·综述·

本草基因组计划研究策略

陈士林[1*], 孙永珍[1], 徐 江[1], 罗红梅[1], 孙 超[1], 何 柳[1],
程翔林[2], 张伯礼[3], 肖培根[1]

(1. 中国医学科学院、北京协和医学院药用植物研究所, 北京 100193; 2. 中国生物技术发展中心 北京 100036;
3. 天津中医药大学, 天津 300193)

摘要: 本草基因组计划 (HerbGP) 是针对具有重大经济价值和典型次生代谢途径的药用植物进行的全基因组测序和后基因组学研究的系列计划。本文从物种选择, 全基因组测序、组装和生物信息学分析, 后基因组研究等方面系统阐述了本草基因组计划的研究策略。该计划将推动一批具有典型次生代谢途径的模式植物研究平台的建立, 促进各种“组学”研究方法在药用植物研究领域中的应用, 推动中国传统药学进入生命科学研究前沿领域。本草基因组计划为占领中药基础研究领域的科技制高点提供了难得的机遇, 还将通过对药用植物有效成分生物合成路径的解析和药用植物优良品种的选育对我国天然药物的研发和中药农业的发展产生巨大而深远的影响。

关键词: 本草基因组计划; 药用植物; 全基因组测序; 后基因组学; 次生代谢
中图分类号: R931.5 **文献标识码**: A **文章编号**: 0513-4870 (2010) 07-0807-06

Strategies of the study on Herb Genome Program

CHEN Shi-lin[1*], SUN Yong-zhen[1], XU Jiang[1], LUO Hong-mei[1], SUN Chao[1], HE Liu[1],
CHENG Xiang-lin[2], ZHANG Bo-li[3], XIAO Pei-gen[1]

(*1. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China; 2. China National Center for Biotechnology Development, Beijing 100036, China; 3. Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China*)

Abstract: Herb Genome Program (HerbGP) includes a series of projects on whole genome sequencing (WGS) and post-genomics research of medicinal plants with unique secondary metabolism pathways or / and those of great medical and pharmaceutical importance. In this paper, we systematically discussed the strategy of HerbGP, from species selection, whole-genome sequencing, assembly and bioinformatics analysis, to post-genomics research. HerbGP will push study on Chinese traditional medicines into the front field of life science, by selecting a series of plants with unique secondary metabolism pathways as models and introducing “omics” methods into the research of these medicinal plants. HerbGP will provide great opportunities for China to be the leader in the basic research field of traditional Chinese medicine. HerbGP shall also have significant impacts on the R&D of natural medicines and the development of medicinal farming by analysis of secondary metabolic pathways and selection of cultivars with good agricultural traits.

Key words: Herb Genome Program; medicinal plants; whole genome sequencing; post-genomics; secondary metabolism

收稿日期: 2010-04-01.

*通讯作者 Tel: 86-10-6289970, Fax: 86-10-62896313, E-mail: slchen@implad.ac.cn

本草基因组计划 (Herb Genome Program, HerbGP) 是针对具有重要经济价值的药用植物和代表不同次生代谢途径的模式药用植物开展的基因组层面的系列研究计划, 主要内容包括全基因组序列的测定、组装和生物信息学分析, 及具有典型次生代谢途径的模式药用植物研究平台的建立和抗病抗逆等优良性状遗传机制的阐明等后基因组学研究。

我国药用植物有 10 000 多种, 约占中药材资源总数的 87%[1], 是所有经济植物中最多的一类。同时, 药用植物也是许多化学药物的重要原料, 目前 1/3 以上的临床用药来源于植物提取物或其衍生物。近年来, 药用植物的基因组学研究已经取得了长足的进步。例如, 国内外已经开展了青蒿[2]、丹参[3]、西洋参[4]、甘草[5]等多种药用植物的大规模转录组研究。但是与模式植物和重要农作物相比, 药用植物的基因组学研究还处于起步阶段。基因组序列包含生物的起源、进化、发育、生理以及与遗传性状有关的一切信息, 是从分子水平上全面解析各种生命现象的前提和基础。从 2000 年第一个高等开花植物——拟南芥的全基因组测序完成以来, 仅有水稻、高粱、大豆等 12 种高等植物的全基因组序列公开发表[6−18] (表 1)。第二代高通量测序技术的出现使测序成本大大降低, 测序时间大大缩短 (图 1), 测序物种开始从模式植物、主要农作物向一般经济作物推广[19], 从而为本草基因组计划的实施奠定了坚实的技术基础。目前, 赤芝、紫芝和丹参基因组的主体测序工作已经完成, 人参基因组计划也已经启动 (表 1)。本草基因组计划将使药用植物的生物学研究进入一个崭新的时代——本草基因组时代。本草基因组计划将会极大地推动前沿生命科学技术在药用植物和中药领域的应用, 在较短时间内实现药用植物和中药学研究的跨越式发展, 使其迅速走到生命科学研究的最前沿。该计划的研究成果将为阐明药用植物有效成分的合成和调控奠定基础, 进而促进植物类药物的筛选和生物合成研究, 同时该计划还将加速药用植物优良品种的选育并促进绿色中药农业的科学化和规模化发展。

