

# 高粱基因组学研究进展

卢 峰<sup>1</sup>, 吕香玲<sup>2</sup>

(1. 辽宁省农业科学院 作物研究所, 辽宁 沈阳 110161; 2. 沈阳农业大学 农学院, 辽宁 沈阳 110161)

**摘要:**高粱是世界上最主要的禾谷类作物之一, 是日趋重要的生物燃料作物。同时, 它也是一种最有害的杂草的先祖, 是许多具有复杂基因组的热带草的植物学模型, 开展其基因组学研究具有重要意义。目前已经构建了高粱种内和种间的遗传图谱, 完成了高粱全基因组的测序, 这为探索高粱基因和功能的对位提供了保障。笔者对高粱基因组遗传图谱、物理图谱及基因组特征、序列分析、QTL 定位及关联遗传学等几方面的研究进展进行了整理和总结, 为开展相关研究工作及想了解相关信息的广大学者提供了参考信息和研究方向的建议。

**关键词:**高粱; 基因组学; 遗传图谱; 物理图谱; 序列分析

中图分类号: S514; Q34 文献标识码: A 文章编号: 1000- 7091( 2008) 增刊- 0149- 04

## Progresses on Sorghum Genomics

LU Feng<sup>1</sup>, LU Xiang ling<sup>2</sup>

(1. Crop Research Institute, Liaoning Academy of Agriculture Sciences, Shenyang 110161, China;

2. College of Agronomy, Shenyang Agricultural University, Shenyang 110161, China)

**Abstract:** Genomics study on sorghum is very meaningful as it is one of the world's leading cereal crops, a biofuel crop with high and growing importance, a progenitor of one of the world's most noxious weeds and a botanical model for many tropical grasses with complex genomes. Up to now, intraspecific and interspecific sorghum genetic maps have been constructed and the whole genome sequencing completed. This provides a foundation for invigorating progress toward relating sorghum genes to their functions. Research progresses on sorghum genome such as genetic map, physical map, sequence analysis, QTL mapping and correlative genetics were collected and summarized in this article. And hopefully it can provide useful reference and suggestion for scholars endeavoring on relevant research works.

**Key words:** Sorghum; Genomics; Genetic map; Physical map; Sequence analysis

作为粮饲兼用的作物, 高粱 (*Sorghum bicolor* (L.) Moench) 是全球农业生态系统中一个重要的保障。高粱抗逆性强, 在一些降雨量很少, 多数粮食作物生长受抑的地区 (如非州东北部和美国南部平原), 占有重要地位。随着有限的淡水资源需求的增加、边际土地利用的增长和全球的气候变化趋势, 高粱这类抗旱作物对于满足全世界日益增多的人口对粮食和饲料的需求必将起到越来越重要的作用。

高粱也是生物燃料作物。在美国, 高粱是仅次于玉米的第二大粮食生产乙醇原料, 并且以需水量少, 市场价格低, 乙醇出产率高等优势越来越受重视。目前, 开发高粱纤维素的生物燃料比以籽粒为原料更有优势, 除了出产率高外, 还能够利用边际性

土地而减少单位面积内生物产量的成本<sup>[1]</sup>。最近对高粱的多年生遗传控制机理和基因组学的研究<sup>[2, 3]</sup>, 为高粱采用多年再生种植方式奠定了基础, 这将有助于进一步降低生物燃料生产成本, 同时也给高粱作为纤维素的生物燃料作物带来更大的期望, 尤其作为青饲和青贮作物的高茎秆含糖量的甜高粱<sup>[4]</sup>更是大有前途。

通过比较遗传学的研究发现, 禾本科、十字花科、豆科、茄科和一些树种的基因组, 虽然经过了数百万年的演化, 它们的基因内容和基因排列次序都是相当保守的, 即相互之间表现出共线性 (Colinearity)。其中, 以禾本科植物的共线性最明显。作为禾谷类作物基因组研究的模式植物, 水稻基因组全序

收稿日期: 2008- 09- 19

基金项目: 国际合作项目 (CFC/FIGG/32); 科技部国家科技支撑项目 (2006BAD02B03)

