

能源植物高粱基因组研究进展

罗洪¹, 张丽敏¹, 夏艳¹, 吴小园^{1,2}, 王聪^{1,2}, 刘智全¹, 景海春¹

1. 中国科学院植物研究所北方资源植物重点实验室, 北京 100093

2. 中国科学院大学, 北京 100049

摘要 回顾了高粱基因组学研究的发展进程, 概述了初期组学数据的积累、参考基因组的破译及新一代测序技术和数据分析方法引领下的组学研究进展; 介绍了高粱基因组的结构, 从比较基因组学的角度, 分析了高粱基因组的进化及其特性; 探讨了高粱功能基因组的研究方法和研究进展, 总结了已经发掘的高粱关键基因和遗传位点, 对高粱组学数据资源进行了归纳。对高粱基因组学的发展方向进行了展望。

关键词 高粱基因组; 新一代测序技术; 比较与功能基因组学; 重要 QTL 和基因; 生物质能源性状

中图分类号 Q-1

文献标志码 A

doi 10.3981/j.issn.1000-7857.2015.16.002

An update on genome research of biofuel sorghum (*Sorghum bicolor*)

LUO Hong¹, ZHANG Limin¹, XIA Yan¹, WU Xiaoyuan^{1,2}, WANG Cong^{1,2}, LIU Zhiquan¹, JING Haichun¹

1. Key Laboratory of Plant Resources, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Research of sorghum genomes has been accelerated by the development of next-generation sequencing technologies and bioinformatics tools and has accumulated a large amount of genomics data. This review describes the structural and sequence characteristics of sorghum genome and discusses its evolutionary background from the perspective of comparative genomics. We also highlight new findings obtained through exploiting the genomic information and future directions of sorghum genome research to develop it into a dedicated biofuel crop.

Keywords sorghum genome; next-generation sequencing; comparative and functional genomics; important QTLs and genes; the biofuel syndrome

高粱(*Sorghum bicolor* L. Moench)是禾本科高粱属二倍体植物($2n=2x=20$), 是世界上继小麦(*Triticum aestivum* Linn.)、玉米(*Zea mays* L.)、水稻(*Oryza sativa* L.)和大麦(*Hordeum vulgare* L.)之后的第五大作物^[1], 为干旱及半干旱地区 30 多个国家的近 5 亿人提供粮食 (FAO 统计数据)。高粱最初的起源地和种植区被广泛认为是苏丹和尼日利亚之间的南撒哈拉 (Sub-Saharan Africa)^[2], 并于公元前 3000 年以前从这个地区扩散进入非洲西部热带地区, 干旱的东北地区及非洲东南地区, 此后高粱又经地中海传入印度、中国等地^[3]。但是就中国“kaoliang”的起源问题, 科学家还存在不同的见解。在世界主要作物中, 高粱是唯一起源于非洲的作物, 在

作物起源进化和驯化研究中起着重要地位^[4]。

高粱是 C₄ 单子叶作物, 光合作用效率高, 生物产量高, 经济效益大, 被称为“高能作物”。根据用途, 高粱可分为籽实高粱, 甜高粱和饲草高粱等几种类型。除了作为重要的粮食作物外, 高粱也可以用于食品加工, 酿酒工业, 青贮饲料和色素工业等, 尤其是近几年全球能源危机的加剧, 甜高粱作为新一代能源作物备受关注。与普通籽实高粱相比, 甜高粱具有生长快、产量高、适应性广、耐逆性强和对土壤及肥料要求不高等优点, 素有“作物中的骆驼 (camel amongst crops)”之称^[5]。此外, 甜高粱茎秆含汁量较高 (60% 以上), 汁液锤度一般在 16%~22%, 每公顷茎秆可产 4500~6000 kg 50° 的白酒,

收稿日期: 2015-06-30; 修回日期: 2015-07-21

作者简介: 罗洪, 博士, 研究方向为生物信息, 电子信箱: hong.luo@ibcas.ac.cn; 景海春 (通信作者), 研究员, 研究方向为甜高粱基因组学及其分子育种设计, 电子信箱: hcjing@ibcas.ac.cn

引用格式: 罗洪, 张丽敏, 夏艳, 等. 能源植物高粱基因组研究进展[J]. 科技导报, 2015, 33(16): 17-26.

因此甜高粱被认为是大有希望的甘蔗糖厂的补充原料。用甜高粱秸秆生产燃料酒精作为汽车的原料可大大减少大气污染,是大有希望的生物新能源。同时,因其茎秆中含有高糖高蛋白,使其成为优良的青贮饲料,对奶牛的产奶量有明显的增产效果。甜高粱含有14%~18%的纤维素,其产量每公顷达7.5~15 t,是纤维素或造纸的好原料。甜高粱的这些显著特性使得其成为土地贫瘠、投入低下的农田生态系统中成功种植的重要经济和能源作物,对国力不足的发展中的非洲和亚洲国家的能源粮食安全与农村发展起着重要作用。

高粱在全世界范围均有种植,中国从南到北、自东至西均可种植甜高粱,其中可种植的未利用地面积达 5919.2×10^4 hm^2 ,主要集中于新疆和内蒙古等省区;而最适宜甜高粱种植的未利用地面积为 286.7×10^4 hm^2 ,主要分布在黑龙江、内蒙古、山东和吉林等地。中国未利用地甜高粱乙醇生产潜力较大,在不考虑其他社会经济限制因素下,总乙醇生产潜力可达 11838.5×10^4 t以上。最适宜未利用地的甜高粱乙醇生产潜力为 573.4×10^4 ~ 2637.8×10^4 t,平均为 1075.2×10^4 t,可满足中国目前E20乙醇汽油84.8%的需求^[6]。因此,甜高粱作为重要的能源作物具有重要的研究价值。

1 高粱基因组学研究的进展

1.1 早期的组学研究探索

高粱组学研究早期,受实验条件和技术的制约,重点主要是收集和构建组学资源,如BAC文库^[7,8]、Cot基因组文库^[9,10]、生物和非生物胁迫下的EST数据库^[11]、分子标记及遗传图谱等。由于对高粱基因组序列知之甚少,初期的高粱遗传图谱大都以限制性片段长度多态性(restriction fragment length polymorphism, RFLP)探针为主要的标记类型。高粱的第一张遗传图谱由Hulbert等^[12]在1990年绘制完成,他们利用玉米基因组中开发的37个RFLP探针对高粱品种ShanguiRed×M91051的F₂群体进行基因型分析,构建了总长为283 cM的8个连锁群。随着研究的深入,高粱的遗传图谱的标记数目、连锁群和图谱的长度均得到大幅度的提高^[13,14]。

随着新技术的发展及人们对基因组结构认识的深入,大量新类型的分子标记发展起来,其中最为常用的是简单序列重复(simple sequence repeat, SSR)、扩增片段长度多态性(amplified fragment length polymorphism, AFLP)和随机扩增多态性DNA(random amplification polymorphic DNA, RAPD)标记。这些分子标记与RFLP相比,多态性更为丰富,数量众多,稳定性更佳,同时操作也比较简便。Tao等^[14]首次将SSR标记用于高粱的遗传图谱中,构建了总长为1400 cM,包含21个连锁群的遗传图谱。Boivin等^[15]将AFLP标记运用到了高粱的遗传图谱中,绘制了总长为1889 cM,共11个连锁群的遗传图谱。另外,研究人员整合多种分子标记用于构建高粱高密度遗传图谱。2002年,Menz等^[16]构建了包含RFLP、SSR和AFLP3种类型的标记的遗传图谱,此图谱包含2926个标记位点,包含了10条连锁群,总长达到1713 cM。Bowers等^[17]

