

作物基因组学研究进展

唐 丁¹, 吕慧颖¹, 王 珏¹, 葛毅强², 魏 珣², 杨维才¹, 程祝宽¹

(¹中国科学院遗传与发育生物学研究所/种子创新研究院, 北京 100101; ²中国农村技术开发中心, 北京 100045)

摘要:农作物基因组学研究的发展,对于有效利用现代分子生物学手段进行物种的遗传改良发挥了重要作用。随着测序技术的发展,已经实现对重要农作物,如水稻、小麦、玉米、大豆、油菜、棉花、蔬菜等作物基因组的测序或重测序,在此基础上完成对控制重要农艺性状基因的克隆和鉴定。本文综述了2017年度主要农作物基因组研究方面取得的一系列重要进展。

关键词:基因组;农艺性状;农作物

Progresses on the Major Crop Genomics

TANG Ding¹, LV Hui-ying¹, WANG Jue¹, GE Yi-qiang², WEI Xun², YANG Wei-cai¹, CHENG Zhu-kuan¹

(¹Institute of Genetics and Developmental Biology, Chinese Academy of Sciences/Innovative Academy of Seed Design, Beijing 100101; ²China Rural Technology Development Center, Beijing 100045)

Abstract: With the completion of genome sequencing on major crops, such as rice, wheat, maize, soybean, rape, cotton and vegetables, the molecular biological techniques are becoming more important and widely used in their genetic improvement. Many genes curtail for agronomic traits have been identified in the past decades. In this review, we summarized the research progresses on the Major Crop Genomics, especially those achieved in 2017.

Key words: genome; agronomic traits; crops

1 基因组学领域论文分析

在 SCIE 数据库共获得 954 篇基因组学研究相关的文献。中国发文量最多,共 339 篇,约占该领域发文总量的一半;美国发文数量为 163 篇,排名第 2;德国发文量为 49 篇,排名第 3;印度、韩国、日本、法国和澳大利亚紧随其后,发文量分别为 48 篇、35 篇、33 篇、29 篇和 28 篇(表 1)。

中国机构表现较突出,共有 6 个机构进入该领域论文数量 TOP 10 行列。中国农业科学院发文量为 31 篇,排名第 1;其次是华中农业大学和中国科学院,论文数量各为 22 篇;南京农业大学发文 19 篇,排名第 3(图 1)。

表 1 基因组学领域主要发文国家

Table 1 Publications in field of Genomic research

排序 Rank	国家 Country	发文量 No. of publications
1	中国 China	339
2	美国 USA	163
3	德国 Germany	49
4	印度 India	48
5	韩国 South Korea	35
6	日本 Japan	33
7	法国 France	29
8	澳大利亚 Australia	28
9	英国 UK	19
10	意大利 Italy	17