作者简介: 卢 峰 (1974- ), 男, 辽宁沈阳人, 博士, 副研究员, 主要从事高粱遗传育种研究。

列已经公布。相比较而言,水稻更能代表起源温带的  $C_3$  光合作用途径的草类。高粱可以作为起源热带的草类的代表,它们都是  $C_4$  光合作用途径,利用生化和形态学的特化作用来保证在高温条件下提高碳同化作用。与水稻相比,高粱与玉米、甘蔗等基因组较大的禾谷类作物亲缘较近。在 500 万年前甘蔗与高粱是共同祖先,仍保持着相似的基因序列,甚至它们的属间杂交能产生很有活力的后代。玉米与高粱分歧后进行了一次全基因组的复制,甘蔗至少进行了两次全基因组的复制<sup>[5]</sup>。与玉米和甘蔗等其他禾谷类作物相比,高粱基因组重复少,基因组小,似乎可以作为禾谷类作物基因组结构、功能和进化研究的又一重要模式植物。

目前,已经完成了高粱全基因组的测序,这为探索高粱基因和功能打下了基础。随着一些高分辨率的高粱遗传、物理和细胞图谱加速构建,重要基因的克隆和功能鉴定日益增多。这些研究进展促进了高粱特有的多种适应逆境胁迫等优异基因资源的发掘及其在作物改良中的应用,更有利于全面认识禾谷类作物基因组的进化、遗传、变异机制并推进禾谷类作物功能基因组学的研究。

## 1 高粱基因组遗传图谱

分子遗传图谱的构建是遗传学研究的重要领域,是系统开展基因组研究的第一步。分子遗传图谱在遗传育种研究中也具有广泛的用途,在拥有一张高密度分子图谱的前提下,通过性状调查和遗传分析,可以将控制某一性状的基因定位于图谱的某一位置,从而获得性状连锁的标记用于标记辅助选择,或者进行目标基因的图位克隆。许多遗传图谱都结合重要农艺性状特别是数量性状基因的定位而构建,使图谱成为基因组研究与育种实践之间相互联系的桥梁。

高粱(*S. bicolor*)为二倍体常异交作物,品种间 DNA 多态性高,这使其较容易进行遗传作图。与高粱的 10 对染色体( $2n=20$ )相对应,构建的遗传图谱连锁群数目应为 10。目前已经获得了高密度的种内(*S. bicolor*)和种间(*S. bicolor* × *S. propinquum*)遗传图谱。其中最经典的两张图谱为: Menz 等利用 BTx623 × IS3620C 重组自交系群体构建的遗传图谱和 Bowers 等人发表的一张由 *S. bicolor* 和 *S. propinquum* 杂交组建的  $F_2$  群体构建的遗传图谱。这两张图谱遗传标记密度较大,标记类型丰富,两张图谱包含高粱 10 个连锁群,提供了 2600 STS 标记(基于已测序的低拷贝探针)、2 454 个 AFLP 标记和 1 375 个

sequence scanned 标记(基于锚定的 BAC 克隆序列)。在高粱连锁群上的 SSR 标记,来自甘蔗、玉米、水稻、大麦、燕麦、拟南芥等物种的同源探针以及水稻、大麦、燕麦和玉米 cDNA 和 gDNA 序列开发的同源标记,是有效用于比较基因组学的参考标记。这两张遗传图谱有一个共用的亲本(*S. bicolor*“BTx623”),本质上是共线性的,对高粱基因组学的深入研究有着重要意义。目前,在此基础上,已经公布了多张用于高粱 QTL 定位分析的遗传连锁图谱。一些高粱自身序列标记也已被直接定位于其它类群上或可以根据序列相似性标绘出来。这就为禾本科作物间基因组的比较研究提供了便利。这些图谱和标记可以从 <http://www.plantgenome.uga.edu/sorghummap> and [www.gamene.org](http://www.gamene.org) 网站上获取。