利用2512个标记位点构建的全长为1059 cM的遗传图谱,平均标记间距仅为0.4 cM。随着芯片技术的发展,一种依赖基因芯片杂交技术来区别基因组座位多态性的标记——多样性序列芯片技术(diversity arrays technology, DAiT),被Mace等运用到高粱遗传图谱中^[18]。

1.2 高粱参考基因组的破解

高粱作为重要的农作物,其基因组的研究早已展开。但限于早期实验条件和技术,高粱全基因组一直未能破解。近10年来,随着基因组序列测定成本的降低,以及基因组生物信息学分析技术的进步,这一情况得到了很大的改善。

2009年初,美国能源部联合基因组研究所(DOE JGI)主持完成了对高粱品种BTx623基因组的测序,组装,以及初步的分析。高粱基因组采用Sanger法测序,建立了插入片段从2~3 kb到108 kb的各种双端读段序列文库,共测得8.5倍覆盖率的读段,结合物理图谱进行了De novo拼接。又通过二代测序方法,补测了大量基因组以及转录组读段,不断对组装和注释进行改善。到目前为止,其基因组序列和注释已经更新到2.1版本。

BTx623作为高粱参考基因组,其最新的序列和注释信息可在JGI建立的植物基因组网站phytozome免费获取。高粱参考基因组为2倍体,有10条染色单体,大小约726 Mb,含约33000个基因,高粱基因组基本功能元件的平均长度见表1。

表1 高粱基因组功能元件长度
Table 1 Characteristics of sorghum genome

名称	平均长度
Gene	2856 bp
Transcript	1426 bp
Exon	267 bp
Intron	419 bp
Protein	409 aa

注:数据源于<http://genome.jgi-psf.org/Sorbi1/Sorbi1.info.html>。

1.3 二代测序技术引领的高粱组学研究新方法

第二代测序(next generation sequencing, NGS)技术的发展,是推动组学研究方法进步的最大动力。第二代测序技术起始于2005年底,当时454公司提出了基于焦磷酸边合成边测序(sequence by synthesis)原理^[19]为基础的测序方法,相对于第一代Sanger测序方法^[20],可以大大提高测序效率并降低测序成本。很快Illumina等公司迅速优化了测序策略,包括采用效率更高的桥式扩增(Bridge PCR)^[21]等方法,并推出了商业化的测序平台。由此,大规模的基因组序列测定研究开始在世界各地的科研机构广泛展开。迄今为止,二代测序技术和平台经过近10年的发展和融合,在测序质量控制和工业化流程方面,已经比较成熟。

第二代测序技术通常包括DNA样品的制备、样品的剪

切、样品的扩增,以及通过荧光信号识别碱基的边合成边测序^[22-24]。至此二代测序的实验工作部分完成,但对得到的数据进行生物信息学分析部分的工作刚刚开始。由于在样品制备过程中,DNA样品被随机打断并扩增,二代测序最终得到的是海量长度一般为100~200 bp的DNA短片段,通常称为“读段(reads)”。由于是随机打断,理论上这些读段在基因组上随机分布,并且互相之间可能有重叠。这些测序读段的总长度之和与基因组总长度的比例,称为“测序深度(sequencing depth)”,如何利用这些海量的短片段信息进行组学研究,是摆在人们面前的一个更关键的课题。

参考基因组序列的破译,进一步拓展了高通量组学研究的应用范畴。很快发现,利用重测序(resequencing)的方法,加之生物信息学分析手段,可以高效准确地识别包括SNP(single nucleotide polymorphism)和INDEL(insertions and deletions)在内的基因组变异^[25]。重测序是指采用比较低的深度(一般只需*De novo*拼接所需深度的几十分之一),对某物种的不同品种进行测序。然后将得到的读段比对到参考基因组上,通过对比对结果的分析,找到目标品种基因组与参考基因组之间的差异。这种方法的优点是可以一次实验测序很多品种,而且由于不再需要进行耗时的*De novo*拼接,数据分析成本降低,极大地提高了研究效率。重测序方法的高效性在植物基因组的应用中也得到了验证^[26,27]。通过这种利用大规模重测序进行基因组变异挖掘和基因型分析的方法(genotyping by sequencing,GBS)^[28,29],大量的基因组变异信息被挖掘。这些信息可以作为高密度的分子标记,用来进行全基因组关联(genome wide association study,GWAS)和群体遗传育种的研究。

转录本测序(RNA-seq)是另一种较常用的研究方法^[30],其准确性和灵活性已经得到广泛的认可,在转录水平研究的领域,正逐渐取代传统的基因表达芯片。RNA-seq的基本原理是先提取待研究样品的RNA,以其为模板合成cDNA并将其打断。接下来的文库制备和测序流程,与一般基因组测序基本类似。不同的是,得到的读段都来自基因的转录本,因此将这些读段比对到基因组上的方法不同于重测序^[31]。通过转录本测序可以进行一系列研究,包括基因边界鉴定和验证,可变剪切挖掘,基因融合探查,编码区变异识别,基因表达水平差异比较,以及全新转录本发现等。另外,对非编码区转录本功能,比如Non-coding RNA及microRNA的研究也是近年来的热点。

2 高粱基因组结构与比较基因组学

2.1 高粱进化与比较基因组学

作为禾本科作物,高粱和水稻一样具有成为模式植物的潜质。而作为热带C₄植物的代表,高粱与很多重要的C₄作物,如玉米、甘蔗(*Saccharum officinarum*)、芒草(*Miscanthus*)、柳枝稷(*Panicum virgatum*)等,有更近的亲缘关系,同属蜀黍族(Andropogoneae)。研究表明,禾本科植物的共同祖先在大

约7000万年前发生过一次基因组复制事件,之后C₃和C₄植物开始分化^[32]。而高粱和玉米在大约1200万年前从一个共同祖先分离,之后玉米的基因组经历了至少一次大规模的全基因组复制事件^[33],因此高粱和玉米共享很多同源基因。而甘蔗和高粱从它们的共同祖先分离可追溯到近至大约500万年前^[34],虽然之后甘蔗至少经历了2次全基因组复制事件^[35],但二者基因组的很多区域仍然具有高度的共线性^[36]。同时,高粱自分化后,未发生全基因组大规模复制,其基因组相对较小,复杂程度低。如玉米基因组大小约2.4 G,甘蔗染色体个数可达 $2n=106\sim 114$ ^[37]。高粱作为禾本科C₄模式植物,与经典C₃模式植物水稻一起,为禾本科比较基因组学研究提供更广阔的前景。

全基因组复制在植物基因组中产生了大量的重复片段,具有高度序列相似性,而这些片段的产生可能是由于基因组之间不规则的重组导致的。Wang等^[38]发现,在水稻高粱发生分化后,部分水稻基因和高粱基因受到非交互重组(nonreciprocal recombination)的影响,而这种重组在染色体末端具有更高的频率。这种基因转化在选择压力上没有变化,也并不会影响GC含量,但这种基因转化会加大水稻、高粱同源基因之间的差异,表明这样的不规则重组能加速进化而非维持保守。

比较甜高粱品系Keller与参考测序品系BTx623基因组的差异发现,串联和片段重复(tandem and segmental duplication)是导致它们基因表达分化的主要因素^[39]。对基因演化的研究发现,山羊草(*Aegilops tauschii*)、二穗短柄草(*Brachypodium distachyon*)、水稻以及高粱中的基因缺失和基因重复率相似,山羊草基因与高粱基因的同线性为64%^[40]。有研究者^[41]对玉米、水稻、高粱和二穗短柄草中的简单重复序列进行分析,发现所占比例分别为2.5%、44.1%、12.3%和28.1%,这暗示相比于高粱和玉米,水稻基因组经历了更少的进化事件。