收稿日期:2018-03-06 修回日期:2018-04-09 网络出版日期:2018-04-17

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20180417.0949.004.html>

基金项目:中国农村技术开发中心“农作物育种行业动态专题研究”项目

第一作者主要从事水稻遗传育种研究, E-mail: tangding@genetics.ac.cn

通信作者:程祝宽,主要从事植物分子细胞遗传研究, E-mail: zkcheng@genetics.ac.cn

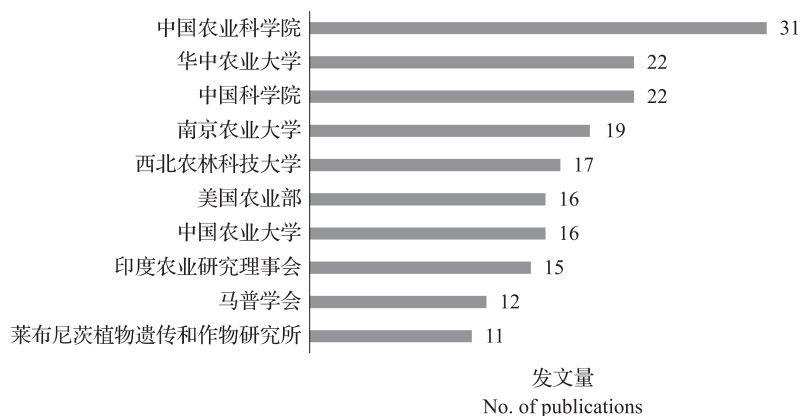


图1 作物基因组学领域主要发文机构

Fig. 1 Institutions who published papers in the field of genomic research

2 基因组学研究态势分析

农作物基因组研究的空前发展正推动着农业的第二次“绿色革命”。全基因组的剖析,可以提供每个农业生物物种或品种全基因组的遗传信息。尤其是对控制重要农艺性状的基因组成情况分析,以及对调控复杂性状的分子网络的解析,皆得益于高效与廉价测序技术的发展。目前已经实现了对重要农作物,如水稻、小麦、玉米、大豆、油菜、棉花、蔬菜等基因组测序或重测序,实现了对控制重要农艺性状关联基因的大规模克隆和鉴定。2017 年度经过全球科学家的努力,在主要农作物基因组研究方面取得了一系列重要进展。

2.1 水稻基因组研究

水稻(*Oryza sativa* L.)既是重要的粮食作物,也是生物学的模式植物。近年来,我国水稻功能基因组学研究一直走在世界前列,高质量的水稻参考基因组序列可谓功不可没。中科院遗传发育所梁承志课题组与四川农业大学李仕贵合作,利用 PacBio 单分子测序技术结合 fosmid 文库测序以及遗传图谱的相关结果,对籼稻品种蜀恢 498 进行测序,并通过 BioNano 光学图谱进行进一步的验证,最终获得了一个长度为 390.3 Mb,共由 17 个连续 DNA 片段(Super-Contig)组成的水稻基因组。该研究组装了蜀恢 498 除了 5 个着丝粒区域和少数几个串联重复序列区域以外的整个基因组,与已公布的水稻及拟南芥等植物基因组相比,蜀恢 498 的基因组具有更高的完整性、更好的连续性以及更低的错误率,是目前所有高等动植物中组装质量最高的基因组。通过对蜀恢 498 与日本晴这两个基因组的基因序列

进行比较发现,这两个品种超过 2/3 的基因在序列上存在差异。同时,比较结果显示两个基因组之间也存在大量的因转座子插入引起的染色体结构变异。此外,研究人员还成功组装出蜀恢 498 完整的线粒体序列,指出了日本晴线粒体序列中存在的错误,并进一步发现了目前日本晴基因组参考序列中掺杂了许多的线粒体和叶绿体序列,相关结果对于日本晴核基因组和线粒体基因组的数据提供了补充和修正。

蜀恢 498 基因组测序工作的完成,对于水稻尤其是籼稻优异等位基因资源的挖掘和利用、籼稻群体的全基因组关联分析的相关研究具有很重要的应用价值,同时对于提高目前高等动植物基因组的组装质量具有重要的指导意义^[1]。

杂草稻(*Oryza sativa* L.)是一类具有栽培稻外表、杂草特性的水稻,因其具有较强的适应性、生育期短、易落粒、种子休眠时间长等特点,在田间能够与栽培稻竞争光、水分和养分,对水稻的产量与品质危害极大。美国的杂草稻从形态学上主要分为 SH(straw hull)、BHA(black hull awned)两大类,以华盛顿大学 Olsen 教授领衔的中美科研人员,通过对美国本土的 18 个 SH 和 20 个 BHA 杂草稻进行全基因组测序,并与 145 种已公布的水稻基因组序列(包括 89 种栽培稻、53 种野生稻和 3 种中国中部杂草稻)进行比对分析,结果表明 SH 和 BHA 类杂草稻分别与东南亚籼稻以及 aus 稻亲缘关系较近,而中国杂草稻(来源于江苏省)则属于中国籼稻类型。通过比较中美杂草稻基因组中栽培稻和野生稻特有 SNPs 的分布比例,研究人员发现中国杂草稻包含大部分栽培稻的特异 SNPs,而美国杂草稻基因组中野

生稻特异 SNPs 比例要高于栽培稻 SNPs, BHA 类杂草稻 SNPs 的分布比例竟与野生稻高度相似, 表明中国杂草稻与栽培稻关系更近, 而美国杂草稻则不然, BHA 类杂草稻与野生稻的亲缘关系更近。从进化时间上看, 3 类杂草稻均要晚于水稻驯化以及品种内部分化时期。

