

本草基因组学

——中药组学的发展与未来

陈士林 等 编著

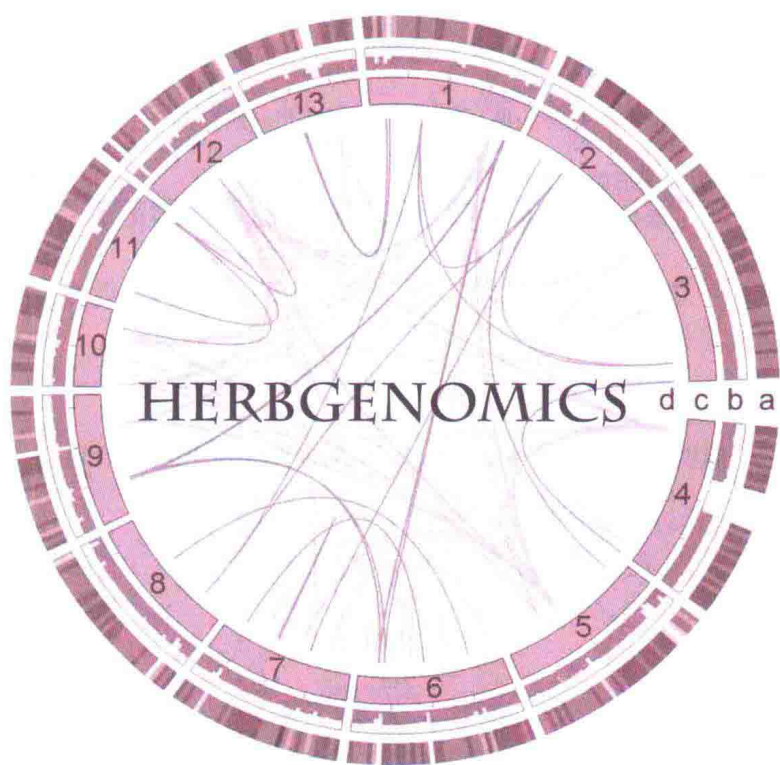


 科学出版社

本草基因组学

——中药组学的发展与未来

陈士林 等 编著



科学出版社

北京

内 容 简 介

本草基因组学作为中草药研究的热点领域,在中药现代化发展中发挥着重要作用。本书作者长期工作在本草基因组学研究一线,结合各自研究工作介绍了本草基因组学的主要研究内容及其应用。本书第一章至第八章为基础篇,包括概述、结构基因组学、转录组学、功能基因组学、蛋白质组学、代谢组学、表观基因组学和宏基因组学等;第九章至第十五章为应用篇,包括药用模式生物、中药合成生物学、基因组辅助育种、DNA 鉴定、中药体内代谢基因组研究、关键实验技术和生物信息学及数据库等。

本书图文并茂,注重科学性、实用性,可为从事本草基因组学教学和研究的、中医药大学的本科生和研究生提供参考和借鉴。

图书在版编目(CIP)数据

本草基因组学:中药组学的发展与未来 / 陈士林等编著. —北京:科学出版社, 2016.12

ISBN 978-7-03-051288-8

I. ①本… II. ①陈… III. ①中草药—基因组—研究 IV. ①R284

中国版本图书馆 CIP 数据核字(2016)第 314831 号

责任编辑:曹丽英 贾冬梅 / 责任校对:刘亚琦

责任印制:肖 兴 / 封面设计:黄华斌

版权所有,违者必究。未经本社许可,数字图书馆不得使用

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京利丰雅高长城印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 12 月第 一 版 开本:889×1194 1/16

2016 年 12 月第一次印刷 印张:20 1/4

字数:598 000

定价:238.00 元

(如有印装质量问题,我社负责调换)

《本草基因组学》编委会

顾 问 肖培根 胡之璧 周宏灏 刘昌孝
张伯礼 吴以岭

主 编 陈士林

副主编 孙 超 徐 江 宋经元

主要编写人员

罗红梅	董林林	段礼新	姚 辉
肖水明	乐 亮	韩建萍	李西文
卢善发	李 滢	王彩霞	徐志超
向 丽	代云桃	廖保生	吴 斌

前言

中医学不仅是中华民族优秀传统文化的重要组成部分和杰出代表，也凝聚着先辈深邃的哲学智慧，反映了中华民族几千年的健康养生理念及其实践经验，是我国人民与疾病长期斗争过程中积累起来的宝贵财富，为中华民族的繁衍生息和人类生命健康做出了巨大贡献。近代以来随着化学和医药工业的发展，从中医学中汲取营养，开发现代工业产品中成药、中药组分或单体成分中药，成为中药现代化的重要组成部分。20世纪70年代中国中医科学院中药研究所屠呦呦教授的研究团队从青蒿中分离纯化出具有良好抗疟疾活性的青蒿素，因此获得2015年诺贝尔生理学或医学奖，成为中药现代化进程中的里程碑，以中药有效成分分离纯化为基础的中医药研究达到巅峰。