2.2 高粱基因组转座子研究

高粱基因组有其独有的特点和复杂性。比如,染色体异染色质含量^[42]460 Mb,约占基因组的63%,远高于水稻的63 Mb,约占基因组的15%;基因组重复序列含量相当高(表2)。如表2所示,高粱基因组重复序列主要是各类转座子,总含量占基因组的62%,远高于水稻的39.5%,但低于玉米的82.1%。而玉米基因组转座子绝大部分属于长末端重复序列(long terminal repeats,LTR)类反转座子,其活跃程度相当高。类似于玉米,高粱基因组的LTR类反转座子比例也相当高。同时,高粱基因组含有一定比例的DNA转座子,比例介于水稻和玉米之间。另外,高粱基因组还包括有3%左右的串联重复序列^[42]。水稻、高粱和玉米中LTR的比例分别为23%、54%和75%,但其全长LTR反转录转座子在全基因组的比例分别为19.6%、36.4%和14.3%^[43]。说明在进化过程中,相对于玉米和水稻,高粱的反转座子序列重组和缺失的概率较小。

表2 高粱、水稻和玉米基因组重复元件
Table 2 Repeat elements of sorghum, maize and rice genome

重复元件类型	占基因组比例/%		
	水稻	高粱	玉米
Retroelement	25.78	54.52	79.44
LTR retrotransposon	23.47	54.43	75.08
non-LTR retrotransposon	1.24	0.04	0.35
DNA transposon	13.67	7.46	2.68
DNA transposon super family	7.04	4.79	0.92
MITE	5.24	1.74	0.32
Helitron	0.33	0.81	1.31
总和	39.50	62.00	82.10

注: 数据源于文献[42]。

Jiang等^[44]在高粱中发现了23915个全长LTR反转录转座子影响到672个基因,在水稻中有7043个全长LTR反转录转座子影响到1343个基因,在玉米中有31172个全长LTR却只影响到424个基因。在水稻中这些基因在GO富集分析中并没有显著的富集,而在高粱中则在DNA/RNA代谢以及核染色质上存在过表达。玉米全长LTR反转录转座子在每个染色体的着丝粒区都有较高的比例,但是高粱全长LTR反转录转座子着丝粒区并不富集。在水稻LTR基因中只有36%表达,而高粱中72.4%的有表达。大部分水稻LTR基因在中性选择下进化为假基因,而在高粱中则经历了正向选择形成有功能的基因。在高粱LTR基因中有812个被认为是siRNA,而在水稻中只有85个,这些siRNA发现与负调控相关。

Sharma等^[45]研究了高粱、玉米、水稻和短柄草等植物中着丝粒反转录转座子的进化,发现着丝粒区的转座子(centromericretrotransposons, CR)不仅在粟种之间具有水平转移(horizontal transfer, HT),并且随后迁移到谷子和高粱中,结合内源元件形成大量的重组体。其研究表明有水稻的CR元件迁移到高粱中,并结合高粱已有的CR元件进行了重组。而现存的LTR反转录转座子就是在转座子元件在水平转移以后,在新的宿主里保留的最佳反转录转座子。也有研究表明,在水平转移的转座子元件中,Route66在水稻和高粱之间迁移^[46]。高粱转座子的高度活性甚至可能导致水平基因转移(horizontal gene transfer, HGT),有证据表明,高粱的寄生植物独脚金(*Striga hermonthica*)就获得了高粱的基因^[47]。

2.3 高粱基因组可变剪切研究

Panihi等^[48]在高粱基因组中发现了2137个可变剪切事件,其中40%是属于内含子保留(比已研究的杨树(*Populus L.*)、水稻和衣藻(*Chlamydomonas*)比例要小),43%是属于复杂型,而其他类型所占的比例很小,这与水稻和拟南芥(*Arabidopsis thaliana*)的情况类似。在对内含子缺失的研究上^[49],发现玉米和高粱比谷子(*Setaria italica*)、水稻和短柄草,

具有更高的可重复内含子缺失比率,而且相邻的内含子和小的内含子更容易缺失。有内含子缺失的基因在生殖细胞或早期胚胎的表达概率更高。两侧外显子的A+T含量和内含子的TG/CG比例明显高于保留下来的内含子,表明甲基化过程可能对内含子缺失起到一定影响。

3 高粱功能基因组学

3.1 高粱不同品种间基因组结构变异研究

在高粱参考基因组发布以后,陆续有不同的研究组对几十个品种的高粱进行了重测序工作,并挖掘了大量的遗传变异信息。Nelson等^[50]首先重测序了8个籽实高粱品种并得到了SNP信息,验证了GBS方法运用在高粱上的可行性。Zheng等^[51]通过对2个甜高粱和1个籽实高粱品种的重测序,得到了包括SNP, INDEL, PAV (presence absence variation)以及CNV (copy number variation)在内的变异信息。通过对比这些变异数据,识别了它们在不同品种高粱基因组区域,基因家族,以及代谢通路中的分布差异。同时发现有将近1500个基因有差别地存在于籽实高粱和甜高粱的基因组中,这些基因主要参与了糖和淀粉的代谢、木质素和香豆素的生物合成、核苷酸的代谢、逆境应答和DNA的损伤修复。这一工作为探索通过重测序将高粱品种表型对应到基因型上的研究方法做出了有益的尝试。Emma等^[52]对包括地方种、改良种、野生高粱、拟高粱4个群体的44个高粱品种进行了重测序,通过野生种和改良种的变异数据识别出725个与驯化和作物改良有关的候选基因^[53]。图1展示了48个高粱品种在全基因组范围内SNP的分布情况,可以发现在染色体着丝粒区域变异频率较低,在染色体的两端变异较高。改良品种、地方品种、野生高粱以及拟高粱之间的SNP分布差异揭示了高粱具有遗传多样性,可加速高粱的遗传改良以及功能位点分析。

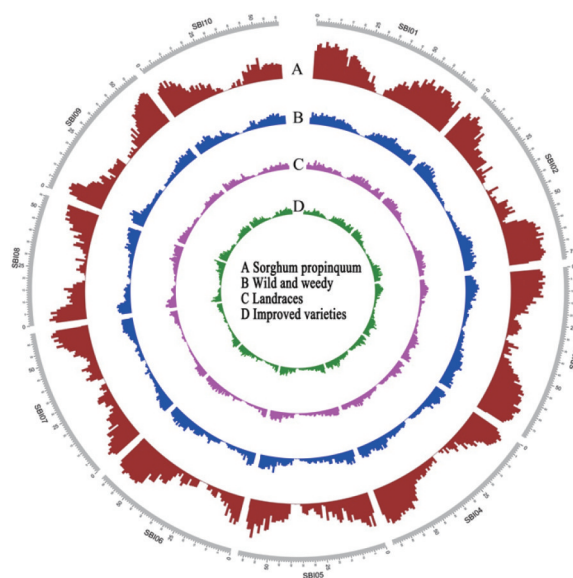


图1 48个高粱的SNP在全基因组中的分布

Fig. 1 Whole-genome distribution of SNP in 48 sorghum lines

大规模挖掘得到的变异数据,还适合作为高密度分子标记,为高粱群体育种提供正向遗传和基因组选择的材料。Zou等^[54]利用3418个SNPs构建了总长为1591 cM,包含10个连锁群的遗传图谱。Bekele等^[55]利用对3个甜高粱和2个籽实高粱重测序得到的数据构建了SNP芯片,并在包含564个基因型的高粱群体中进行测试,验证了其可靠性。Emma等整合了高粱物理及遗传图谱,将771 QTLs遗传位点定位到图谱上,其中利用构建的高粱NAM群体对前人研究的开花相关的遗传位点进行分析及挖掘^[53,56]。

