



“十二五”国家重点出版规划

中华民族基因组 多态现象研究

中华民族遗传结构与亲缘关系

丛书总主编 李生斌 梁德生
本卷主编 李生斌

 西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS





国家出版基金项目

NATIONAL PUBLICATION FOUNDATION

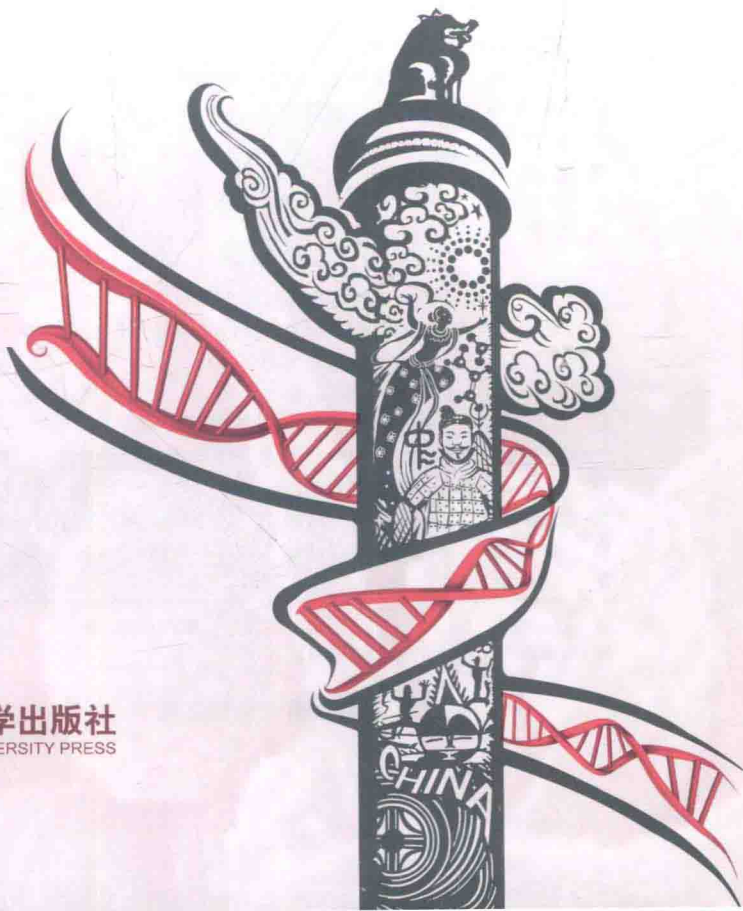
“十二五”国家重点出版规划

中华民族基因组
多态现象研究

中华民族遗传结构与亲缘关系

丛书总主编 李生斌 梁德生

本卷主编 李生斌



西安交通大学出版社

XI'AN JIAOTONG UNIVERSITY PRESS

图书在版编目(CIP)数据

中华民族遗传结构与亲缘关系/李生斌主编. —西安:
西安交通大学出版社, 2015. 10
(中华民族基因组多态现象研究/李生斌, 梁德生总主编)
ISBN 978-7-5605-8044-9

I. ①中… II. ①李… III. ①中华民族-群体-遗传
结构 ②中华民族-亲缘关系-研究 IV. ①Q347

中国版本图书馆CIP数据核字(2015)第254724号

书 名 中华民族遗传结构与亲缘关系
丛书总主编 李生斌 梁德生
本 卷 主 编 李生斌
责 任 编 辑 吴 杰

出 版 发 行 西安交通大学出版社
(西安市兴庆南路10号 邮政编码710049)
网 址 <http://www.xjtupress.com>
电 话 (029)82668357 82667874(发行中心)
(029)82668315(总编办)
传 真 (029)82668280
印 刷 中煤地西安地图制印有限公司

开 本 787 mm×1092 mm 1/16 印张 32.75 字数 817千字
版 次 2016年6月第1版 2016年6月第1次印刷
书 号 ISBN 978-7-5605-8044-9/Q·29
定 价 430.00元