科学总结和深入研究中医学对丰富世界医学事业、推进生命科学研究具有重要意义。基因和基因组是现代生命科学研究的中枢，进行中药的生物遗传背景研究，突破现代医药工业面临的资源浪费与枯竭、环境污染与破坏等世界性难题，实现中药绿色、可循环、可持续性开发和利用成为21世纪中药现代化发展的首要问题，回归中药原物种的生物学研究将成为推动中医学实现跨越式发展的原动力。2015年作者在*Science*增刊发表题为“本草基因组学解读传统药物的生物学机制”的文章，首次提出在获取基因组和基因遗传信息的基础上，通过对基因功能的研究和开发来解决中药研究中面临的一系列难题。其中包括建立用于次生代谢产物生物合成及其调控研究的药用模式生物体系，通过合成生物学生产重要的天然药物或新药原料，利用基因组辅助育种培育中草药优良品种，开发用于鉴别中药材混伪品的分子标记，揭示药材道地性的生物学本质，建立本草数据库保护珍贵的药用植物基因资源，以及研究人类基因组遗传多态性和肠道微生物的多样性对中药体内代谢和药效的影响等。

本书对近10年来本草基因组学研究进行了细致的梳理和总结，并对未来的研究思路 and 方向进行了思考和探讨。本书第一章至第八章为基础篇，第九章至第十五章为应用篇。基础篇包括概述、结构基因组学、转录组学、功能基因组学、蛋白质组学、代谢组学、表观基因组学和宏基因组学等，论述了本草基因组学的核心研究内容。应用篇包括药用模式生物、中药合成生物学、基因组辅助育种、DNA鉴定、中药体内代谢基因组研究、关键实验技术和生物信息学及数据库等内容，重点论述本草基因组学的主要应用，介绍本草基因组研究的关键实验技术、生物信息学方法及可查询的相关数据库。希望本书能够为从事本草基因组学研究的科学人员及中医药大学的本科生和研究生提供参考及借鉴。

感谢科学出版社在本书出版过程中付出的辛苦努力。本书是课题组集体努力的结果，是在大量发表论文的基础上编著而成。感谢陈士林、孙超、宋经元、徐江对本书的整体架构、章节编写定位和修改校正付出的努力。感谢撰稿人第一章陈士林、宋经元，第二章徐江、李西文，第三章罗红梅、向丽，第四章孙超、徐志超，第五章乐亮，第六章段礼新、代云桃，第七章罗红梅、卢善发、吴斌，第八章韩建萍、董林林，第九章徐江、陈士林，第十章孙超、王彩霞，第十一章董林林、宋经元，第十二章姚辉、李西文，第十三章肖水明，第十四章徐江、陈士林，以及第十五章李滢、廖保生付出的艰辛和汗水。感谢周宏灏、肖培根、胡之璧、张伯礼、刘昌孝、吴以岭等院士的指导帮助，感谢漆小泉、陈之端、杨秀伟、朱平等著名专家学者对本书的指导和建议，限于我们的经验和能力，书中难免存在疏漏和不足之处，敬请广大读者提出批评和建议，以利再版时改进。

作者

2016年9月

Introduction

Traditional Chinese medicine (TCM) has contributed greatly to improving human health. However, the biological characteristics and molecular mechanisms of TCM in treatment of human diseases remain largely unknown. Genomics plays an important role in modern medicine and biology. Here, we introduce genomics and other related omics to the study of herbs to propose a new discipline, herbgenomics, that aims to uncover the genetic information and regulatory networks of herbs and to clarify their molecular mechanism in the prevention and treatment of human diseases.

Herbgenomics includes herbal structural genomics, functional genomics, transcriptomics, proteomics, metabonomics, epigenomics and metagenomics. Genomic information, together with transcriptomic, proteomic, and metabolomic data, can therefore be used to predict secondary metabolite biosynthetic pathways and their regulation, triggering a revolution in discovery-based research aimed at understanding the genetics and biology of herbs. Herbgenomics provides an effective platform to support chemical and biological analyses of complex herbal products that may contain more than one active component.

Herbgenomics is now being applied to many areas of herb related biological research to help understand the quality of traditional medicines and for molecular herb identification through the establishment of an herbal gene bank. Moreover, functional genomics can contribute to model herb research platforms, geoherbal research, genomics-assisted herb breeding, and herbal synthetic biology, all of which are important for securing the future of the medicinal plants and their active compounds. In addition, herbgenomics will facilitate the elucidation of the targets and mechanism of herbs in disease treatment and provide support for personalized precise medicine. Herbgenomics will accelerate the application of cutting-edge technologies in herbal research and provide an unprecedented opportunity to revolutionize the use and acceptance of traditional herbal medicines.