除了SNP标记,更大片段的结构变异标记也被开发出来。Zhang等通过PCR及杂交等方法在高粱基因组内挖掘出51个大片获得与缺失变异(larger-size presence/absence variants, lsPAVs, >30 kb),分析这些变异的类别、数目、分布和遗传多样性,发现基因组间lsPAVs在自然的选择压力下存在着广泛的遗传多样性,这种现象与lsPAV影响的基因功能相关^[57]。Shen等分析发生在基因上的小片段PAVs(genic small-size PAVs, gss PAVs)变异特征,构建了首张高粱的PAV标记遗传图谱,并利用此图谱完成了不同地区种植的高粱群体的重要农艺学性状的QTLs定位分析^[58,59]。

近年来,使用GWAS方法对高粱进行研究的策略已经比较成熟。Morris等^[60]选取了971个遍布全球、适应于不同农业环境条件的高粱品种,通过酶切测序的方法高效经济地得到了数十万个SNP位点信息,通过这些SNPs构成的全基因组SNP图谱,并结合高粱品种的地理分布及形态特点,阐明了它们的群体结构及高粱由非洲传播到亚洲的方式,并且以此为基础进行了GWAS分析,定位了多个与高粱株高以及花序结构相关的候选基因。另外,有研究报道了对高粱籽实多酚含量^[61],以及高粱对茎腐病抗性^[62]的全基因组关联分析。

从这些结果中可以看出,高粱在基因组水平上的遗传变异研究已经取得了比较大的进展,得到了大量的基因组序列和结构变异的信息,同时利用这些信息完成了高粱表型变异和群体驯化及进化的初步研究,而接下来的工作则是更为高效地利用这些信息探究影响高粱重要性状的基因,真正意义上为高粱的育种工作提供便利。

3.2 高粱转录水平功能研究

高粱转录研究主要聚焦于各种胁迫条件下基因表达的变化,识别应激候选基因以及它们的调控机制。在非生物胁迫方面,Dugas等^[63]利用RNA-seq对高粱在渗透压和脱落酸胁迫下的转录本进行了测序,识别了一些相关的功能基因、转录因子以及基因调控网络。Johnson等^[64]利用表达谱芯片研究了高粱在热和干旱胁迫条件下的转录本,展示了在多种胁迫同时存在时,信号传导途径和基因调控网络的应激模式与单一胁迫下的差异。在生物胁迫方面,Mizuno^[65]利用RNA-seq对高粱在靶斑病(target leaf spot)胁迫下的转录本进行了研究,发现了一些串联重复旁系同源基因在应激过程中高度活跃,并且激活了柠檬酸循环通路中的乙醛酸分路,使其功

能从能量供给转换为细胞合成。在非胁迫研究方面,Jiang等^[66]利用RNA-seq测序甜高粱与籽实高粱的转录组,探索了不同品种高粱的基因组差异对其表达差异的影响。

利用转录本测序方法研究高粱miRNA是近年来的另一热点。传统上,高粱miRNA的识别主要采用同源基因预测的方法,很难发现新的miRNA家族,二代测序的普及大大推动和扩展了这一领域的研究进程。Calvino等^[67]对甜高粱和籽实高粱进行miRNA测序,识别了9个新的miRNA基因家族,并发现了不同高粱品种中miRNA基因表达差异对花期和茎秆含糖量的影响。Zhang等^[68]通过对甜高粱miRNA测序又识别了13个新的基因家族,并进行了时序表达分析和miRNA调控目标基因识别。Katiyar等^[69]针对新的测序数据,改进了生物信息学方法,为高粱miRNA预测提供了更多样的手段。

3.3 高粱关键基因及遗传位点的挖掘

高粱作为重要的粮食和能源作物,在20世纪90年代中期就已经开始了QTL定位的研究。自Paterson和Pereira等在1995年首次对高粱的种子大小、穗柄长和成熟时间等农艺学性状进行了QTL的定位^[70-72],目前已经挖掘出与植株形态建成、品质特性及抗逆性相关的遗传位点约858个^[73]。在植株形态建成方面,尤其是对于控制株高遗传位点的挖掘上,目前认为主要由4个QTLs控制,分别为*Dw1*、*Dw2*、*Dw3*和*Dw4*^[74-76],其中只有*Dw3*的基因被定位和克隆。控制*Dw3*的基因能够编码一个ATP结合的盒转运蛋白,通过突变这个基因,可以降低生长素的转运,使得高粱的茎节变得较短,从而影响高粱的整体株高^[77]。对于开花时间的研究,在早期鉴定为由4个QTLs控制,分别是*Ma1*、*Ma2*、*Ma3*和*Ma4*^[78,79],之后又有2个QTLs(*Ma5*和*Ma6*)被鉴定,但是只有*Ma1*的控制基因*SbPRR37*被克隆,这个基因通过调节高粱对光周期的响应来控制高粱的成熟期^[80]。高粱抗逆性的QTL定位研究主要集中在高粱的持绿性、抗旱性和耐冷性,而抗旱性的研究则是以基于高粱在干旱胁迫下的持绿能力作为主要指标的。高粱持绿性的QTL定位工作开展较早,涉及到高粱生长的不同时期,定位到的QTL位点范围广泛,在高粱的10条染色体上都有发现。尽管如此,现在比较稳定的持绿性QTLs主要有4个位点,分别是位于3号染色体的*stg1*和*stg2*,2号染色体的*stg3*,及5号染色体的*stg4*^[81,82]。

除此之外,挖掘出控制高粱抗病虫害^[83-85]、茎秆纤维含量^[86,87]、含糖量^[88]、籽实产量及品质^[89,90]及生物能源相关^[91-93]的QTL等。虽然大量的QTL位点被挖掘,但仅少数基因得到克隆,如控制落粒的基因*sh1*^[94],控制单宁合成的相关基因*Tannin1*^[95],调控高粱开花时间的*SbPRR37*^[96],高粱雄性不育恢复基因*Ry1*^[97]和耐铝毒害基因*AluSB*^[98],这些为主要农作物遗传育种提供重要的基因资源(表3)。随着高粱基因组测序工作的深入开展,将加速整合物理图谱及遗传图谱信息,这些有助于高粱关键遗传位点的深度挖掘和加速分子遗传育种工作。

表3 高粱重要基因

Table 3 Gene list of Sorghum

性状	表型	基因
形态特征	茎秆特征	<i>Tb1</i> ^[99] , <i>Dw2</i> ^[100,101] , <i>Dw3</i> ^[77] , <i>Sb.Ht9.1</i> ^[102]
	叶片特征	<i>Bmr6</i> ^[103] , <i>Bmr12</i> ^[104] , <i>Bmr2</i> ^[105]
	生育期	<i>Ma1</i> ^[100,101] , <i>Ma3</i> ^[106] , <i>Ma4</i> ^[107]
	植株颜色	<i>P</i> or <i>Q</i> ^[108]
	育性	<i>Rf2</i> ^[109] , <i>Rf1</i> ^[110]
	颖壳包被度 落粒性	<i>Gc</i> ^[111] <i>Sh1</i> ^[94]
品质特性	胚乳特征	<i>Sh2</i> ^[112] , <i>Bt2</i> ^[112] , <i>Ae1</i> ^[112] , <i>Wx</i> ^[112] , <i>O2</i> ^[112]
	籽实品质	<i>Y</i> ^[113] , <i>B2</i> ^[114] , <i>R</i> ^[115] , <i>I</i> ^[116] , <i>SbBADH2</i> ^[117]
	种子储存	<i>Gama-kafir</i> ^[118] , <i>Beta-kafir</i> ^[118] , <i>Delta-kafir</i> ^[118]
	单宁含量	<i>Tannin1</i> ^[95]
抗逆性	叶片持绿性	<i>stg1</i> , <i>stg2</i> , <i>stg3</i> , <i>stg4</i> ^[115,119]
	耐铝毒害	<i>AltSB</i> ^[120]
	氨转运相关 激素响应	<i>SbAMT3</i> ; <i>1</i> ^[121] , <i>SbAMT4</i> ^[121] <i>SbIAA1</i> ^[122] , <i>SbGH3-13</i> ^[122] , <i>SbLBD32</i> ^[122]