## 2 高粱基因组物理图谱及基因组特征

物理图谱是利用限制性内切酶将染色体切成片段,再根据重叠序列确定片段间连接顺序,以及遗传标志之间物理距离[碱基对(bp)或千碱基(kb)或兆碱基(Mb)]的图谱。DNA 物理图谱是顺序测定的基础,也可理解为指导 DNA 测序的蓝图。广义地说, DNA 测序从物理图谱制作开始,它是测序工作的第一步。

高粱基因组的大小是 700(基于科特分析)<sup>[8]</sup> ~ 772 Mbp(基于流式细胞计量术)<sup>[9]</sup>。这表明高粱基因组的大小比水稻的基因组大 60%, 但仅相当于玉米和人类基因组大小的 1/4。DNA 复性动力学分析表明高粱基因组是由 16% 的折回 DNA、15% 的高度重复 DNA(个别家系的基因组可达 5 200 个拷贝)、41% 的中度重复 DNA(平均 72 拷贝)和 24% 的低拷贝 DNA。在高 Cot 值时仍保持单链状态的 DNA 占大约 4%,估计是已被损坏的 DNA<sup>[8]</sup>。

高粱是最先建立并发布 BAC 文库的被子植物。BTx623、*S. propinquum* 和 IS3620C 都有可用的高覆盖度的 BAC 文库,其中 BTx623 分别为 *Hind* III(12X) 和 *Bam*HI(8X) 酶切; *S. propinquum* 分别为 *Eco*R I (~7X) 与 *Hind* III (~7X) 共酶切(13~14X); IS3620C 为 *Hind* III 酶切(~9X)。BTx623 BACs 基于琼脂糖电泳获得 69 545 个指纹,利用 7 292 个探针锚定出 211 558 个杂交位点。同样地, *S. propinquum* BACs 基于琼脂糖电泳获得 40 957 个指纹,利用 7 481 个探针锚定出 189 735 个杂交位点。并通过附加的 contig 末端 BACs HICF 填充间隙。这些信息都已经作为可查阅的物理图谱被整合于网站上(<http://algodon.tamu.edu/sorghumdb/sorghum96/preview.html>)。

大约有 456 个 *S. propinquum* 和 303 个 *S. bicolor* BAC contigs(占 BACs 的 41%, 占单拷贝位点的 80%) 似乎位于常染色质区域。随着额外锚定, 归属于常染色质的基因组比例可能上升。BAC contigs 占 BACs 的 41%, 然而高粱基因组 DNA 只有 24% 是单拷贝或低拷贝<sup>[8]</sup>。这一研究结果表明, 高粱常染色质是由低拷贝序列和重复 DNA 序列组成。

### 3 高粱基因组序列分析

利用鸟枪法对美国的优良高粱自交系 BTx623 进行测序, 获得了 10.5 百万 reads, 覆盖 8 倍基因组, 序列数据现存于 NCBI Trace Archive。早期的分析已经证明高粱基因组序列对高粱进行完整而高质量的注释将是一个基础。在初步的整合中, 超过 97% 的高粱蛋白质编码基因(ESTs) 被定位于 250 个长的连锁群上。它们中的绝大部分可以通过遗传和物理图谱进行连接、排序和定向, 以用于重新构建完整的染色体。对高粱甲基化过滤序列的初步装配, 高粱、玉米和甘蔗转录装配以及拟南芥和水稻蛋白质组的校正确保了装配的精准度和正确的蛋白质编码位点。

来源于简化代表性测序的附加资料将对基因组序列中表达部分的鉴定有很大作用。目前, 大约 204 000 个序列表达标签代表了高粱的基因空间, 它们中多数簇集为 22 000 个 unigenes, 并且代表了来自几个基因型的 20 多个文库<sup>[10]</sup>。约 500 000 个甲基化过滤的 reads 片段已被组合成 contigs (SAMIs, <http://magi.plantgenomics.iastate.edu>), 而这些 reads 片段提供了基于甲基化估计的基因空间的 1X 覆盖度<sup>[11]</sup>。

