. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 2010, 42(11): 961–969. [DOI]
- [71] Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA, Gunter LE, Ranjan P, Vining K, Pomraning KR, Wilhelm LJ, Pellegrini M, Mockler TC, Freitag M, Gerald A, El-Kassaby YA, Mansfield SD, Cronk QCB, Douglas CJ, Strauss SH, Rokhsar D, Tuskan GA. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytol*, 2012, 196(3): 713–725. [DOI]
- [72] Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet*, 2011, 43(10): 956–963. [DOI]
- [73] Xu Q, Chen LL, Ruan XA, Chen DJ, Zhu AD, Chen CL, Bertrand D, Jiao WB, Hao BH, Lyon MP, Chen JJ, Gao S, Xing F, Lan H, Chang JW, Ge XH, Lei Y, Hu Q, Miao Y, Wang L, Xiao SX, Biswas MK, Zeng WF, Guo F, Cao HB, Yang XM, Xu XW, Cheng YJ, Xu J, Liu JH, Luo OJ, Tang ZH, Guo WW, Kuang HH, Zhang HY, Roose ML, Nagarajan N, Deng XX, Ruan YJ. The draft genome of sweet orange (*Citrus sinensis*). *Nat Genet*, 2013, 45(1): 59–66. [DOI]
- [74] Wang YS, Lin SQ. Resequencing 10 accession wild and 10 accession cultivated loquat (*Eriobotrya japonica*): SNP marker explored and used for preliminary analysis of population genomics. In: Proceedings of the Sixth National Symposium on Loquat (Abstracts). Beijing: Chinese Society for Horticultural Science, 2013: 34–40.
- 王云生, 林顺权. RAD-tag 重测序 10 份野生和 10 份栽培枇杷: SNP 标记开发及初步群体基因组学分析. 见: 第六届全国枇杷学术研讨会论文(摘要)集. 北京: 中国园艺学会, 2013: 34–40. [DOI]
- [75] Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*, 1996, 273(5281): 1516–1517. [DOI]
- [76] Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 2005, 6(2): 95–108. [DOI]
- [77] Flint-Garcia SA. Genetics and consequences of crop domestication. *J Agric Food Chem*, 2013, 61(35): 8267–8276. [DOI]
- [78] Robinson MR, Wray NR, Visscher PM. Explaining additional genetic variation in complex traits. *Trends Genet*, 2014, 30(4): 124–132. [DOI]
- [79] Zhang H, Wang ZP, Wang SZ, Li H. Progress of genome wide association study in domestic animals. *J Anim Sci Biotechnol*, 2012, 3: 26. [DOI]
- [80] Zhou JP, Pei ZY, Chen YB, Chen RS. Strategies of genome-wide association study based on high-throughput sequencing. *Hereditas (Beijing)*, 2014, 36(11): 1099–1111. 周家蓬, 裴智勇, 陈禹保, 陈润生. 基于高通量测序的全基因组关联研究策略. *遗传*, 2014, 36(11): 1099–1111. [DOI]
- [81] Xu J, Yuan YB, Xu YB, Zhang GY, Guo XS, Wu FK, Wang Q, Rong TZ, Pan GT, Cao MJ, Tang QL, Gao SB, Liu YX, Wang J, Lan H, Lu YL. Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC Plant Biol*, 2014, 14: 83. [DOI]
- [82] Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci USA*, 2013, 110(2): 453–458. [DOI]