Author

September, 2016

目 录

基 础 篇

第一章 概述	3
第一节 本草基因组学的产生和发展	3
第二节 本草基因组学研究内容	8
第三节 本草基因组学的实践应用	12
参考文献	14
第二章 结构基因组学	17
第一节 中草药核基因组研究	18
第二节 中草药叶绿体基因组研究	40
第三节 中草药线粒体基因组研究	51
参考文献	64
第三章 转录组学	67
第一节 根及根茎类中草药转录组	69
第二节 皮类中草药转录组	78
第三节 叶及花果类中草药转录组	82
第四节 全草类中草药转录组	86
第五节 真菌类中草药转录组	91
参考文献	95
第四章 功能基因组学——天然产物生物合成途径解析	100
第一节 天然产物生物合成途径解析策略	101
第二节 萜类生物合成途径解析	104
第三节 生物碱类合成途径解析	110
第四节 黄酮类生物合成途径解析	116
参考文献	118
第五章 蛋白质组学	123
第一节 中草药蛋白质组研究	124
第二节 中药作用靶点的蛋白质组研究	128
参考文献	132
第六章 代谢组学	135
第一节 代谢组学研究方法	136
第二节 中草药鉴别和质量控制的代谢组学研究	139
第三节 中草药有效成分合成的代谢组学研究	144
第四节 中草药抗逆代谢组学研究	147
参考文献	148
第七章 表观基因组学	151
第一节 中草药非编码 RNA	152
第二节 中草药 DNA 甲基化与 RNA 编辑	157
参考文献	162
第八章 宏基因组学	165
第一节 宏基因组学研究策略与应用前景	166

第二节 中草药土壤宏基因组	168
第三节 中草药内生菌宏基因组	177
参考文献	179

应 用 篇

第九章 药用模式生物研究	185
第一节 药用模式生物研究概述	186
第二节 药用模式真菌——灵芝	191
第三节 药用模式植物——丹参	196
第四节 其他相关模式生物	202
参考文献	204
第十章 中药合成生物学研究	208
第一节 中药合成生物学体系优化策略	208
第二节 萜类合成生物学研究	213
第三节 生物碱合成生物学研究	217
第四节 黄酮合成生物学研究	221
参考文献	223
第十一章 基因组辅助育种研究	229
第一节 基因组辅助优质高产中草药的选育	230
第二节 基因组辅助中草药抗逆育种	234
参考文献	237
第十二章 DNA 鉴定研究	239
第一节 中药 DNA 条形码鉴定	241
第二节 叶绿体全基因组超级条形码	257
第三节 基于分子杂交等技术的中药鉴定	261
参考文献	268
第十三章 中药体内代谢基因组研究	270
第一节 中药基因组研究	272
第二节 肠道宏基因组研究	275
参考文献	278
第十四章 关键实验技术	281
第一节 中草药基因组关键技术	283
第二节 中药基因功能研究关键技术	290
第三节 蛋白质组学关键技术	295
第四节 代谢组学关键技术	298
第五节 合成生物学关键技术	300
参考文献	303
第十五章 生物信息学及数据库	305
第一节 生物信息学分析方法	306
第二节 基因组相关数据库	309
参考文献	313
索引	315

基础篇

概 述

本草基因组学 (herbgenomics) 是从基因组水平研究中药及其对人体作用的一门前沿学科, 利用组学技术研究中药基原物种的遗传信息及其调控网络, 阐明中药防治人类疾病的分子机制。研究内容涉及结构基因组、转录组、功能基因组、蛋白质组、代谢组、表观基因组、宏基因组、药用模式生物、基因组辅助分子育种、中药合成生物学、DNA 鉴定、中药体内代谢基因组研究和生物信息学及数据库等多个方面。

传统中医药是中华文明的重要组成部分, 也是中国古代科学的瑰宝。近代以来中药研究在形态识别、化学物质基础揭示、药效作用分析、资源调查、人工栽培等多个方面发展迅速, 古老的中药焕发出新的活力。但是由于缺乏中药原植物基因组信息, 中药学和现代生命科学之间缺乏沟通的桥梁, 新兴的前沿生命科学技术很难应用于传统中医药研究, 严重制约了中医药的进一步发展。对于中药道地性形成的遗传机制, 以及道地性和药性的相互关系缺乏深入了解, 已严重影响了中国道地药材的资源保护和新品种选育, 中药道地性形成的遗传基础研究急需加强; 中药基因资源是珍贵的国家战略资源, 国际竞争形势严峻, 韩国、美国、日本等国家已启动许多中药基原物种全基因组研究, 对我国传统中药研究领域造成极大挑战。由于大多数药用植物有效成分含量低, 天然药物分离提取需要消耗大量原料, 对天然资源造成一定破坏, 也使得许多天然药物的生产成本大幅增加。