1　本草基因组计划的全基因组测序、组装和分析策略

我国药用植物资源丰富, 种类繁多, 因此药用植物全基因组计划测序物种的选择应该综合考虑物种的经济价值和科学意义, 并按照基因组从小到大、从简单到复杂的顺序进行测序研究。在测序平台的选择上应以高通量测序平台为主, 以第一代测序技术为辅。图 2 显示了本草基因组计划的全基因组测序策略。

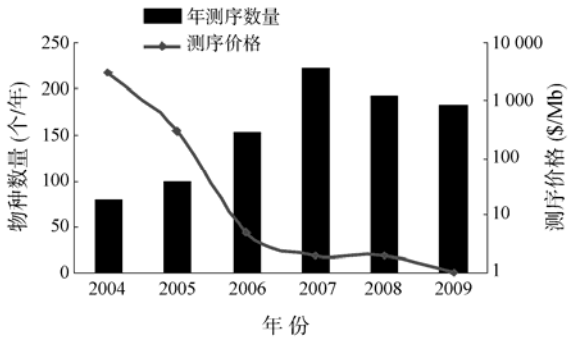

图 1　单位长度测序价格和新增测序物种数随时间变化图

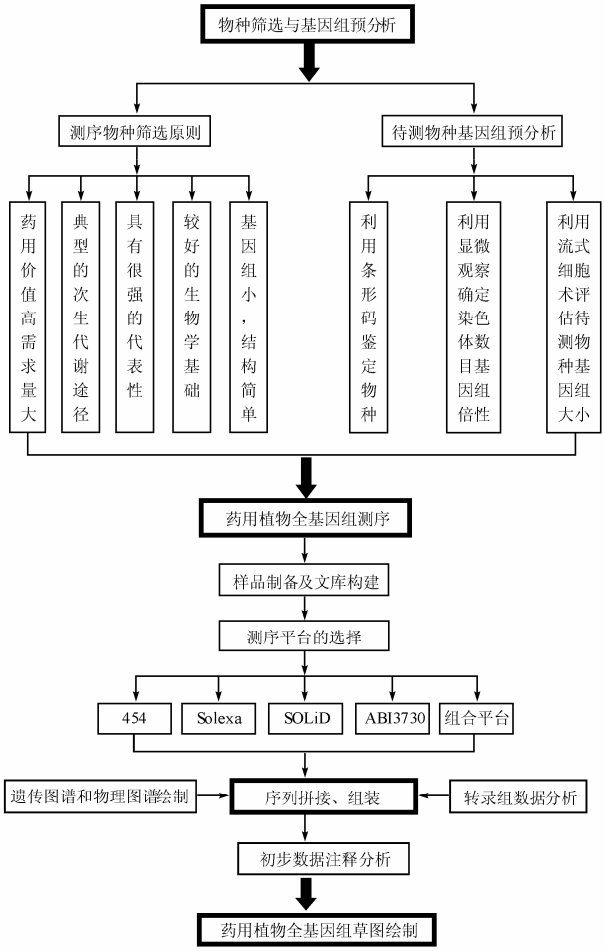

图 2　本草基因组计划全基因组测序策略

1.1　测序物种的筛选原则　本草基因组计划测序物种筛选的基本原则为: ① 名贵大宗中药材的基源植物或重要化学药物的来源植物; ② 药效成分比较清晰, 具有典型的次生代谢途径的代表植物; ③ 含药用植物较多的植物分类单元中的代表植物, 例如豆科的甘草和茄科的枸杞; ④ 具有成为模式植物的潜质, 具有较好的生物学研究基础; ⑤ 优先选择遗传背景清晰、基因组较小且结构相对简单的二倍体植物。