表4 高粱基因组学数据库

Table 4 Databases of sorghum genome

名称	分类	访问地址
Phytozome	序列注释	http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Sbicolor
Gramene	序列注释	http://archive.gramene.org/species/sorghum/sorghum_intro.html
Ensembl	序列注释	http://plants.ensembl.org/Sorghum_bicolor/Info/Index/
PlantGDB	序列注释	http://www.plantgdb.org/SbGDB/
PlantTFDB	转录因子	http://www.plantgdb.org/SbGDB/
miRBASE	microRNA	http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=sbi
MOROKOSHI	转录本	http://sorghum.riken.jp/morokoshi/
SorGSD	基因组变异	http://sorgsd.big.ac.cn
GRIN	种质资源	http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?454806

5 问题与展望

高粱基因组学经过这些年的发展,在新一代组学测序技术和组学分析方法的引领下,已经积累了一些成果。但与其他研究开展比较早、投入比较大的模式植物,如拟南芥、水稻和玉米等比较,依然处于初级阶段,存在一些问题和困难,同时也有广大的扩展空间。

目前,越来越多的高粱品种进行了重测序,利用重测序所得到的遗传变异信息(SNP、Indels、PAV及CNV等)可对高粱重要农艺性状(包括生物胁迫和非生物胁迫的耐受性以及生物能源产量相关的性状)的遗传位点或基因进行快速定位,但其精确定位还需要结合分析方法及实验设计的改进。甜高粱作为一种重要的能源植物,许多重要的性状,如秸秆含糖量与含汁量、生物量、胁迫(盐碱、低温、干旱、病虫害等)应答、生产性状(抗倒伏、分蘖、抗除草剂等)等都为复杂性状,统称为生物质能源综合症(biofuel syndrome)^[73],涉及到基因组多个位点的复杂控制与交互作用。下一代测序技术可

4 高粱组学数据库资源

随着高粱组学研究广泛深入的展开,产出了大量的基因组数据信息。建立组学数据库,将这些信息收集、分类、共享,对于高粱研究具有积极的意义。如表4所示,目前高粱组学数据库主要包括几类:一类是基因组序列和注释信息数据库,除了上文提到的Phytozome外, Gramene、Ensembl、PlantGDB也都提供高粱参考基因组的序列和注释信息,可以通过基因名称等关键词搜索相关信息,并可通过基因组浏览器进行各种可视化操作和数据下载;另外一类是高粱基因组二级注释数据库,主要是通过文献挖掘和生物信息预测,对基因组某一类功能原件的数据进行收集整理,包括PlantTFDB、miRBASE等;第三类数据库收集了主要由新一代测序技术得到的不同高粱品种转录本和基因组的变异数据,比如MOROKOSHI和SorGSD。另外,国际种质资源库GRIN收集了大量高粱品种的育种信息。

从基因组层面分析序列与结构变异,建立不同高粱品种间从籽实产量到株高、茎秆含糖量等表型变异与基因组变异关联,从而解析复杂性状的调控位点,加快甜高粱优良性状整合,获得集高生物量、高含糖量以及高抗逆性为一体的超级甜高粱,为生物能源产业提供原材料。目前仅仅拼接完成了一个籽实高粱品种(BTx623)的参考基因组,而且基因组拼接质量有待提高,依然存在很多未定位或者空缺的区段,这对后续进行高效精准的研究产生了很大的限制。因此,一方面需要对参考基因组进行更精确的拼接和注释,另一方面需要参照拟南芥、水稻及玉米等模式植物,对多个具有代表性的高粱品种,如与籽实高粱表型差异很大的甜高粱品种,或者具有代表性的地方品种(如中国吉林的Ji2731)进行基因组的De novo拼接,获得多个参考基因组乃至高粱的泛基因组“Pan Genome”。虽然二代测序技术的成本已经大大降低,但是由De novo方法获得精确的参考基因组依然是一项需要较大人力物力投入的工作。为此,新的低成本测序技术与新的

高效率分析方法需要应用到高粱中。首先,应利用三代或准三代测序技术(如PacBio^[123]),获得更长的测序读长,以应对高粱基因组杂合度以及重复序列含量较高的问题。另外,还可以利用单分子光学图谱技术(如Bionano^[124]和OpGen^[125]),以比传统BAC文库方法更经济和高效的方法构建高粱的物理图谱,辅助基因组拼接。

随着越来越多高粱基因组学数据的积累,如何有效利用这些大数据资源进行有效的分析也是亟待探索和解决的问题。首先要鼓励和倡导研究机构开发高粱的二级专业数据库,并积极共享和整合数据资源。其次,目前大多数的生物信息组学工具和算法,都是基于模式生物(比如人类基因组)来开发的。高粱和模式生物的基因组特性不同,这些工具和算法并非完全适用。因此,为得到更为精确可靠的数据,需要针对高粱基因组的特性,利用实验验证,对这些工具进行参数测试和结果矫正。第三,要善于利用并行大数据分析和云端存储技术,以应对越来越多的大规模全基因组关联分析和基因组选择分析所带来的硬件需求问题。

参考文献(References)