本草基因组学作为新兴学科, 主要研究内容包括系统发掘中药活性成分合成及优良农艺性状相关基因, 为中药道地品种改良和基因资源保护奠定基础; 建立含有重要活性成分的中药原物种基因组研究体系, 为中药药性研究提供理论基础, 对传统药物学理论研究和应用具有重要意义; 从基因组层面阐释中药道地性的分子基础, 推动中药创新药物研发, 为次生代谢产物的生物合成和代谢工程提供技术支撑; 创新天然药物研发方式, 为优质高产药用植物品种选育奠定坚实基础, 推动中药农业的科学发展, 对揭示天然药物形成的生物学本质具有重要价值。

本草基因组学将会极大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。本草基因组学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。本草基因组学将开辟中药研究和应用的全新领域, 把握历史性机遇, 大大提高我国开发中药资源的能力, 增强中药基础研究实力, 提高中药研究的自主创新能力, 加速中药现代化进程。

第一节 本草基因组学的产生和发展

1. 本草基因组学的产生

从“神农尝百草, 一日而遇七十毒”的传说到现存最早的中药学著作《神农本草经》(又称为《本草经》), 从世界上现存最早的国家药典《新修本草》(《唐本草》)到本草学巨著《本草纲目》, 两千多年来本草学的发展反映中国人民在寻找天然药物、利用天然药物方面积累了丰富的丰富经验。中药学是中国医药学的伟大宝库, 对世界医药学发展做出了巨大贡献。随着现代科学技术的发展, 特别是人类基因组计划 (human

genome project, HGP) 的提出和完成, 对人类疾病的认识和治疗开启了全新的篇章, 在此背景下中药学研究逐渐深入到基因组水平, 从而导致本草基因组学的产生。

1977 年 Sanger 完成首个物种全基因组测序, 即噬菌体 Φ X174 基因组, 大小为 5.836 kb; 人类基因组计划由美国科学家于 1985 年率先提出, 1990 年正式启动, 六国共同参与, 2000 年完成, 是一项规模宏大, 跨国、跨学科的科学探索工程, 其旨在测定组成人类染色体 (指单倍体) 中所包含的 30 亿个碱基对组成的核苷酸序列, 从而绘制人类基因组图谱, 并且辨识其载有的基因及其序列, 达到破译人类遗传信息的最终目的。2000 年破译的拟南芥 (*Arabidopsis thaliana*) 全基因组, 大小为 125Mb, 作为第一个植物全基因组测序在植物科学史上具有里程碑意义。中国药用植物有 11 146 种, 约占中药材资源总数的 87%, 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, 其中最著名的青蒿素来源植物是黄花蒿。

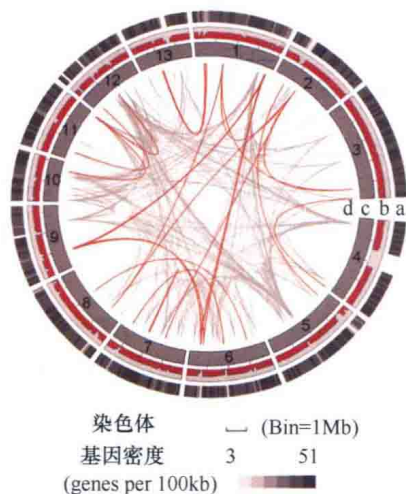


图 1-1 灵芝基因组结构信息 (Chen et al., 2012)

a. 灵芝的 GC 含量; b. 灵芝的基因密度;
c. 灵芝的染色体组成; d. 灵芝的基因组
重复区域

中国学者应用光学图谱和新一代测序技术, 完成染色体水平的灵芝基因组精细图绘制 (图 1-1), 通过基因组解析提出灵芝为首个中药基原的药用模式真菌, 文章发表在《自然通讯》上, 期刊编辑部以特别图片 (featured image) 形式进行了推介 (图 1-2), 认为该论文表明灵芝对于研究传统菌类中药的次生代谢途径及其调控是一个有价值的模式系统。论文发表当天, 《今日美国》(USA Today) 以“揭秘中国‘仙草’基因组”为题报道了该研究成果 (图 1-3), 该论文被 Nature China 网站选为中国最佳研究 (图 1-4)。灵芝基因组图谱的公布为开展灵芝三萜等有效成分的合成研究提供了便利, 随着这些合成途径的逐步解析, 使得通过合成生物学合成灵芝有效成分成为可能。同时, 对灵芝生长发育和抗病抗逆关键基因的发掘及认知, 将推动灵芝的基因组辅助育种研究, 加速灵芝新品种的培育, 并为灵芝的科学栽培和采收提供理论指导。