综合考虑以上因素, 紫芝、赤芝、茯苓、丹参、

Table 1 Completed plant genome projects and medicinal plant genome projects in progress. [a]AGI: The Arabidopsis Genome Initiative; CAS: Chinese Academy of Sciences; IRGSP: The International Rice Genome Sequencing Project; IPGC: The International Populus Genome Consortium; MGSC: The Maize Genome Sequencing Consortium; FIPC: The French–Italian Public Consortium; HPGC: Hawaiian Papaya Genome Consortium; Uni-Freiburg: The University of Freiburg; DOE-JGI: The U.S. Department of Energy Joint Genome Institute; CAAS: The Chinese Academy of Agricultural Science; UGA: The University of Georgia; IBI: The International Brachypodium Initiative; ICGC: The International Citrus Genomics Consortium; MBGP: The Multinational Brassica Genome Project; HAGSC: HudsonAlpha Genome Sequencing Center; IBSC: The International Barley Genome Sequencing Consortium; TIGR (JCVI): The Institute of Genome Research (The J. Craig Venter Institute); IMPLAD: Institute of Medicinal Plant Development; PAG: Plant and Animal Genome Conference

Items	Species	Genomesize /Mb	Sequecing methods	Coverage (when published)	Status (when published)	Organizations in charge [a]	Information resourses
Completed plant genome projects	*Arabidopsis thaliana*	125	Sanger	\	Draft	AGI	Nature, 2000, 408 (6814): 796−815
	Oryza sativa.indica	466	Sanger	4.0×	Draft	CAS	Science, 2002, 296 (5565): 79−92
	Oryza sativa.japonica	420	Sanger	6.0×	Draft	IRGSP	Science, 2002, 296 (5565): 92−100
	Populus trichocarpa	485	Sanger	7.5×	Draft	IPGC	Science, 2006, 313 (5793): 1596−1604
	Physcomitrella patens	511	Sanger	9.0×	Draft	Uni-Freiburg & DOE-JGI et al	Science, 2008, 319 (5859): 64−69
	Zea mays	2 300	Sanger	6.0×	Highquality draft	MGSC	Science, 2009, 326 (5956): 1112−1115
	Vitis vinifera	487	Sanger	8.4×	High-quality draft	FIPC	Nature, 2007, 449 (7161): 463−467
	Carica papaya	372	Sanger	3.0×	Draft	HPGC	Nature, 2008, 452 (7190): 991−996
	Sorgum bicolor	730	Sanger	8.5×	High-quality draft	UGA & DOE-JGI et al	Nature, 2009, 457 (7229): 551−556
	Cucumis sativus	243.5	Sanger + Solexa	4.0×+68.3×	Draft	CAAS et al	Nature Genetics, 2009, 41 (12): 1275−1281
	Glycine max	1 100	Sanger	8.0×	Draft	HAGSC & DOE-JGI et al	Nature, 2010, 463 (7278): 178−183
	Brachypodium distachyon	272	Sanger	9.4×	High-quality draft	IBI	Nature, 2010, 463 (7282): 763−768
Medicinal plants genome projects	*Chlorophytum borivilianum*	540	Sanger	\	In progress	Nandan Biomatrix Ltd	NCBI
	Selaginella moellendorfii	212	Sanger	7.0×	In progress	DOE-JGI	PAG XVI, 2008
	Citrus sinensis	380	Sanger + 454	2.1×+30×	In progress	ICGC	PAG XVIII, 2010
	Brassica juncea	\	454	\	In progress	MBGP	PAG XVIII, 2010
	Hordeum vulgare	5 000	454 + Solexa	\	In progress	IBSC	PAG XVII, 2009
	Ricinus communis	400	Sanger	4.0×	In progress	TIGR (JCVI)	NCBI
	Ganoderma lucidum	\	\	\	Assembled	IMPLAD	NCBI
	Ganoderma sinesis	\	\	\	Assembled	IMPLAD	NCBI
	Salvia miltiorrhiza	600	\	\	Assembled	IMPLAD	NCBI
	Panax ginseng	3 200	\	\	In progress	IMPLAD	NCBI
	Panax Notoginseng	\	\	\	In progress	IMPLAD	NCBI