进一步通过对 3 类杂草稻中与水稻驯化以及改良相关的基因进行分析, 在 *PROG1* (直立株型基因)、*sh4* (落粒基因)、*OsLGI* (穗型基因) 这 3 个代表驯化早期的基因中, 3 类杂草稻的变异位点与栽培稻一致, 表明了杂草稻起源于驯化后的栽培稻祖先, 支持了去驯化 (de-domestication) 是杂草稻起源的这一观点。而在大部分改良型基因位点, SH 型杂草稻变异和中国杂草稻以及栽培稻一致, 而 BHA 类杂草稻与野生稻更为一致, 表明杂草稻的进化在水稻驯化的早期和晚期都有可能发生。研究人员进一步分析了形成杂草稻特性的遗传学基础, 发现杂草稻中维持基本生物过程的基因并没有显著增加, 与组织发育、与抗逆相关的基因可能在不断富集, 这与杂草稻有着较快的生长速度以及较强的抗逆性可能不无关系。同时, 与杂草稻适应性的相关基因通常成簇存在, 形成少数的基因组岛 (genomic islands), 暗示了杂草稻进化过程的选择仅发生在基因组相对较小区域的位点上。这些基因组岛与目前已定位的与杂草稻驯化相关的 QTL 位置吻合^[2]。

中国是一个水稻生产大国, 杂草稻对我国粮食安全的危害近年来日趋严重。无独有偶, 浙江大学樊龙江等课题组也对国内的杂草稻起源与进化进行了相关的报道。研究人员通过收集我国境内杂草稻危害最为严重地区: 江苏、广东、辽宁和宁夏 4 地的 155 份杂草稻和 76 份当地栽培稻品种进行了全基因组重测序和群体分析, 发现 4 地的杂草稻群体各自相互独立起源于栽培稻, 其中江苏、广州杂草稻起源于籼稻, 而辽宁、宁夏的杂草稻则起源于粳稻, 且在起源过程中遭受了强烈的短期遗传瓶颈效应。进一步分析发现虽然 4 个杂草稻群体起源多样, 但是在 7 号染色体 6.0~6.4 Mb 区间内存在趋同进化区域, 包含决定种皮颜色基因 *Rc* 和一系列编码水稻过敏性蛋白在内的 15 个关键基因。这些基因可能在杂草稻的起源和进化过程中扮演了重要的角色。通过对等位基因的频率变化分析发现, 已有变异 (standing variation) 对于杂草稻的环境快速适应性起到了关键作用, 而新突变 (new mutation) 对于杂草稻尤其是籼型杂草稻的进化作用明显。同时研究还

表明杂草稻基因组上发现很多区域受到了平衡选择信号, 这有助于杂草稻产生更多的遗传多态性以适应复杂的生存环境^[3]。

以上两篇报道揭示了美国和中国杂草稻的起源、进化以及环境适应性的机制, 加深了人们对农作物驯化和去驯化遗传机制认识, 对于世界范围内的杂草稻的控制与防治乃至全球粮食安全都有着非常积极的意义。

2.2 小麦基因组研究

小麦是全球最重要的粮食作物之一, 小麦的稳产和增产对我国乃至全世界粮食安全的影响举足轻重。近年来由于全球气候变化、环境变化的影响, 小麦生产面临严峻的挑战, 对于小麦的育种和品种改良工作提出了新的要求。普通小麦 (*Triticum aestivum* L.) 是 3 个不同亚基因组形成的异源六倍体物种 (AABBDD), 由早期的野生二粒小麦与粗山羊草天然杂交而来。其基因组非常庞大且结构异常复杂, 富含大量的重复序列, 这些特点使得小麦基因组学的研究远远落后于水稻和玉米等二倍体植物, 严重制约了小麦功能基因组学研究和育种工作的深入。