2009 年, 陈士林团队提出本草基因组计划, 即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。全基因组测序、组装和分析策略: 测序物种的筛选原则, 待测物种基因组预分析, 测序平台的选择, 遗传图谱和物理图谱的绘制, 全基因组的组装及生物信息学分析; 模式药用植物突变体库的建立和基因功能研究; 药用植物有效成分的合成及其调控研究; 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。在此基础上, 详细介绍了本草基因组方法学研究: 全面介绍物种基因组大小、染色体数目测定方法、第二代高通量测序

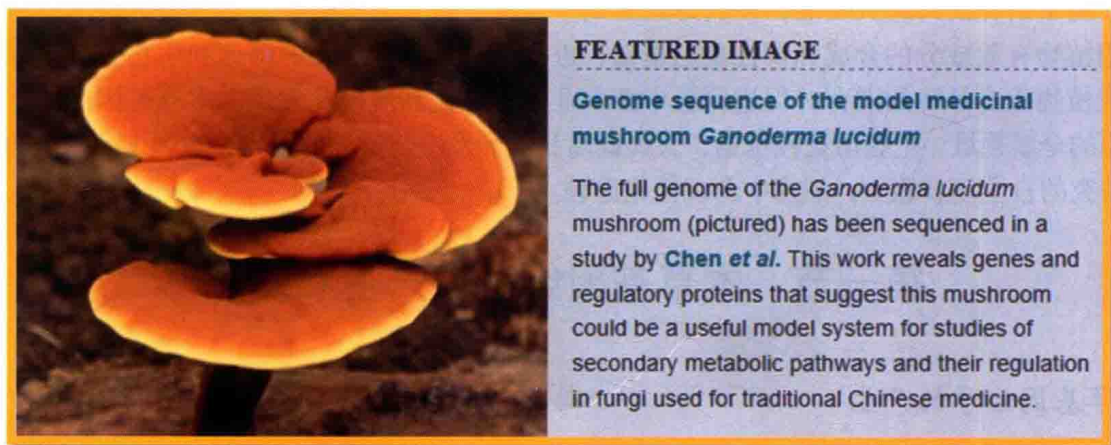


图 1-2 《自然通讯》以特别图片形式发表灵芝基因组研究成果 (Chen et al., 2012)



图 1-3 《今日美国》报道：“揭秘中国‘仙草’基因组”



图 1-4 灵芝基因组论文被 Nature China 网站选为中国最佳研究

方法、全基因组组装和基因组注释方法、基因组比较等生物信息学分析手段、简要阐述重测序在药用植物全基因组研究中的应用方法。由此，本草基因组学逐渐形成和完善，包括中草药结构基因组、功能基因组、基因组辅助分子育种、中草药表观基因组、中草药宏基因组、中草药蛋白质组学、中药合成生物学、中药代谢组、药物基因组学、中草药生物信息学及数据库等内容。基于分子生物学和基因组学的药用植物鉴别是当前研究的活跃领域，用于鉴别的分子生物学和基因组学技术有扩增片段长度多态性 (AFLP)、限制性内切酶片段长度多态性 (RFLP)、随机扩增多态性 DNA (RAPD)、DNA 微阵列技术 (microarray)、DNA 条形码 (barcoding) 等，基于基因组鉴别的分子基础是植物分子系统发育关系反映物种进化关系。

在这些技术中, 药用植物 DNA 条形码鉴定策略及关键技术是最受关注的方向, 中药材 DNA 条形码分子鉴定指导原则已列入《中华人民共和国药典》(简称《中国药典》) 2010 版增补本Ⅲ和《中国药典》2015 版。

2. 本草基因组学的发展

2015 年国际期刊《科学》(*Science*) 增刊详述“本草基因组解读传统药物的生物学机制”, 提出本草基因组学为药用模式生物、道地药材研究、基因组辅助育种、中药合成生物学、DNA 鉴定、基因数据库构建等提供理论基础和技术支撑(图 1-5)。目前, 药用植物基因组学与生物信息学已经进入快速发展阶段, 必将对传统药物学产生巨大影响。国内外已经开展青蒿、丹参、西洋参、甘草等多种药用植物的大规模转录组研究。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, 是从分子水平全面解析各种生命现象的前提和基础。第二代高通量测序技术的飞速发展及第三代单分子测序技术的兴起使测序成本大大降低、测序时间大大缩短, 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、丹参及铁皮石斛等重要药用植物的基因组已被公布, 人参、苦荞、穿心莲、紫苏等中草药基因组图谱也完成绘制。为解析丹参的遗传背景, 陈士林团队联合国内外著名高校和研究机构, 通过联合测序技术完成了丹参基因组图谱的组装, 丹参基因组的完成代表着首个鼠尾草属物种基因组图谱的成功绘制。进化分析显示, 丹参与芝麻的亲缘关系更近, 估计其分化时间约为 6700 万年前(图 1-6)。丹参基因组的发表推动药用模式植物研究体系的确立。本草基因组学将使中草药生物学研究进入一个崭新的时代——本草基因组时代。