### 4 高粱 QTL 定位及关联遗传学

为了实现基因组研究与育种实践间的联系, 高粱连锁图谱被应用于许多性状的定位和标记。种间的群体可以有效地用在与驯化相关的基因的鉴定, 如种子大小、落粒性<sup>[12]</sup>、分蘖和根茎等<sup>[13]</sup>。植株高度和花期也一直很受重视<sup>[14, 15]</sup>。由于高粱以杂交种的推广利用为主, 所以在育性恢复方面的分子遗传控制机制的研究越来越多<sup>[16]</sup>。对于高粱的各种疾病, 主要虫害、寄生杂草等抗性基因也已被标记了<sup>[17, 18]</sup>。与非生物胁迫有关的基因和 QTLs, 如生殖后期耐旱性(持绿性)<sup>[19]</sup>、收获前发芽和耐铝性<sup>[20]</sup>等研究都有较大的进展。由于高粱在生物能源上的利用价值越来越受到重视, 进行了茎秆含糖量 QTL 的定位<sup>[21]</sup>。

对现存种质多样性水平和模式的更好了解有助

于高粱特异基因的分析 and 培育适应环境的特用高粱, 并能更多地实现高粱序列分析的价值。因为具有连锁不平衡的中间模式和自花授粉的交配体系, 所以高粱很适合关联作图方法。在美国国家植物种质系统和 ICRISAT 有广泛的 *ex situ* 高粱种质。美国的科研人员和 GCP 第一子项目组正在进行补充关联遗传学面板的鉴定工作。目前, 在序列的 3.3 Mb 处发现了超过 750 个 SSR 等位基因和 1 402 个 SNP 等位基因, 它们可以在 CSGR(Comparative Grass Genomics Center) 相关数据库中免费获得。对高粱 DNA 序列变异的大量研究表明, 单倍型变异很低(即使核苷酸变异很高也是如此)。在 17 个种质(accessions) 中调查了平均长度 671 bp 的区域, 单倍型的中间数为 3, 众数是 2<sup>[22]</sup>。因此普通的序列变异在少量种质样品中即可获得。

### 5 工作展望

作为禾谷类作物的一个模式基因组, 高粱基因组的测序已经完成, 高密度遗传图谱、表达图谱及基因功能鉴定等研究也取得了显著进展。这些研究成果可为玉米、小麦、大麦等禾谷类作物基因组研究提供极有价值的信息、理论及技术。

此外, 高粱是从含糖的谷类植物中分化出来的, 这一类群中包括甘蔗、约翰逊草和 *Microstegium*。他们要么在基因组大小, 要么在染色体数目上发生了一定的变化。如大多数的 *parasorghums* 的染色体数目都出现了明显的减少, 由祖先的 20 变为 10; 甘蔗(*Saccharum*) 出现了两次以上的染色体加倍, 属于这一类群中的少数; 一部分染色体数目变化的种还有自然发生(*Sorghum halepense*) 和人为介导(*Saccharum* cultivars) 的多倍体化, 这一类群中与高粱具有相同染色体数的种间在基因组大小上有很大的变异(*S. bicolor* and *S. propinquum* versus *nitidum*)<sup>[23]</sup>。因此, 对于高粱基因组的研究, 有助于对甘蔗、约翰逊草、*parasorghums* 等这一特殊类群成员的基因组大小和结构的进化机制、水平、模式进行了解, 有助于揭示高粱基因组演变到目前状态的路径, 也为进一步对甘蔗和其他有重要经济性状的成员的研究打下基础。

对高粱属的研究也为在杂草及其散布的生物学方面提供了新的认识。由地下茎滋生导致的营养体繁衍和由落粒引起的种子传播使约翰逊草[*Sorghum halepense* (L.) Pers,  $2n = 2x = 40$ ] 成为世界上最为有害的杂草之一。它是双色高粱(*Sorghum bicolor*) 和拟高粱(*S. propinquum*) 的种间杂交种, 并遗传了后者的地下茎的特性。约翰逊草产生一种系统素

(Systemin) 是形成根状茎的遗传物质。约翰逊草作为杂草的这个特性对于许多遗传复杂的作物, 如饲草、草皮和营养体作物来讲恰恰是很有利用价值的。因此, 对高粱该方面开展分子生物学方面的研究可以为了解相关的杂草生物学和改良大批的其他饲草、草皮和营养体作物提供机会。