野生二粒小麦 (*T. turgidum* subsp. *dicoccoides* (Körn.) Thell) 是栽培小麦的四倍体祖先种。以色列等多国研究人员对野生二粒小麦种质 Zavitan 进行了全基因组鸟枪法测序, 结合 3D Hi-C 数据和遗传图谱信息将 10.5 Gb 基因组序列组装到 14 条染色体中, 其中 10.1 Gb 序列能锚定到 14 条染色体上。同时, 研究人员对野生二粒小麦不同发育阶段的各个组织进行了 RNA 测序。建立了 65012 个高可信度的基因模型并分析了相关基因的分布。同源性分析进一步表明, 在亚基因组间的, 有 72.3% 的基因都有其同源基因。全基因组中, 有 82.2% 的序列注释为转座子序列, 而且这个比例在不同亚基因组间大致相当, 主要由长末端重复逆转录转座子组成。研究人员还对一个大麦关键驯化基因 *TtBtr1* 进行研究, 在野生小麦驯化的过程中, 麦穗变得不易破碎, 从而表现为不易落粒。通过构建定位群体, 研究人员发现了调控麦穗脆性表型的基因组区域, 并最终推测驯化小麦含有的 *TtBtr1-A* 和 *TtBtr1-B* 等位基因变异可能引起蛋白质功能丧失, 导致麦穗不易破碎。同时作者还检测了二粒小麦中可能受到选择的驯化区域, 发现与野生二粒小麦相比, 栽培二粒小麦间遗传多样性仅略微降低^[4]。

粗山羊草 (*Aegilops tauschii* Coss.) 是六倍体小

麦 D 基因组的二倍体祖先,是小麦重要的遗传资源。美国加州大学等研究机构基于 BAC 序列结合全基因组重测序序列、PacBio 技术以及遗传图谱并利用 BioNano 单分子光学图谱技术进行验证对粗山羊草基因组序列进行测序和组装,最终将约 4 Gb 占基因组 95.2% 的序列组装到 7 条染色体上。这些染色体是由 12 条祖先染色体通过嵌套的染色体插入(NCI,nested chromosome insertion)引起非整倍体减少进化而来。粗山羊草的 1D、2D、4D 和 7D 染色体主要是通过 NCI 这种形式形成的,5D 染色体与可能起源于水稻 9 号染色体以及 12 号染色体对应的祖先染色体短臂融合,随后 5DL 的 Os9 部分与 4DS 的 Os3 部分相互易位形成。

粗山羊草基因组中转座子序列约占 84.4%,其中主要是占 65.9% 长末端重复反转录转座子(LTR-RTs),Gypsy 和 CACTA 分别是最丰富的 RNA 和 DNA 转座子超家族。而新发现的 1113 个 TE 家族中大多数的拷贝数较少。通过对基因组序列的分析,研究人员共注释了 83117 个基因,包括 39622 个高可信度的基因以及 43495 个低可信度的基因,其中 38775 个高可信度基因可锚定到染色体上,但这些基因中只有 5050 个是单拷贝基因。通过与短柄草、水稻、大麦、高粱以及拟南芥的基因组中注释的基因相比,粗山羊草的基因具有最长的平均外显子和转录本,而外显子平均数目却少于上述物种。同时研究人员还对小麦特异基因和抗病基因进行了分析,并研究了基因在染色体上的分布规律以及与重组率之间的关系,这些研究成果为小麦新基因的发掘和应用、小麦的品种改良、小麦的进化与多倍体研究以及比较基因组学研究都有着重要的意义^[5]。

对于粗山羊草的测序工作,中国农科院贾继增研究团队也在同时进行,他们利用最新的技术,对小麦 D 基因组的供体粗山羊草进行重新测序与组装,成功将 92.5% 基因组序列锚定到染色体上,完成了 D 基因组序列的测定。研究人员通过对粗山羊草基因组的序列进行分析,绘制了包括基因位点分布、基因表达、假基因分布、甲基化分布、重组率、microRNAs 分布以及转座因子(TE)分布信息的基因组特征图,重点分析了 TE 对基因组进化、基因结构以及基因表达的影响。研究发现,粗山羊草基因组中有近 1/2 的基因中携带有 TE,而 TE 通常还会抑制基因的正常表达。研究结果还表明在粗山羊草基因组中近期发生了一波基因复制事件。该研究还首次将小麦分子标记和之前检测到的重要农艺性状基因和

QTL 整合到小麦 D 基因组上,获得一个完整的高精度整合图谱。将极大促进小麦基因克隆和分子育种工作^[6]。

2017 年,小麦结构基因组学取得了长足的进步和发展,相关研究成果为小麦的功能基因组学和蛋白组学的研究奠定了坚实的基础,将对小麦的分子生物学研究以及分子设计育种工作产生巨大的影响。