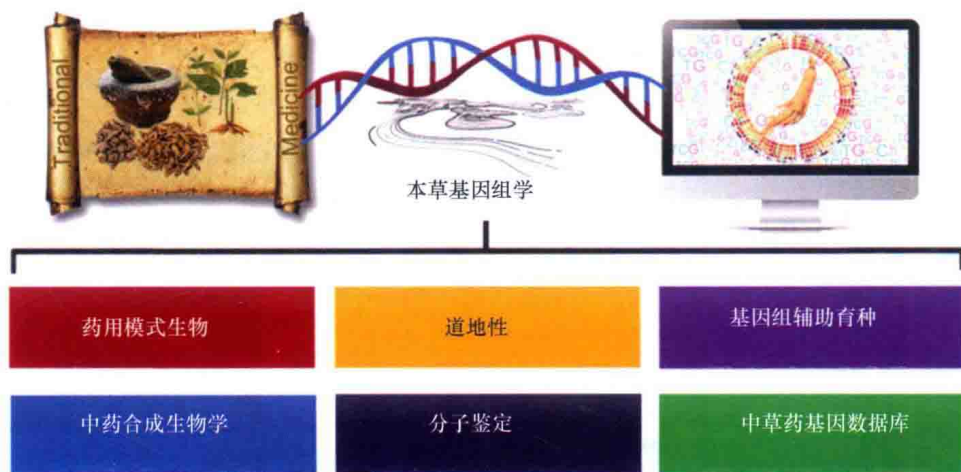
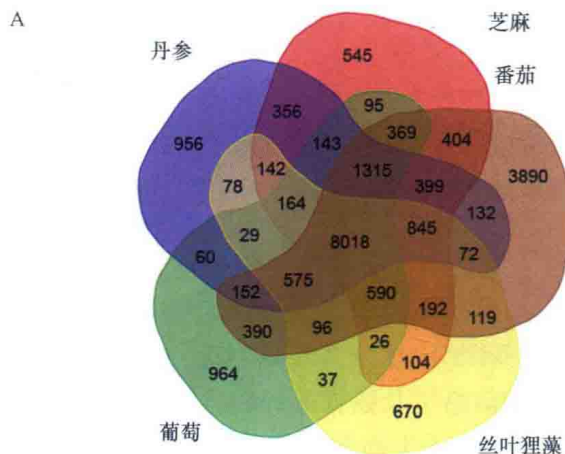


图 1-5 本草基因组学搭建传统药物和现代生命科学研究桥梁



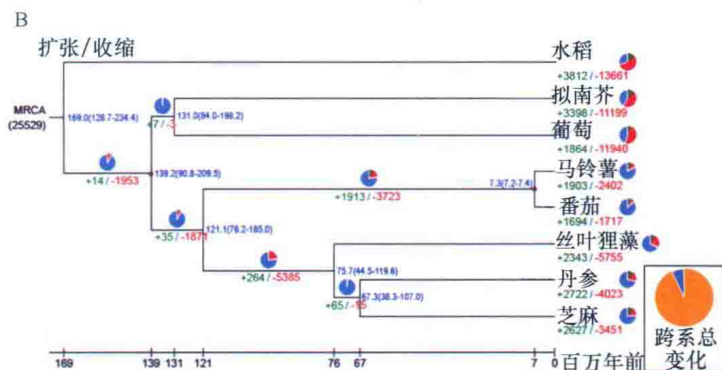


图 1-6 丹参基因组的进化分析 (Xu et al., 2016)

A. 丹参、狸藻、芝麻、番茄和葡萄基因组注释基因家族的进化关系; B. 水稻、拟南芥、葡萄、番茄、马铃薯、狸藻、芝麻和丹参 8 个基因组的 1824 单拷贝基因构建系统进化树, 预测其基因组分化时间

3. 学科内涵和外延

根据本草基因组学产生和发展的过程, 主要从三个方面确定学科的内涵, 即理论体系、实验技术和应用方向 (图 1-7)。本草基因组学形成了高度综合的理论体系, 包括从基因组水平研究本草的九大内容:

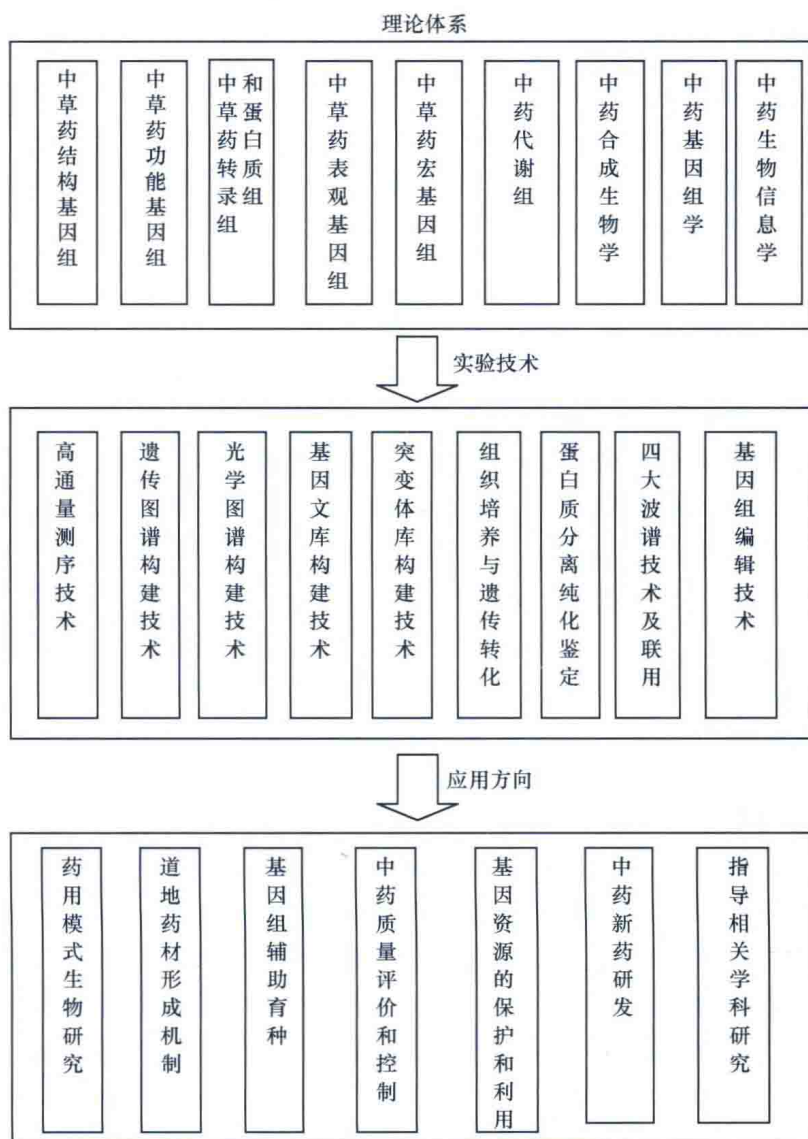


图 1-7 本草基因组学的学科内涵

中草药结构基因组、中草药功能基因组、中草药转录组和蛋白质组、中草药表观基因组、中草药宏基因组、中药代谢组、中药合成生物学、中药基因组学、中药生物信息学。本草基因组学实验主要包括九大技术：高通量测序技术、遗传图谱构建技术、光学图谱构建技术、基因文库构建技术、突变库构建技术、组织培养与遗传转化、蛋白质分离纯化与鉴定、四大波谱技术及联用、基因组编辑技术。基于本草基因组学的理论体系和实验技术，形成了该学科的七大应用方向：药用模式生物研究、阐明道地药材形成机制、基因组辅助育种、中药质量评价和控制、基因资源的保护和利用、中药新药研发、指导相关学科研究。

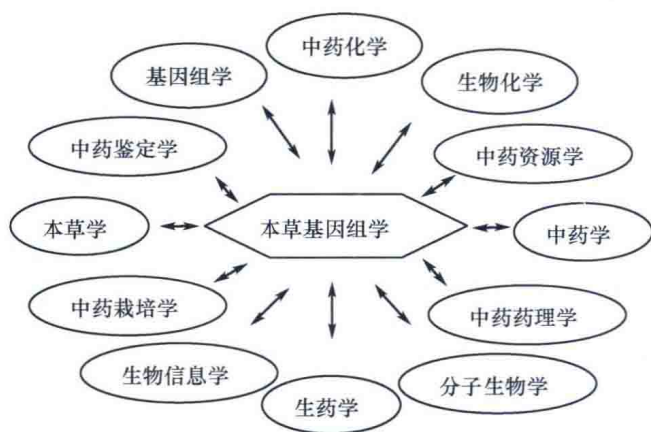


图 1-8 本草基因组学的学科外延

本草基因组学的学科外延与本草学、中药学、基因组学、生物信息学、分子生物学、生物化学、生药学、中药资源学、中药鉴定学、中药栽培学、中药药理学、中药化学等密切相关（图 1-8）。本草学和中药学为本草基因组学奠定了深厚的历史基础和人文基础，为本草基因组学研究对象的确定提供了丰富的候选材料；基因组学和生物信息学为本草基因组学提供了前沿理论和技术支撑，分子生物学、生物化学、中药化学则为本草基因组学提供基础理论和基本实验技术支持，生药学、中药资源学、中药鉴定学、中药栽培学与本草基因组学互相支撑发展，各学科的侧重点不同；