- [1] Belton P S, Taylor J R N. Sorghum and millets: Protein sources for Africa[J]. Trends in Food Science & Technology, 2004, 15(2): 94-98.
- [2] Mann J A, Kimber C T, Miller F R. The origin and early cultivation of sorghums in Africa[EB/OL]. [2015-07-13]. <http://hdl.handle.net/1969>.
- [3] De Wet J. Systematics and evolution of Sorghum sect. Sorghum (Gramineae) [J]. American Journal of Botany, 1978, 65(4): 477-484.
- [4] Doebley J F, Gaut B S, Smith B D. The molecular genetics of crop domestication[J]. Cell, 2006, 127(7): 1309-1321.
- [5] Zhang Limin, Liu Zhiqian, Chen Bingru, et al. Current status and application prospects of Csweet sorghum breeding in China[J]. Journal of China Agricultural University, 2012, 17(6): 76-82.
- [6] 张彩霞, 谢高地, 李士美, 等. 中国能源作物甜高粱的空间适宜分布及乙醇生产潜力[J]. 生态学报, 2010, 30(17): 4765-4770.
Zhang Caixia, Xie Gaodi, Li Shimei, et al. Spatial suitability and its bio-ethanol potential of sweet sorghum in China[J]. Acta Ecologica Sinica, 2010, 30(17): 4765-4770.
- [7] Woo S S, Jiang J, Gill B S, et al. Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor[J]. Nucleic Acids Research, 1994, 22(23): 4922-4931.
- [8] Lin Y R, Zhu L, Ren S, et al. A Sorghum propinquum BAC library, suitable for cloning genes associated with loss-of-function mutations during crop domestication[J]. Molecular Breeding, 1999, 5(6): 511-520.
- [9] Peterson D G, Schulze S R, Sciara E B, et al. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery[J]. Genome Research, 2002, 12(5): 795-807.
- [10] Peterson D G, Wessler S R, Paterson A H. Efficient capture of unique sequences from eukaryotic genomes[J]. Trends in Genetics, 2002, 18(11): 547-550.
- [11] Pratt L H, Liang C, Shah M, et al. Sorghum expressed sequence tags identify signature genes for drought, pathogenesis, and skotomorphogenesis from a milestone set of 16, 801 unique transcripts [J]. Plant Physiology, 2005, 139(2): 869-884.
- [12] Hulbert S H, Richter T E, Axtell J D, et al. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes[J]. PNAS, 1990, 87(11): 4251-4255.
- [13] Whitkus R, Doebley J, Lee M. Comparative genome mapping of sorghum and maize[J]. Genetics, 1992, 132(4): 1119-1130.
- [14] Tao Y Z, Jordan D R, Henzell R G, et al. Construction of a genetic map in a sorghum recombinant inbred line using probes from different sources and its comparison with other sorghum maps[J]. Australian Journal of Agricultural Research, 1998, 49(4): 729-736.
- [15] Boivin K, Deu M, Rami J F, et al. Towards a saturated sorghum map using RFLP and AFLP markers[J]. Theoretical and Applied Genetics, 1999, 98(2): 320-328.
- [16] Menz M A, Klein R R, Mullet J E, et al. A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP (R), RFLP and SSR markers[J]. Plant Molecular Biology, 2002, 48(5): 483-499.
- [17] Bowers J E, Abbey C, Anderson S, et al. A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses[J]. Genetics, 2003, 165(1): 367-386.
- [18] Mace E S, Xia L, Jordan D R, et al. DArT markers: Diversity analyses and mapping in *Sorghum bicolor*[J]. BMC Genomics, 2008, 9: 26.
- [19] Margulies M, Egholm M, Altman W E, et al. Genome sequencing in microfabricated high-density picolitre reactors[J]. Nature, 2005, 437(7057): 3376-3380.
- [20] Hunkapiller T, Kaiser R, Koop B, et al. Large-scale and automated DNA sequence determination[J]. Science, 1991, 254(5028): 59-67.
- [21] Fedurco M, Romieu A, Williams S, et al. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies[J]. Nucleic Acids Research, 34(3): e22.
- [22] Metzker M L. Sequencing technologies—The next generation[J]. Nature Reviews Genetics, 2010, 11(1): 31-46.
- [23] Mardis E R. Next-generation DNA sequencing methods[J]. Annual Review of Genomics and Human Genetics, 2008, 9(1): 387-402.
- [24] Shendure J, Ji H. Next-generation DNA sequencing[J]. Nature Biotechnology, 2008, 26(10): 1135-1145.
- [25] Van Tassel C P, Smith T P L, Matukumalli L K, et al. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries[J]. Nature Methods, 2008, 5(3): 247-252.
- [26] Ganai M W, Altmann T, Röder M S. SNP identification in crop plants [J]. Current Opinion in Plant Biology, 2009, 12(2): 211-217.
- [27] Rounsley S D, Last R L. Shotguns and SNPs: How fast and cheap sequencing is revolutionizing plant biology[J]. Plant Journal, 2010, 61(6): 922-927.
- [28] Nielsen R, Paul J S, Albrechtsen A, et al. Genotype and SNP calling from next-generation sequencing data[J]. Nature Reviews Genetics, 2011, 12(6): 443-451.
- [29] Spindel J, Wright M, Chen C, et al. Bridging the genotyping gap: Using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations[J]. Theoretical and Applied Genetics, 2013, 126(11): 2699-2716.
- [30] Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics[J]. Nature Reviews Genetics, 2009, 10(1): 57-63.
- [31] Haas B J, Zody M C. Advancing RNA-seq analysis[J]. Nature Biotechnology, 2010, 28(5): 421-423.
- [32] Paterson A H, Bowers J E, Feltus F A, et al. Comparative genomics of

- grasses promises a bountiful harvest[J]. *Plant Physiology*, 2009, 149(1): 125-131.
- [33] Swigonova Z, Lai J, Ma J, et al. Close split of sorghum and maize genome progenitors[J]. *Genome Research*, 2004, 14(10A): 1916-1923.
- [34] al-Janabi S M, Honeycutt R J, McClelland M, et al. A genetic linkage map of *Saccharum spontaneum* L. SES 208[J]. *Genetics*, 1993, 134(4): 1249-1260.
- [35] Ming R, Liu S C, Lin Y R, et al. Detailed alignment of saccharum and sorghum chromosomes: Comparative organization of closely related diploid and polyploid genomes[J]. *Genetics*, 1998, 150(4): 1663-1682.
- [36] Wang J, Roe B, Macmill S, et al. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes[J]. *BMC Genomics*, 2010, 11(1): 261.
- [37] 郑成木. 甘蔗核型及其染色体数目变化的研究[J]. *热带作物学报*, 1993, 14(1): 47-51.
Zheng Chengmu. Karyotypes and variations of chromosome number in sugar cane[J]. *Chinese Journal of Tropical Crops*, 1993, 14(1): 47-51.
- [38] Wang X, Tang H, Bowers J E, et al. Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization[J]. *Genome Research*, 2009, 19(6): 1026-1032.
- [39] Jiang S Y, Ma Z G, Vanitha J, et al. Genetic variation and expression diversity between grain and sweet sorghum lines[J/OL]. *BMC Genomics*, 2013, 14: doi: 10.1186/1471-2164-14-18.
- [40] Massa A N, Wanjugi H, Deal K R, et al. Gene space dynamics during the evolution of *Aegilopstauschii*, *Brachypodiumdistachyon*, *Oryza sativa*, and *Sorghum bicolor* genomes[J]. *Molecular Biology and Evolution*, 2011, 28(9): 2537-2547.
- [41] Jin T, Chen J, Zhu L, et al. Comparative mapping combined with homology-based cloning of the rice genome reveals candidate genes for grain zinc and iron concentration in maize[J]. *BMC Genetics*, 2015, 16: 17: doi: 10.1186/s12863-015-0176-1.
- [42] Paterson A H, Bowers J E, Bruggmann R, et al. The *Sorghum bicolor* genome and the diversification of grasses[J]. *Nature*, 2009, 457(7229): 551-556.
- [43] Schnable P S, Ware D, Fulton R S, et al. The B73 maize genome: Complexity, diversity, and dynamics[J]. *Science*, 2009, 326(5956): 1112-1115.
- [44] Jiang S Y, Ramachandran S. Genome-wide survey and comparative analysis of LTR retrotransposons and their captured genes in rice and sorghum[J]. *PLoS One*, 2013, 8(7): e71118.
- [45] Sharma A, Presting G G. Evolution of centromeric retrotransposons in grasses[J]. *Genome Biology and Evolution*, 2014, 6(6): 1335-1352.
- [46] Roulin A, Piegu B, Fortune P M, et al. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae[J]. *BMC Evolutionary Biology*, 2009, 9: 58-67.
- [47] Yoshida S, Maruyama S, Nozaki H, et al. Horizontal gene transfer by the parasitic plant strigahermonthica[J]. *Science*, 2010, 328(5982): 1128.
- [48] Panahi B, Abbaszadeh B, Taghizadeghan M, et al. Genome-wide survey of alternative splicing in *Sorghum bicolor*[J]. *Physiology and Molecular Biology of Plants: An International Journal of Functional Plant Biology*, 2014, 20(3): 323-329.
- [49] Wang H, Devos K M, Bennetzen J L. Recurrent loss of specific introns during angiosperm evolution[J]. *PLOS Genetics*, 2014, 10(12): doi: 10.1371/journal.pgen.1004843.
- [50] Nelson J C, Wang S, Wu Y, et al. Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum[J]. *BMC Genomics*, 2011, 12: 352.
- [51] Zheng L Y, Guo X S, He B, et al. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*)[J]. *Genome Biology*, 2011, 12(11): R114.
- [52] Mace E S, Tai S, Gilding E K, et al. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum[J]. *Nature Communications*, 2013: doi: 10.1038/ncomms3320.
- [53] Gilding E K, Frère C H, Cruickshank A, et al. Allelic variation at a single gene increases food value in a drought-tolerant staple cereal[J]. *Nature Communications*, 2013: doi: 10.1038/ncomms2450.
- [54] Zou G, Zhai G, Feng Q, et al. Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods[J]. *Journal of Experimental Botany*, 2012, 63(15): 5451-5462.
- [55] Bekele W A, Wieckhorst S, Friedt W, et al. High-throughput genomics in sorghum: From whole-genome resequencing to a SNP screening array[J]. *Plant Biotechnology Journal*, 2013, 11(9): 1112-1125.
- [56] Mace E S, Jordan D R. Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement[J]. *Theoretical and Applied Genetics*, 2011, 123(1): 169-191.
- [57] Zhang L M, Luo H, Liu Z Q, et al. Genome-wide patterns of large-size presence/absence variants in sorghum[J]. *Journal of Integrative Plant Biology*, 2014, 56(1): 24-37.
- [58] Shen X, Liu Z Q, Mocoer A, et al. PAV markers in Sorghum bicolor: genome pattern, affected genes and pathways, and genetic linkage map construction[J]. *TAG Theoretical and Applied Genetics Theoretische and AngewandteGenetik*, 2015, 128(4): 623-637.
- [59] Mocoer A, Zhang Y M, Liu Z Q, et al. Stability and genetic control of morphological, biomass and biofuel traits under temperate maritime and continental conditions in sweet sorghum (*Sorghum bicolor*)[J]. *Theoretical and Applied Genetics*, 2015: doi: 10.1007/s00122-015-2538-5.
- [60] Morris G P, Ramu P, Deshpande S P, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum[J]. *PNAS*, 2012, 110(2): 453-458.
- [61] Rhodes D H, Hoffmann L, Rooney W L, et al. Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) moench] germplasm[J]. *Journal of Agricultural and Food Chemistry*, 2014, 62(45): 10916-10927.
- [62] Adeyanju A, Little C, Yu J, et al. Genome-wide association study on resistance to stalk rot diseases in grain sorghum[J]. *G3: Genes/Genomes/Genetics*, 2015: doi: 10.1534/g3.114.016394.
- [63] Dugas D, Monaco M, Olson A, et al. Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid[J]. *BMC Genomics*, 2011, 12(1): 514.
- [64] Johnson S, Lim F L, Finkler A, et al. Transcriptomic analysis of Sorghum bicolor responding to combined heat and drought stress[J]. *BMC Genomics*, 2014, 15(1): 456.
- [65] Mizuno H, Kawahigashi H, Kawahara Y, et al. Global transcriptome analysis reveals distinct expression among duplicated genes during sorghum-Bipolarissorghicolainteraction[J]. *BMC Plant Biology*, 2012,