2.3 大麦基因组研究

大麦(*Hordeum vulgare* L.)是世界上驯化最早的饲料和粮食作物,也是遗传学研究的模式植物之一。作物全球第四大禾谷类作物,大麦在我国国民经济发展中具有很重要的地位。大麦结构基因组学研究是大麦功能基因组学和遗传育种研究取得突破性成果的重要前提。但是大麦基因组非常庞大,结构复杂并富含转座因子,因此全基因组测序工作难度很大。李承道和张国平等,综合运用包括 BAC 测序、Hi-C 及 BioNano 等多种最先进的测序和组装技术,组装完成了一个目前最为完整的包含 4.79 Gb 的大麦 Morex 高质量参考基因组序列,其中 94.8% 的组装序列被定位到大麦的各染色体上。通过分析鉴定了 39734 个高置信度基因,预测了 19908 个长链非编码 RNA 和 792 个 microRNA 前体基因位点,并利用 Hi-C 技术对大麦染色体特征及行为规律进行探索和验证。大麦基因组主要由高拷贝重复序列组成,约 80.8% 的序列为转座因子,作者同时解析了重复元件以及基因在染色体上分布的特点,重点分析大麦麦芽品质相关基因家族的特点,明确了相关基因的变异类型,通过对现代核心种质资源的 SNPs 进行剖析,指出了大麦基因组中易受遗传侵蚀的区域,为拓宽栽培大麦日趋狭窄的基因库提供了相应的策略。高质量的大麦基因组参考序列的发表,为大麦优质基因资源的克隆和利用提供了强有力的工具,也为高品质大麦育种指明了方向^[7]。

大麦是有稃大麦(皮大麦)和裸大麦的统称,我们习惯所说的大麦是指皮大麦,而裸大麦在各地称谓不同,在青藏高原则被称作青稞,青稞具有广泛适应性和强抗逆性等特点,是青藏高原地区主要的粮食作物。浙江大学张国平课题组利用第三代测序技术 PacBio 对藏区种植面积最大的青稞品种藏青 320 进行测序分析,从头组装出 4.84 Gb 的基因组,其中 4.59 Gb 可锚定到大麦 7 条染色体上。分析表明该基因组共有 46787 个具有高可信度的基因,其中 31564 个基因信息在 39 个野生和栽培大麦 RNA 测

序数据库以及 NCBI 非冗余蛋白数据库中获得了验证。此外,研究人员以青稞作为参考基因组,比较皮大麦(Morex)和裸大麦(藏青 320)基因组间的差异,显示二者亲缘关系较近。藏青 320 的参考序列的公布填补了大麦 Morex 基因组序列的部分缺口。相关结果对于大麦的遗传改良具有重要意义^[8]。

2.4 玉米基因组研究

玉米(*Zea mays* L.)是世界上分布最广的农作物之一,自 2009 年玉米自交系 B73 的测序完成以来,参考基因组版本不断更新,但是由于玉米基因组结构非常复杂,存在高密度的重复序列以及高活性的转座子,这为玉米功能基因组的测序和组装提出了新的挑战和提升空间。2017 年,美国研究人员使用先进的单分子实时测序和高分辨率光学制图技术对玉米近交系 B73 进行测序,从头组装出 2106 Mb 大小的参考基因组。与之前的测序组装技术相比,新技术增加了重叠群长度,改进了基因间隔与着丝粒区域的组装质量,更新的基因注释。玉米基因组主要由转座元件构成,本研究共鉴定出 130604 个结构完整的逆转录转座子(1268 Mb),有助于我们理解玉米转座子的发生进化史。通过比较最新的 B73 基因组图谱与另外两个自交系 Ki11 系和 W22 系基因组图谱,发现后两个品系的基因组与 B73 的基因组序列差异巨大,分别只有 32% 和 39% 的序列能匹配到 B73 基因组序列上。表明玉米不同自交系之间基因组变异很大,具有良好的可塑性,为其不断适应新的环境提供遗传基础和可能^[9]。