订购热线: (029) 82665248 (029) 82665249
投稿热线: (029) 82665546

版权所有 侵权必究



国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

中华民族基因组多态现象研究



中华民族基因组多态现象研究
编撰委员会

顾 问

杨焕明 夏家辉 贺 林 樊代明
李昌钰 刘 耀 丛 斌

主任委员

李生斌

副主任委员

梁德生 于 军 赖江华 邬玲仟 魏曙光

丛书总主编

李生斌 梁德生

丛书执行主编

李生斌

丛书总审

杨焕明 夏家辉 樊代明 李昌钰
贺 林 刘 耀 丛 斌 闫剑群

丛书编者

(按姓氏笔画排列)

于 军	万立华	马丽霞	马瑞玉	王 剑
王江峰	邓林贝	叶 健	丛 斌	巩五虎
吕卫刚	朱永生	伍新尧	郭玲仟	刘 沁
刘 静	刘 耀	刘梦莹	刘新社	闫剑群
许冰莹	孙 斌	孙宏斌	负克明	严 恺
杜 宏	李 卓	李 波	李 莉	李 晔
李 涛	李帅成	李生斌	李秀萍	李昌钰
李晓忠	李浩贤	杨 爽	杨 璞	杨焕明
吴元明	余 兵	沈亦平	张 月	张 林
张 锐	张玉荣	张幼芳	张杨慧	张秀清
张保华	张洪波	张淑杰	陈 腾	陈荣誉
苟建重	范 歆	林彭思远	欧 拉	罗 莉
罗小梅	罗仕玉	罗静思	周 秦	郑 辉
郑海波	官方霖	郝好英	周 林	胡 亮
胡华莹	胡珺洁	钟秋连	贺 明	袁海明
夏家辉	顾珊智	党永辉	徐 兰	高利生
高树辉	郭 婧	郭佑民	郭若兰	席 惠
通木尔	黄景峰	黄燕茹	梅利斌	曹英西
常家祯	阎春霞	盖 楠	梁德生	彭 洁
彭 莹	董 妍	韩 卫	曾兰兰	曾晓峰
赖 跃	赖江华	谭 虎	曾 博	樊代明
薛晋杰	魏贤达	魏曙光		





中华民族遗传结构与亲缘关系

编撰委员会

主 编

李生斌

副主编

李 波 官方霖 张秀清 李帅成

主 审

于 军

编 者

(按姓氏笔画排列)

王泳钦	田 钊	朱 峰	朱永生	伏东科	刘清波
李 畅	李 波	李帅成	李生斌	张 倩	张秀清
张建波	张洪波	陈 波	官方霖	徐月红	常 辽
		赖江华	魏曙光		





丛书总策划

(按姓氏笔画排列)

王强虎 吴 杰 魏曙光

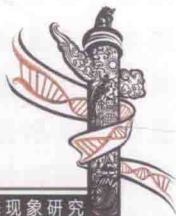
丛书编辑

吴 杰 赵文娟 王银存 田 滢 王 坤



总序

中华民族基因组多态现象研究



个体基因组之间的多态和变异现象，从基因水平上揭示了群体、个体之间差异的本质。基因组多态现象（genomic variation），或称DNA多态性（DNA polymorphisms），是指在一个生物群体基因组中，经常同时存在两种或两种以上的等位基因（allele）或基因型（genotype），且每种类型的变异频率都较高，不能由重复突变来维持。一般认为，基因组DNA序列中某些特定位点的变异频率超过1%的则称为多态性或者多态现象，这些变异频率大于1%的序列或者片段就被称为DNA多态性位点（polymorphic locus）；其余变异频率低于1%的被称为突变，这些序列或者片段就被称为突变位点（mutant locus）。

基因组多态性的本质，就是在生物进化过程中，各种原因引起染色体DNA的核苷酸排列顺序发生了改变，即产生了基因水平上DNA片段大小和DNA序列在个体间的差异，一般发生在基因序列中的非编码区。DNA多态性主要有片段长度多态性和序列多态性两大类，前者指等位基因间片段长度差异，后者指等位基因间的碱基序列差异。



2001年,由美国、英国、德国、日本、法国和中国共同参与的国际人类基因组计划(Human Genome Project, HGP)完成,我们荣幸参加了人类基因组计划的中国1%任务。HGP的完成,推动了创新生物技术的发展,为生命科学揭开了崭新的篇章,产生了巨大的经济效益和社会效益;为国际跨领域合作创造了“共有、共为、共享”精神财富;让科学家首次从一个基因、一个蛋白、一种标记、一种功能的单一研究,转变成为使用基因组科学,全面地、系统地、从分子到整体功能地揭示生命奥秘、探索医疗应用、服务司法实践。