人参、三七等 10 余种药用植物被筛选作为本草基因组计划的第一批测序物种。其中丹参因为基因组小(约 600 Mb)、生长周期短、组织培养和遗传转化体系成熟等原因, 被认为是研究二萜合成的理想模式植物[20]。丹参全基因组的测序完成将会进一步推动丹参作为第一个药用模式植物地位的确定。

1.2 待测物种基因组预分析 由于多数药用植物都缺乏系统的分子遗传学研究, 因此在开展全基因组计划之前进行基因组预分析非常必要。基因组预分析的主要内容包括: ① 利用条形码等技术对满足筛选原则的待测物种进行鉴定[21, 22]; ② 通过观察有丝分裂中期染色体确定待测物种的染色体倍性和条数; ③ 采用流式细胞术[23, 24]或脉冲场电泳技术估测物种的基因组大小, 为测序平台的选择提供参考。

1.3 测序平台的选择 在过去 20 年中, Sanger 双脱氧酶法一直是测序技术的“金标准”。近年来, 第二代高通量测序技术逐步成熟, 新启动的测序计划多以第二代测序平台为主 (表 1)。表 2 显示了不同测序平台的最新的技术比较[25, 26]。

由于药用植物丰富的多样性, 不同物种的基因

组大小和复杂程度可能千差万别, 因此药用植物的全基因组测序可以根据经费预算和基因组预分析结果, 灵活选择不同的测序平台或平台组合 (图 2)。在基因组较小的物种测序计划中可以选择 GS FLX 或 Illumina GA 测序平台[27]。对于复杂的植物大基因组可以选择两种或以上的测序平台进行鸟枪法和双末端 (pair-end) 测序, 同时构建大片段插入文库如 BAC、Fosmid 等方法进行测序, 然后对获得的数据进行组合拼接。目前, 利用 GS FLX 的鸟枪法测序完成基因组的初步组装, 产生 454 contigs, 然后利用 Illumina GA 或 SOLiD 的双末端测序数据确定 454 contigs 之间的顺序和方向, 形成 scaffolds。最后利用 Illumina GA 或 SOLiD 数据填充部分 contigs 之间的空隙, 是一个比较合理和经济的测序策略。

1.4　遗传图谱和物理图谱的绘制　遗传图谱和物理图谱在植物复杂的大基因组组装中具有重要作用。借助于遗传图谱或物理图谱中的分子标记, 可以将测序拼接产生的 scaffolds 按顺序定位到染色体上。

遗传图谱又称连锁图谱, 是指基因或 DNA 标记在染色体上的相对位置和遗传距离。RFLP、SSR、RAPD 和 AFLP 等分子标记都可以用于遗传图谱的构建[2]。由于遗传图谱的构建需要遗传关系明确的亲本和子代株系, 因此其在大多数药用植物中的应用受到限制。

物理图谱描绘 DNA 上可以识别的标记位置和相互之间的距离 (碱基数目)。目前物理图谱的绘制多是基于 BAC 文库, 通过限制性酶切指纹图谱、荧光原位杂交等技术将 BAC 克隆按其在染色体上的顺序排列, 不间断地覆盖到染色体上的一段区域[28]。此外, HAPPY 作图[29]、光学图谱[30]等不依赖于 BAC 文库的方法也可以用于药用植物物理图谱的绘制。

1.5　全基因组的组装及生物信息学分析　随着第二代测序技术的快速发展, 用于短序列拼接的生物信息学软件大量涌现, 常用软件包括用于罗氏 454 数据拼接的 GS Assembler 及第三方软件 Velet、Euler、SOAP 等[31]。

基因组草图组装完成后, 可利用生物信息学方法对基因组进行分析和注释, 为后续功能基因组研究提供丰富的资源。例如, 可以通过 GeneScan[32]、FgeneSH[33]等工具发现和预测基因; 利用 BLAST 同源序列比对或 InterProScan[34]结构域搜索等方法对基因进行注释; 利用 GO 分析对基因进行功能分类[35]; 利用 KEGG 对代谢途径进行分析等[36]。