现代栽培玉米是大约 9000 多年前由生长在低海拔地区的野生大刍草驯化而来。众所周知,驯化过程对作物来说,意味着优异的基因资源丢失以及遗传多样性丧失。利用野生资源进行作物的遗传改良越来越受到育种学家的重视。而大刍草(*Zea mays ssp. mexicana* Iltis)可以与现代栽培玉米进行自由杂交。从而实现优异基因渗透,为现代栽培玉米育种和改良提供了一条新途径。华中农大严建兵等课题组利用玉米 Mo17 与大刍草类蜀黍的回交重组自交系群体,结合二代和三代测序技术,组装出了高质量的玉米 Mo17 和 *mexicana* 基因组,基因组大小分别为 2.04 Gb 和 1.20 Gb。预测的高可行度基因为 40003 个和 31387 个。通过比较发现同属玉米属的 Mo17、B73 以及 *mexicana* 三者基因组存在较大的结构变异,研究人员进一步调查了 895 个近交系,发现 10.7% 的玉米基因组与 *mexicana* 基因组存在基因渗透,暗示了大刍草可能为玉米的环境适应性

提供了基因资源。本研究可深入挖掘来自野生玉米材料中的优异基因,为玉米的遗传改良提供了宝贵的基因资源^[10]。

2.5 棉花基因组研究

棉花是重要的天然纤维和油料作物,也是研究多倍体进化和作物驯化的重要模式植物。南京农业大学张天真等课题组通过将栽培棉与野生棉对比,绘制出了棉花表观遗传基因的“甲基化基因图谱”,对野生棉和栽培棉之间超过 1200 万个的差异甲基化胞嘧啶进行分析,鉴定出 519 个表观等位基因(epialleles),这些基因可能在异源四倍体棉花的进化和驯化过程中发挥作用。同时本研究还重点分析了光周期敏感基因 *COL2* 在野生棉和栽培棉中因甲基化水平的差异,而导致不同的光周期敏感性和适合种植区域,表明了 DNA 甲基化在棉花驯化过程中的重要作用以及育种上的用途。棉花甲基化基因图谱展示了棉花基因组在进化过程中的 DNA 甲基化变化的一系列特点,为帮助研究人员选育出高产、优质、多抗的棉花新品种提供了重要参考线索^[11]。通过对 318 份棉花地方品种和现代改良品种(系)的全基因组重测序,分析了基因组的遗传变异和种群结构,揭示了现代改良棉花品种高产优质的遗传基础和演化规律。通过对中国地区种植的 258 份棉花种质进行全基因组关联分析,鉴定了 119 个与产量、纤维品质、黄萎病抗性等关联位点。并发现了两个参与乙烯途径的基因与棉花增产相关。本研究为棉花精准育种和改良提供了基因组学基础,具有重要的理论和应用价值^[12]。

目前,生产上主要棉花栽培种为异源四倍体陆地棉(*Gossypium hirsutum* L.),陆地棉具有很长的驯化和栽培历史,长期的人工驯化选择了一些优异的变异,改变了陆地棉的主要农艺性状,但却不可避免得造成陆地棉遗传资源的流失。中国的陆地棉遗传多样性较为单一,遗传资源狭窄,使得棉花遗传育种工作长期以来进展缓慢。张献龙等从世界各地收集了 31 份棉花野生种和 321 份栽培种进行全基因组重测序。对包括单碱基多态性(SNP),插入/缺失(InDel)和结构变异(SV)进行检测构建了陆地棉的基因组变异图谱。通过将野生种与驯化种进行比较,分析控制陆地棉的纤维产量和品质这些性状改变的遗传学基础,在全基因组范围内鉴定了 93 个驯化选择区间囊括了 1777 个基因,涉及植株形态、产量、纤维品质以及黄萎病抗性等诸多农艺性状。研究人员还对 267 份棉花材料进行全基因组关联分

析,鉴定出 19 个与纤维质量相关的显著位点,其中 16 个位点以前未见报道,可进一步用于对棉花纤维品质的遗传改良,同时该研究还对非编码区的调控变异进行分析,鉴定了大量启动子上的顺式调控元件和增强子元件,此外还提供了棉花的驯化过程 A 和 D 两个亚基因组存在不对称选择的证据,以上结果为棉花优异等位基因的挖掘和农艺性状的改良提供了重要参考^[13]。