- 12(1): 121.
- [66] Jiang S Y, Ma Z, Vanitha J, et al. Genetic variation and expression diversity between grain and sweet sorghum lines[J]. BMC Genomics, 2013, 14(1): 18.
- [67] Calvino M, Bruggmann R, Messing J. Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems[J]. BMC Genomics, 2011, 12(1): 356.
- [68] Zhang L, Zheng Y, Jagadeeswaran G, et al. Identification and temporal expression analysis of conserved and novel microRNAs in Sorghum[J]. Genomics, 2011: doi: 10.1016/j.ygeno.2011.08.005.
- [69] Katiyar A, Smita S, Chinnusamy V, et al. Identification of miRNAs in sorghum by using bioinformatics approach[J]. Plant Signaling & Behavior, 2012, 7(2): 246–259.
- [70] Paterson A H, Lin Y R, Li Z K, et al. Convergent domestication of cereal crops by independent mutations at corresponding genetic-loci [J]. Science, 1995, 269(5231): 1714–1718.
- [71] Paterson A H, Schertz K F, Lin Y R, et al. The weediness of wild plants—Molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) pers[J]. Proceedings of the National Academy of Sciences of the United States of America, 1995, 92(13): 6127–6131.
- [72] Pereira M G, Lee M. Identification of genomic regions affecting plant height in sorghum and maize[J]. Theoretical and Applied Genetics, 1995, 90(3–4): 380–388.
- [73] Anami S E, Zhang L M, Xia Y, et al. Sweet sorghum ideotypes: Genetic improvement of stress tolerance[J]. Food and Energy Security, 2015, 4 (1): 3–24.
- [74] Quinby J R, Karper R E. Inheritance of height in sorghum[J]. Agronomy Journal, 1954, 46(5): 211–216.
- [75] Feltus F A, Hart G E, Schertz K F, et al. Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations[J]. Theoretical and Applied Genetics, 2006, 112(7): 1295–1305.
- [76] Upadhyaya H D, Wang Y H, Gowda C L L, et al. Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection[J]. Theoretical and Applied Genetics, 2014, 127(6): 1461.
- [77] Multani D S, Briggs S P, Chamberlin M A, et al. Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants [J]. Science, 2003, 302(5642): 81–84.
- [78] Quinby J R, Karper R E. The inheritance of 3 genes that influence time of floral initiation and maturity date in milo[J]. Journal of the American Society of Agronomy, 1945, 37(11): 916–936.
- [79] Quinby J R. 4th maturity gene locus in sorghum[J]. Crop Science, 1966, 6(6): 516.
- [80] Murphy R L, Klein R R, Morishige D T, et al. Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum[J]. PNAS, 2011, 108(39): 16469–16474.
- [81] Reddy N R R, Ragimasalawada M, Sabbavarapu M M, et al. Detection and validation of stay-green QTL in post-rainy sorghum involving widely adapted cultivar, M35-1 and a popular stay-green genotype B35[J]. BMC Genomics, 2014: doi: 10.1186/1471-2164-15-909.
- [82] Haussmann B I G, Mahalakshmi V, Reddy B V S, et al. QTL mapping of stay-green in two sorghum recombinant inbred populations[J]. Theoretical and Applied Genetics, 2002, 106(1): 133–142.
- [83] Tao Y Z, Hardy A, Drenth J, et al. Identifications of two different mechanisms for sorghum midge resistance through QTL mapping[J]. Theoretical and Applied Genetics, 2003, 107(1): 116–122.
- [84] Punnuri S, Huang Y, Steets J, et al. Developing new markers and QTL mapping for greenbug resistance in sorghum *Sorghum bicolor* (L.) Moench[J]. Euphytica, 2013, 191(2): 191–203.
- [85] Satish K, Madhusudhana R, Padmaja P G, et al. Development, genetic mapping of candidate gene-based markers and their significant association with the shoot fly resistance quantitative trait loci in sorghum *Sorghum bicolor* (L.) Moench[J]. Molecular Breeding, 2012, 30(4): 1573–1591.
- [86] Shiringani A L, Friedt W. QTL for fibre-related traits in grain x sweet sorghum as a tool for the enhancement of sorghum as a biomass crop [J]. Theoretical and Applied Genetics, 2011, 123(6): 999–1011.
- [87] Saballos A, Sattler S E, Sanchez E, et al. Brown midrib2 (*Bmr2*) encodes the major 4-coumarate: Coenzyme a ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench)[J]. The Plant Journal, 2012, 70(5): 818–830.
- [88] Shiringani A L, Frisch M, Friedt W. Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench [J]. Theoretical and Applied Genetics, 2010, 121(2): 323–336.
- [89] Kumar T, Dweikat I, Sato S, et al. Modulation of kernel storage proteins in grain sorghum (*Sorghum bicolor* (L.) Moench) [J]. Plant Biotechnology Journal, 2012, 10(5): 533–544.
- [90] Mkandawire N L, Kaufman R C, Bean S R, et al. Effects of Sorghum (*Sorghum bicolor* (L.) Moench) tannins on α -amylase activity and in vitro digestibility of starch in raw and processed flours[J]. Journal of Agricultural and Food Chemistry, 2013, 61(18): 4448–4454.
- [91] Brown P J, Klein P E, Bortiri E, et al. Inheritance of inflorescence architecture in sorghum[J]. Theoretical and Applied Genetics, 2006, 113(5): 931–942.
- [92] Guan Y A, Wang H L, Qin L, et al. QTL mapping of bio-energy related traits in Sorghum[J]. Euphytica, 2011, 182(3): 431–440.
- [93] Felderhoff T J, Murray S C, Klein P E, et al. QTLs for energy-related traits in a sweet×grain sorghum *Sorghum bicolor* (L.) Moench mapping population[J]. Crop Science, 2012, 52(5): 2040–2049.
- [94] Lin Z, Li X, Shannon L M, et al. Parallel domestication of the Shattering1 genes in cereals[J]. Nature Genetics, 2012, 44(6): 720–724.
- [95] Wu Y, Li X, Xiang W, et al. Presence of tannins in sorghum grains is conditioned by different natural alleles of Tannin1[J]. PNAS, 2012, 109(26): 10281–10286.
- [96] Murphy R L, Klein R R, Morishige D T, et al. Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum[J]. PNAS, 2011, 108(39): 16469–16474.
- [97] Klein P E, Klein R R, Vrebalov J, et al. Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement [J]. Plant Journal, 2003, 34(5): 605–621.
- [98] Caniato F F, Guimaraes C T, Schaffert R E, et al. Genetic diversity for aluminum tolerance in sorghum[J]. Theoretical and Applied Genetics, 2007, 114(5): 863–876.
- [99] Kebrom T H, Burson B L, Finlayson S A. Phytochrome B represses Teosinte Branched1 expression and induces sorghum axillary bud outgrowth in response to light signals[J]. Plant Physiology, 2006, 140 (3): 1109–1117.