2　本草基因组计划的后基因组研究策略

后基因组学是基因组测序完成后研究基因组的基因功能、基因之间相互关系和调控机制的学科。本草基因组计划的后基因组研究将根据全基因组序列提供的信息, 充分利用各种“组学”方法, 兼顾正向遗传学和反向遗传学的研究思路, 对药用植物的功能基因进行充分的发掘和研究。根据药用植物的特点, 本草基因组计划的后基因组学研究内容主要集中在模式药用植物研究平台的构建、次生代谢产物合成和调控机制的解析及抗病抗逆等优良性状的遗传机制研究等方面。图 3 显示了本草基因组计划的后基因组研究策略。

2.1　模式药用植物突变体库的建立和基因功能研究　突变体法是利用基因标签技术 (Gene Tagging)[37], 由 T-DNA 或转座子 (transposon) 等已知序列 DNA 片段的插入导致目标基因失活或激活等功能性变化。拟南芥[38]、水稻[39]等重要模式植物均具有大规模的 T-DNA 插入突变体库, 利用这些突变体库发掘了大量生长发育、抗逆性、代谢相关的重要基因。丹参等模式药用植物全基因组序列的测定和大规模突变体库的建立将为药用植物研究提供丰富的资源和材料,

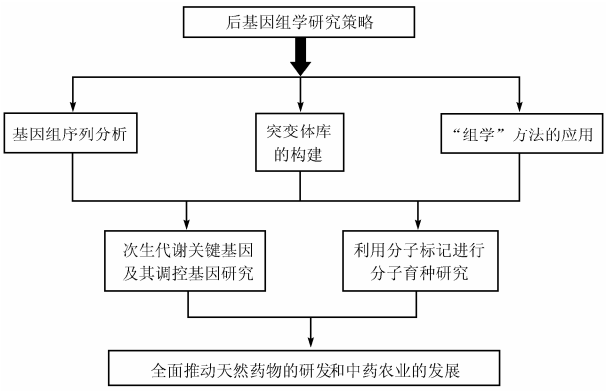

图 3　本草基因组计划后基因组研究策略

Table 2　Comparison of different sequencing platforms

Platform	Read length /bp	Throughput per cycle /Gb	Time expense /d	Sequencing expense /$/Mb
ABI3730	1 000	5.6×10^{-5}	N/A	～500
Roche GS FLX Titanium	400	0.6	0.35	～60
Illumina Hiseq2000	2×100	200	8	<2
ABI/SOLiD 4	2×50	100	12−6	<2

并大大推动药用植物功能基因, 尤其是次生代谢相关基因的发掘进程[40]。目前丹参的突变体库的构建工作正在进行中。

2.2 药用植物有效成分的合成及其调控研究 虽然药用植物有效成分的化学和药理学研究已经具有良好的基础, 但是其生物合成途径和调控方面的研究还很薄弱, 目前该领域的研究主要集中在长春花、青蒿和甘草等少数物种, 例如 Collu 等[41]报道了将候选细胞色素 P450 基因转化长春花的悬浮细胞, 验证了其具有 10-香叶醇羟化酶的催化功能; Seki 等[42]利用昆虫异源体内共表达方法和酵母体外表达检验酶活力的方法对甘草中的三萜甘草酸合成关键酶基因进行了鉴定, 这些研究多采用单基因研究策略。本草基因组计划将会推动转录组学、蛋白组学和代谢组学等“组学”方法在药用植物次生代谢途径和次生代谢调控研究中的应用, 为次生代谢相关基因的“批量化”发掘奠定基础。研究成果将会对次生代谢产物的生物合成及代谢工程、高药效成分种质资源的选育等应用领域产生直接影响。

2.3 药用植物抗病抗逆等优良性状的遗传机制研究及优良品种选育 控制药用植物重要农艺性状的基因, 尤其是与生长发育、抗逆抗病、重要遗传性状及种质性状控制相关的基因是药用植物中一类重要的功能基因, 也是本草基因组计划的重要研究内容。利用基因组注释信息, 发掘优良基因, 运用基因工程的手段打破生殖隔离, 培育药效成分含量高的具有优良农艺性状的新品种, 为药效成分的大量提取和广泛临床应用奠定了基础。全基因组序列的从头测序完成, 也将为转录组分析和基因组重测序研究提供参考序列, 同时第二代和第三代测序技术则为种群内不同个体的转录组和全基因组的重测序研究提供了有力的技术保障。通过对种内或品种间种群个体的重测序可以快速、准确、大规模地发现 SNP、SSR 和 InDel 等遗传分子标记 (genetic molecular markers, GMMS), 从而加速遗传分子标记和优良性状的遗传连锁研究, 快速发现药用植物的表型、生理特征与基因型的关系, 提高育种工作的效率。此外, 突变体库中的一些具有抗逆、抗病、高产等优良性状的突变株系以及转基因植株也是良好的新种质资源。