2.6 大豆重要性状遗传网络解析

大豆(*Glycine max*(L.) Merr.)是人类蛋白质和油类的主要来源之一,是一种非常重要的粮油饲料作物。目前我国大豆主要依赖进口,大豆生产面临非常严峻的形势,这为大豆育种提出了巨大的挑战。中国科学院遗传发育所田志喜等课题组通过对 800 多份大豆材料进行全基因组关联分析,深入解析了大豆 84 个农艺性状间的遗传调控网络,确定了 245 个重要的遗传基因位点,明确了 14 个与脂肪酸积累相关的基因,为大豆的分子设计育种提供重要的理论基础。利用连锁不平衡分析,发现大豆的 51 个不同农艺性状可通过 115 个关联位点相互联系起来,形成复杂的、多性状、多位点的遗传调控网络,其中 23 个关联位点起到了关键调控作用,并对其中部分位点在不同性状耦合中的作用进行了验证。大豆的产量性状是一个多基因控制的复杂性状,该研究通过对大豆产量性状进行解析,明确其遗传调控网络 and 关键调控单元,对大豆的分子设计育种具有重要的指导意义^[14]。

2.7 珍珠粟基因组研究

珍珠粟(*Cenchrus americanus*(L.) Morrone)是一种异花授粉的二倍体植物,主要分布在非洲、印度及南亚等地,具有耐干旱、耐贫瘠、耐酸等优良特性。作为一种 C_4 作物,珍珠粟具有非常高的光合效率和生物学产量,而且营养丰富,在半干旱地区的粮食种植产业中有着重要的地位。研究人员采用二代测序技术,对珍珠粟全基因组进行了测序、成功组装了 7 条染色体。根据估计,珍珠粟的基因组大小为 1.76 Gb,其中 80% 以上的基因组序列为重复序列,存在超过 38000 个基因,并注释了 27000 多个基因。科研人员在珍珠粟基因组中验证了 378 个与植物抗性相关的 NBS 基因,其中有 26.2% 和 25.7% 分别位于 4 号和 1 号染色体,与报道的霜霉病抗性位点吻合。同时还发现与耐热和耐旱相关的基因家族在基因组中富集。通过对珍珠粟的驯化和进化史进行分析,发现栽培群体可能起源于西非中部的野生群

体。并鉴定了栽培品种在驯化过程的受选择区域以及相关基因,通过对珍珠粟近交种质群体的 288 个测交后代进行了 20 个性状的 GWAS 分析,发现每穗实粒数等 15 个产量性状与 1000 多个标记存在关联,这些标记将可能用于珍珠粟的遗传育种。珍珠粟基因组相关研究成果的公布将会有助于人们了解珍珠粟的遗传资源,并最终加速珍珠粟的品种培育^[15]。

2.8 藜麦基因组研究

藜麦(*Chenopodium quinoa* Willd.)为一年生四倍体草本植物,具有耐干旱、盐碱、病虫害等特性,同时藜麦富含极高的营养价值是开发潜力巨大的一种作物资源。但是目前对于藜麦的基因组研究却相对缺乏,以沙特为首的国际团队,综合利用 PacBio 测序技术、结合 Bionano 光学图谱、Hi-C 技术以及遗传图谱组装得到了 1.39 Gb 大小的高质量的藜麦参考基因组序列(基因组大小估值 1.45 ~ 1.50 Gb),获得注释的基因大约有 44776 个,该基因组 64% 由重复序列构成,包含大量的长末端转座因子。为进一步了解藜麦的基因组结构和进化史,作者对藜麦 A 基因组二倍体 *C. pallidicaule* Aellen 和 B 基因组二倍体 *C. suecicum* J. Murr 进行了测序,并对亚基因组的结构和组成进行了分析,初步估计藜麦的四倍化大约发生在 3.3 百万 ~ 6.3 百万年前。并通过对 15 个来源于高原和沿海的藜麦样本以及祖先种进行了重测序,进化树分析表明藜麦是在一次独立事件中从 *C. hircinum* Schrad. 驯化而来,而高原藜麦和沿海藜麦可能在各自的环境中独立驯化而来。最后 D. E. Jarvis 等^[16]对藜麦皂素产生的相关基因进行分析,为藜麦的遗传改良奠定了基础。