- [100] Klein R R, Mullet J E, Jordan D R, et al. The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping[J]. *Crop Science*, 2008, 48: S12-S26.
- [101] Lin Y R, Schertz K F, Paterson A H. Comparative-analysis of QTLs affecting plant height and maturity across the poaceae, in reference to an interspecific sorghum population[J]. *Genetics*, 1995, 141(1): 391-411.
- [102] Brown P J, Rooney W L, Franks C, et al. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes[J]. *Genetics*, 2008, 180(1): 629-637.
- [103] Saballos A, Ejeta G, Sanchez E, et al. A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies SbCAD2 as the brown midrib6 gene[J]. *Genetics*, 2009, 181(2): 783-795.
- [104] Bout S, Vermerris W. A candidate-gene approach to clone the sorghum brown midrib gene encoding caffeic acid O-methyltransferase[J]. *Molecular Genetics and Genomics*, 2003, 269(2): 205-214.
- [105] Saballos A, Sattler S E, Sanchez E, et al. Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench) [J]. *Plant Journal*, 2012, 70(5): 818-830.
- [106] Childs K L, Miller F R, CordonnierPratt M M, et al. The sorghum photoperiod sensitivity gene, Ma(3), encodes a phytochrome B[J]. *Plant Physiology*, 1997, 113(2): 611-619.
- [107] Hart G E, Schertz K F, Peng Y, et al. Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters[J]. *Theoretical and Applied Genetics*, 2001, 103(8): 1232-1242.
- [108] Klein R R, Rodriguez-Herrera R, Schlueter J A, et al. Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum[J]. *Theoretical and Applied Genetics*, 2001, 102(2-3): 307-319.
- [109] Jordan D R, Mace E S, Hanzell R G, et al. Molecular mapping and candidate gene identification of the Rf2 gene for pollen fertility restoration in sorghum [*Sorghum bicolor* (L.) Moench][J]. *Theoretical and Applied Genetics*, 2010, 120(7): 1279-1287.
- [110] Klein R R, Klein P E, Mullet J E, et al. Fertility restorer locus Rf1 of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12[J]. *Theoretical and Applied Genetics*, 2005, 111(6): 994-1012.
- [111] Srinivas G, Satish K, Madhusudhana R, et al. Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum[J]. *Theoretical and Applied Genetics*, 2009, 118(8): 1439-1454.
- [112] de AlencarFigueiredo L F, Sine B, Chantreau J, et al. Variability of grain quality in sorghum: Association with polymorphism in Sh2, Bt2, SssI, AeI, Wx and O2[J]. *Theoretical and Applied Genetics*, 2010, 121(6): 1171-1185.
- [113] Knoll J, Gunaratna N, Ejeta G. QTL analysis of early-season cold tolerance in sorghum[J]. *Theoretical and Applied Genetics*, 2008, 116(4): 577-587.
- [114] Rami J F, Dufour P, Trouche G, et al. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench)[J]. *Theoretical and Applied Genetics*, 1998, 97(4): 605-616.
- [115] Xu W W, Subudhi P K, Crasta O R, et al. Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench)[J]. *Genome*, 2000, 43(3): 461-469.
- [116] Tao Y Z, Hanzell R G, Jordan D R, et al. Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments[J]. *Theoretical and Applied Genetics*, 2000, 100(8): 1225-1232.
- [117] Yundaeng C, Sonta P, Tangphatsornruang S, et al. Gene discovery and functional marker development for fragrance in sorghum (*Sorghum bicolor* (L.) Moench)[J]. *Theoretical and Applied Genetics*, 2013, 126(11): 2897-2906.
- [118] Laidlaw H K C, Mace E S, Williams S B, et al. Allelic variation of the beta-, gamma- and delta- kafirin genes in diverse Sorghum genotypes[J]. *Theoretical and Applied Genetics*, 2010, 121(7): 1227-1237.
- [119] Borrell A K, Oosterom E J, Mullet J E, et al. Stay green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns[J]. *New Phytologist*, 2014.
- [120] Magalhaes J V, Liu J, Guimaraes C T, et al. A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum[J]. *Nature Genetics*, 2007, 39(9): 1156-1161.
- [121] Koegel S, Lahmidi N A, Arnould C, et al. The family of ammonium transporters (AMT) in *Sorghum bicolor*: Two AMT members are induced locally, but not systemically in roots colonized by arbuscularmycorrhizalfungi[J]. *New Phytologist*, 2013, 198(3): 853-865.
- [122] Wang S, Bai Y, Shen C, et al. Auxin-related gene families in abiotic stress response in *Sorghum bicolor*[J]. *Functional & Integrative Genomics*, 2010, 10(4): 533-546.
- [123] Carneiro M, Russ C, Ross M, et al. Pacific biosciences sequencing technology for genotyping and variation discovery in human data[J]. *BMC Genomics*, 2012, 13(1): 375.
- [124] Hastie A R, Dong L, Smith A, et al. Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilopstauschii* genome[J]. *PLoS One*, 2013, 8(2): e55864.
- [125] Dong Y, Xie M, Jiang Y, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*)[J]. *Nature Biotechnology*, 2013, 31(2): 135-141.

(责任编辑 王媛媛)

《科技导报》“卷首语”栏目征稿

“卷首语”栏目每期邀请一位中国科学院院士和中国工程院院士就重大科技现象、事件,以及学科发展趋势、科学研究热点和前沿问题等,撰文发表个人的见解、意见和评论。本栏目欢迎院士投稿,每篇文章约2000字,同时需提供作者学术简历、工作照和签名电子文档。投稿邮箱:kjdbbjb@cast.org.cn。