3 展望

自古以来药用植物及其次生代谢产物都与人类的生存、生活和健康息息相关, 但是人们对于药用植物中次生代谢的途径和调控却知之甚少。与核酸、蛋白等生物大分子相比, 次生代谢产物的种类和合成方式都呈现出了更多的丰富性和复杂性。由于次生代谢途径的多样性, 目前的模式生物无法满足次生代谢研究的需要。本草基因组计划通过选择丹参等具有典型次生代谢途径的植物进行基因组测序, 推动一批模式药用植物研究平台的建立, 为研究复杂多维的次生代谢网络奠定基础, 继核酸、蛋白和糖生物学之后, 推动次生代谢生物学的发展。

随着第二代测序技术的逐步成熟及第三代测序技术的出现, 测序的成本和难度将会进一步大幅下降, 将会有更多的药用植物物种被纳入到本草基因组计划中。随着测序物种的增加, 通过比较基因组学进行次生代谢途径研究将成为可能。同时, 比较基因组学将对物种的起源和进化研究产生重大影响。

本草基因组计划为占领中药基础研究领域的科技制高点提供了难得的机遇, 将推动功能基因组学、蛋白组学和代谢组学等现代生命科学技术在药用植物研究领域中的应用, 将是中药现代化的重要组成部分。其研究成果将为次生代谢产物的生物合成和代谢工程, 及优质高产药用植物品种的选育奠定坚实的基础, 推动中国中药产业的健康科学发展。

References

[1] Chinese Medicinal Corporation. Chinese Traditional Medicine Resource Records (中国中药资源) [M]. Beijing: Science Press, 1995.

[2] Graham IA, Besser K, Blumer S, et al. The genetic map of *Artemisia annua* L. identifies loci affecting yield of the anti-malarial drug artemisinin [J]. Science, 2010, 327: 328−331.

[3] Li Y, Sun C, Luo HM, et al. Transcriptome characterization for *Salvia miltiorrhiza* using 454 GS FLX [J]. Acta Pharm Sin (药学学报), 2010, 45: 524−529.

[4] Sun C, Li Y, Wu Q, et al. De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis [J]. BMC Genomics, 2010, 11: 262.

[5] Li Y, Luo HM, Sun C, et al. EST analysis reveals putative genes involved in glycyrrhizin biosynthesis [J]. BMC Genomics, 2010, 11: 268.

[6] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* [J]. Nature, 2000, 408: 796−815.

[7] International Rice Genome Sequencing Project. The map-based sequence of the rice genome [J]. Nature, 2005, 436: 793−800.

[8] Goff SA, Ricke D, Lan TH, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*) [J]. Science, 2002, 296: 92−100.

[9] Yu J, Hu SN, Wang J, et al. A draft Sequence of the rice genome (*Oryza sativa* L. ssp. *indica*) [J]. Science, 2002, 296: 79−92.

[10] Tuskan GA, Difazio D, Jansson S, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray) [J]. Science, 2006, 313: 1596−1604.

[11] Rensing SA, Lang D, Zimmer AD, et al. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants [J]. Science, 2008, 319: 64−69.

[12] Schnable PS, Ware D, Fulton RS, et al. The B73 maize genome: complexity, diversity, and dynamics [J]. Science, 2009, 326: 1112−1115.

[13] Jaillon O, Aury JM, Noel B, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla [J]. Nature, 2007, 449: 463−467.

[14] Ming R, Hou SB, Feng Y, et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus) [J]. Nature, 2008, 452: 991−996.

[15] Paterson AH, Bowers JE, Bruggmann R, et al. The *Sorghum bicolor* genome and the diversification of grasses [J]. Nature, 2009, 457: 551−556.