# 参考文献:

- [ 1 ] Farrell A E, Plevin R J, Turner B T, *et al.* Ethanol can contribute to energy and environmental goals[ J ]. Science, 2006, 311(5760): 5069– 5081.
- [ 2 ] Hu F Y, Tao D Y, Sacks E, *et al.* Convergent evolution of perennality in rice and sorghum[ J ]. Proceedings of the National Academy of Sciences of the United States of America, 2003, 100(7): 4050– 4054.
- [ 3 ] Kresovich S, Barbazuk B, Bedell J A, *et al.* Toward sequencing the Sorghum genome. A U. S. National Science Foundation-sponsored workshop report[ J ]. Plant Physiology, 2005, 138(4): 1898– 1902.
- [ 4 ] 卢庆善. 高粱学[ M ]. 北京: 中国农业出版社, 1999: 381– 415.
- [ 5 ] 王 磊, 陈景堂, 张祖新. 主要禾谷类作物比较基因组学研究策略与进展[ J ]. 遗传, 2007, 29(9): 1055– 1060.
- [ 6 ] 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, RFLP and SSR markers[ J ]. Plant Molecular Biology, 2002, 48( 5– 6 ): 483– 499.
- [ 7 ] 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: 367– 386.
- [ 8 ] Peterson D G, Schulze S R, Sciana 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.
- [ 9 ] Anumuganathan K, Earle E. Estimation of nuclear DNA content of plants by flow cytometry[ J ]. Plant Molecular Biology Reporter, 1991, 9 ( 3 ): 208– 218.
- [ 10 ] 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.
- [ 11 ] Bedell J A, Budiman M A, Nunberg A, *et al.* Sorghum genome sequencing by methylation filtration[ J ]. PLoS Biology, 2005, 3(1): 13.

- [ 12 ] Paterson A H, Lin Y R, Li Z, *et al.* Convergent domestication of cereal crops by independent mutations at corresponding genetic loci[ J ]. Science, 1995, 269( 5231 ): 1714– 1718.
- [ 13 ] 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 halpense* ( L. ) Pers[ J ]. Proceedings of the National Academy of Sciences of the United States of America, 1995, 92( 13 ): 6127– 6131.
- [ 14 ] 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.
- [ 15 ] Ulanich P E, Childs K L, Morgan P W, *et al.* Molecular markers linked to Ma( 1 ) in sorghum[ J ]. Plant Physiology, 1996, 111: 709.
- [ 16 ] 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.
- [ 17 ] Mutengwa C S, Tongoono P B, Sithole-Niang I. Genetic studies and a search for molecular markers that are linked to Striga asiatica resistance in sorghum[ J ]. African Journal of Biotechnology, 2005, 4( 12 ): 1355– 1361.
- [ 18 ] Nagaraj N, Reese J C, Tuinstra M R, *et al.* Molecular mapping of sorghum genes expressing tolerance to damage by greenbug (Homoptera: Aphididae) [ J ]. Journal of Economic Entomology, 2005, 98( 2 ): 595– 602.
- [ 19 ] 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.
- [ 20 ] Magalhaes J V, Garvin D F, Wang Y, *et al.* Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae[ J ]. Genetics, 2004, 167(4): 1905– 1914.
- [ 21 ] BIAN Yur long, Yazaki Seiji, Inoue Maiko, *et al.* QTLs for sugar content in sweet sorghum( *Sorghum bicolor* L. Moench ) [ J ]. Agricultural Sciences in China, 2006, 5( 10 ): 736– 744.
- [ 22 ] Hamblin M T, Salas Fernandez M G, Casa A M, *et al.* Equilibrium processes cannot explain high levels of short and medium range linkage disequilibrium in the domesticated grass *Sorghum bicolor* [ J ]. Genetics, 2005, 171( 3 ): 1247– 1256.
- [ 23 ] Price H J, Dillon S L, Hodnett G, *et al.* Genome evolution in the genus *Sorghum*( Poaceae) [ J ]. Annals of Botany, 2005, 95(1): 219– 227.