2.9 橡胶草基因组研究

橡胶草(*Taraxacum kok-saghyz* L. E. Rodin)是多年生的二倍体草本植物,根部可产生高质量的天然橡胶和菊糖。具有地理适应范围广、生长周期短、组织培养及基因编辑容易等特点,被认为是一种理想的产胶备选经济作物和科学研究的模式植物。中科院遗传发育所李家洋课题组利用 PacBio 单分子测序技术独立组装完成了大小为 1.29 Gb 橡胶草基因组草图,分析表明橡胶草基因组包含 46731 个预测基因和多达 68.56% 的重复序列。鉴于橡胶草的自交不亲和性,本研究通过对基因组杂合区进行检测,发现了橡胶草基因组中与自交衰退相关的可能候选区域。此外通过比较产胶植物与非产胶植物之间的基因组,鉴定了橡胶草中橡胶合成途径和菊糖

合成途径的相关基因及其表达情况,并阐述了橡胶合成过程中两个关键基因家族 CPT/CPTL 和 REF/SRPP 的进化进程。该研究成果标志着对橡胶草分子生物学研究进入了后基因组时代,将推动我国橡胶产业的发展^[17]。

参考文献

- [1] Du H L, Yu Y, Ma Y F, et al. Sequencing and *de novo* assembly of a near complete *indica* rice genome[J]. *Nature Communications*, 2017, 8:15324
- [2] Li L F, Li Y L, Jia Y L, et al. Signatures of adaptation in the weedy rice genome[J]. *Nature Genetics*, 2017, 49:811-814
- [3] Qiu J, Zhou Y J, Mao L F, et al. Genomic variation associated with local adaptation of weedy rice during de-domestication[J]. *Nature Communications*, 2017, 8:15323
- [4] Avni R, Nave M, Barad O, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication[J]. *Science*, 2017, 357:93-97
- [5] Luo M C, Gu Y Q, Puiu D, et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*[J]. *Nature*, 2017, 551:498-502
- [6] Zhao G Y, Zou C, Li K, et al. The *Aegilops tauschii* genome reveals multiple impacts of transposons[J]. *Nature Plants*, 2017, 3:946-955
- [7] Mascher M, Gundlach H, Himmelbach A, et al. A chromosome conformation capture ordered sequence of the barley genome[J]. *Nature*, 2017, 544:427-433
- [8] Dai F, Wang X L, Zhang X Q, et al. Assembly and analysis of a *qingke* reference genome demonstrate its close genetic relation to modern cultivated barley[J]. *Plant Biotechnology Journal*, 2017, <https://doi.org/10.1111/pbi.12826>
- [9] Jiao Y P, Peluso P, Shi J H, et al. Improved maize reference genome with single-molecule technologies[J]. *Nature*, 2017, 546:524-527
- [10] Yang N, Xu X W, Wang R R, et al. Contributions of *Zea mays* subspecies *mexicana* haplotypes to modern maize[J]. *Nature Communications*, 2017, 8:1874
- [11] Song Q X, Zhang T Z, Stelly D M, et al. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons[J]. *Genome Biology*, 2017, 18:99
- [12] Fang L, Wang Q, Hu Y, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits[J]. *Nature Genetics*, 2017, 49:1089-1098
- [13] Wang M J, Tu L L, Lin M, et al. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication[J]. *Nature Genetics*, 2017, 49:579-587
- [14] Fang C, Ma Y M, Wu S W, et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean[J]. *Genome Biology*, 2017, 18:161
- [15] Varshney R K, Shi C C, Thudi M, et al. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments[J]. *Nature Biotechnology*, 2017, 35:969-976
- [16] Jarvis D E, Ho Y S, Lightfoot D J, et al. The genome of *Chenopodium quinoa* [J]. *Nature*, 2017, 542:307-312
- [17] Lin T, Xu X, Ruan J, et al. Genome analysis of *Taraxacum kok-saghyz* Rodin provides new insights into rubber biosynthesis[J]. *National Science Review*, 2017, <https://doi.org/10.1093/nrs/nwx101>