[16] Huang SW, Li RQ, Zhang ZH, et al. The genome of the cucumber (*Cucumis sativus* L) [J]. Nat Genet, 2009, 41: 1275−1281.

[17] Schmutz J, Cannon SB, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean [J]. Nature, 2010, 463: 178−183.

[18] The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon* [J]. Nature, 2010, 463: 763−768.

[19] Edwards D, Batley J. Plant genome sequencing: applications for crop improvement [J]. Plant Biotechnol J, 2009, 8: 2−9.

[20] Wang QH, Chen AH, Zhang BL. *Salviae Miltiorrhiza*: a model organism for Chinese Traditional Medicine genomic studies [J]. Acta Chin Med Pharm (中医药学报), 2009, 37: 1−3.

[21] Chen SL, Yao H, Han JP, et al. Validation of the ITS2 Region as a novel DNA barcode for identifying medicinal plant species [J]. PLoS One, 2010, 5: e8613.

[22] Pang XH, Song JY, Zhu YJ, et al. Applying plant DNA barcodes for Rosaceae species identification [J]. Cladistics, 2010, 26: 1−6.

[23] Van Duren M, Morpurgo R, Dolezel J, et al. Induction and verification of autotetraploids and diploid banana (Musa acuminata) by *in vitro* techniques [J]. Euphytica, 1996, 88: 25−34.

[24] Dolezel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry [J]. Nat Protoc, 2007, 2: 2233−2244.

[25] Metzker ML. Sequencing technologies—the next generation [J]. Nat Rev Genet, 2010, 11: 31−46.

[26] Illumina. http://www.illumina.com/systems/hiseq_2000.ilmn [OL].

[27] De Schutter K, Lin YC, Tiels P, et al. Genome sequence of the recombinant protein production host *Pichia pastoris* [J]. Nat biotechnol, 2009, 27: 561−566.

[28] Vu GT, Dear PH, Caligari PD, et al. BAC-HAPPY mapping (BAP mapping): a new and efficient protocol for physical mapping [J]. PLoS One, 2010, 5: e9089.

[29] Hamilton EP, Dear PH, Rowland T, et al. Use of Happy mapping for the higher order assembly of the *Tetrahymena* genome [J]. Genomics, 2006, 88: 443−451.

[30] Latreille P, Norton S, Goldman BS, et al. Optical mapping as a routine tool for bacterial genome sequence finishing [J]. BMC Genomics, 2007, 8: 321.

[31] Imelfort M, Edwards D. De novo sequencing plant genomes using second-generation technologies [J]. Brief bioinform, 2009, 10: 609−618.

[32] Lynn AM, Jain CK, Kosalai K, et al. An automated annotation tool for genomic DNA sequences using GeneScan and BLAST [J]. J Genet, 2001, 80: 9−16.

[33] Salamov AA, Solovyev VV. Ab initio gene finding in *Drosophila* genomic DNA [J]. Genome Res, 2000, 10: 516−522.

[34] Zdobnov EM, Apweiler R. InterProScan-an integration platform for the signature-recognition methods in InterPro [J]. Bioinformatics, 2001, 53: 847−848.

[35] Joslyn CA, Mniszewski SM, Fulmer A, et al. The gene ontology categorizer [J]. Bioinformatics, 2004, 20: 169−177.

[36] Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG [J]. Nucleic Acids Res, 2006, 34: 354−357.

[37] McCallum CM, Comai L, Greene EA, et al. Targeted screening for induced mutations [J]. Nat Biotechnol, 2000, 18: 455−457.

[38] Sessions A, Burke E, Presting G, et al. A high-throughput *Arabidopsis* reverse genetics system [J]. Plant Cell, 2002, 14: 2985−2994.

[39] Sallaud C, Gay C, Larmande P, et al. High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics [J]. Plant J, 2004, 39: 450−464.

[40] van der Fits L, Memelink J. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism [J]. Science, 2000, 289: 295−297.

[41] Collu G, Unver N, Peltenburg-Looman AMG, et al. Geraniol 10-hydroxylase1, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis [J]. FEBS Lett, 2001, 508: 215−220.

[42] Seki H, Ohyama K, Sawai S, et al. Licorice *β*-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin [J]. Proc Natl Acad Sci USA, 2008, 105: 14204−14209.