中药药理学、中药化学为本草基因组学的应用提供技术支持。与以上各学科相呼应，本草基因组学促进本草学和中药学从经典走向现代、从传统走向前沿，为中医药更好服务大众健康提供强大知识和技术支撑，扩大了基因组学和生物信息学的研究对象及应用领域，为分子生物学、生物化学、中药化学走向实践应用提供了生动案例，推动生药学、中药资源学、中药鉴定学、中药栽培学从基因组和分子水平开展研究，为中药药理学的深入研究提供理论和技术支持。

第二节 本草基因组学研究内容

本草基因组学借助基因组研究最新成果，开展中草药结构基因组、中草药功能基因组、中草药转录组和蛋白质组、中草药表观基因组、中草药宏基因组、中药合成生物学、中药代谢组、中药基因组学、中草药生物信息学及数据库等理论研究，同时对基因组研究相关实验技术在本草学中的应用与开发进行评价，推动本草生物学本质的揭示，促进遗传资源、化学质量、药物疗效相互关系的认识，以下详细阐述本草基因组学的研究内容。

1. 中草药结构基因组研究

中国药用资源种类繁多，因此药用物种全基因组计划测序物种的选择应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第二代及第三代高通量测序平台为主，以第一代测序技术为辅。图 1-9 显示了适合于当前阶段的中草药基因组研究策略。

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，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。近年来，综合考虑以上因素，紫芝、赤芝、丹参、茯苓、人参、三七、紫苏等十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 600Mb）、生长周期短、组织培养和遗传转化

体系成熟等原因,被认为是研究中药活性成分生物合成理想的模式植物。丹参全基因组测序的完成推动丹参成为第一个药用模式植物。

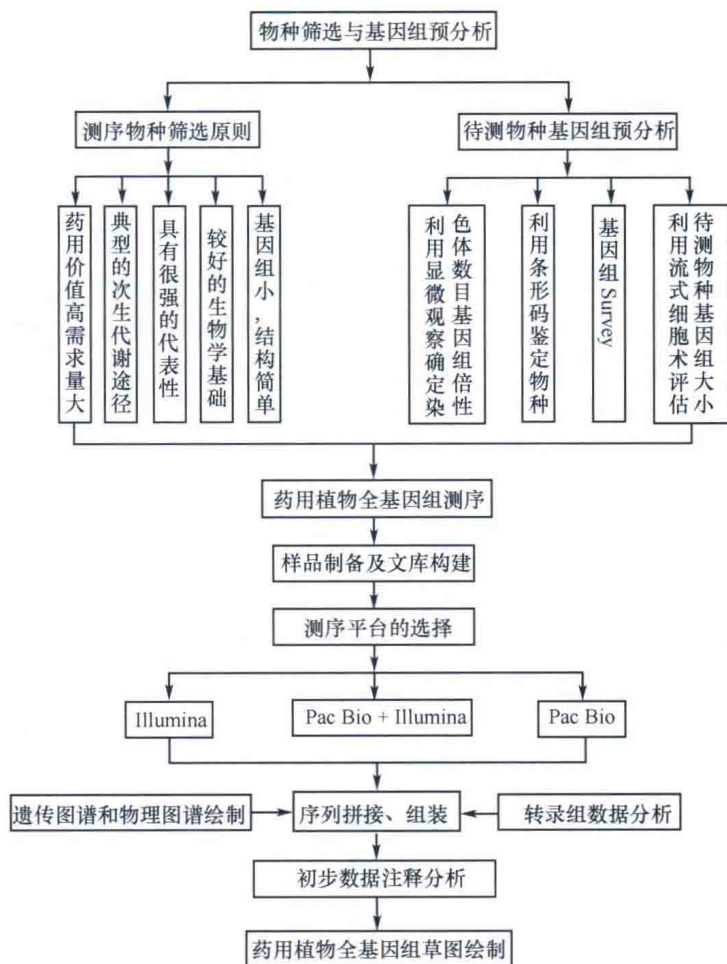


图 1-9 本草基因组计划全基因组测序策略 (陈士林 等, 2010)

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; 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; ④基因组 Survey 测序, 在大规模全基因组深度测序之前, 首先对所选药用植物进行低覆盖度的 Survey 测序, 用来评价其基因组大小、复杂度、重复序列、GC 含量等信息。

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。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 1-9)。在基因组较小的物种测序计划中一般选择 Illumina Hiseq 测序平台。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混合测序拼接, 如构建三代测序大片段文库辅助二代测序组装。

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。遗传图谱又称为连锁图谱, 是指基因或 DNA 标记在染色体上的相对位置和遗传距离。RFLP、简单重复序列 (SSR)、RAPD 和 AFLP 等分子标记都可以用于遗传图谱的构建。由于构建遗传图谱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。物理图谱可以描绘 DNA 上可识