HGP告诉我们,人类基因组包含了24条双链DNA分子(1~22号常染色体DNA与X、Y性染色体DNA),共由大约31.6亿个碱基对组成,基因数目约为3万~3.5万个(不是先前估计的10万个基因),这些编码基因的DNA序列占到人类基因组的2%,大部分非编码基因的序列占到98%(之前人们认为这些非编码基因序列是垃圾DNA)。完整了解全基因组编码基因序列和非编码基因序列的结构变异现象,有助于理解基因的表达调控,细胞的产生、分化,个体发育机制,以及生物的进化;有助于发掘各种疾病的生物标记,例如各种遗传病、肿瘤、出生缺陷、代谢紊乱等的诊断与防治;有助于个体识别、健康预测、个体化医疗、精准医学的新技术创建,例如各种个体基因组分型、亲缘鉴定、种族溯源、系谱分析、游离DNA分型等;有助于了解人与人之间只有0.1%的序列差异,就是这0.1%的序列差异,决定了人与人之间对疾病的易感性、对药物和环境因素的反应性不同。

长期以来,科学家们一直聚焦于人类基因组中2%编码序列的变异与功能,由此开辟了表达谱、外显子组、蛋白组、代谢组、功能组等新兴研究,并在生命、健康、医学、进化、遗传、制药、预防等领域取得了前所未有的巨大成就,引领着自然科学、社会科学领域诸如哲学、数学、化学、物理等基础科学的快速发展。但对于占人类基因组98%的非编码序列的变异与功能却知之甚少。通过国际人类基因组计划(HGP)、国际

千人基因组计划、单倍体型图计划 (HapMap)、人类基因组多样性计划 (HGDP) 和中国人群基因组多态性结构研究, 科学家们开始意识到, 人类基因组存在着多种可遗传的变异方式, 即基因组存在多种形式的个体和种群差异, 这种差异性的揭示, 开辟了人类针对个体特征、群体遗传结构和复杂疾病致病机制研究的新时代, 使目前绘制一张几乎覆盖全人类的基因组遗传变异图谱, 包括所有的在人群中出现频率不低于1%的变异, 以及那些出现频率还不到0.5%的位于基因之内的变异, 构建世界上最大的人类基因组变异的目录成为可能。人类基因组的非编码区蕴藏着每个人的个体特征, 记录着人类共同的历史演变, 同一种遗传标记在不同的种族、民族和地区的人群中其多态性分布存在着差异, 因此有必要对我国不同民族和地区的群体多态性分布进行调查, 以获得详细可靠的群体遗传学资料。这些资料是法医分子遗传学个体识别及亲子鉴定概率计算中不可缺少的基础性科学依据。但遗憾的是, 上述研究计划并未涵盖世界上所有人群, 也无法使我们系统地认识中华民族群体遗传多态性结构特征和变异规律, 因此, 中华民族群体的基因组多态性特征和变异规律的研究只能由国人自己来完成。对中华民族遗传资源的研究、开发与利用, 是一项具有重大意义而又异常艰辛的工作。这项工作可以为阐明中华民族的起源、演化和发展提供积极的启示; 也将为研究遗传因素在疾病的发生、发展过程中所扮演的角色以及其在法医学领域的应用提供极具价值的参考; 同时为我们从DNA分子水平上详细分析中华民族群体基因组多态性结构特征和变异规律提供科学依据。

“中华民族基因组多态现象研究”丛书聚焦非编码序列的变异与功能, 研究这些中立区域的DNA在人类个体识别、人类群体溯源、人类起源进化及疾病药物效应的个体差异, 帮助我们从新的角度学习和理解我们的基因组, 发现和开发大有希望的组学生物标记 (bio-marker) 或优化已知的生物标记及其检测方法, 例如开发新的血液/组织相关的生物标记, 基因/网络/通路相关的生物标记用于疾病检测和个体诊断。



“中华民族基因组多态现象研究”丛书分为5卷，系统介绍了中华民族的人文、地理与历史演变，剖析人文历史与地理环境对群体基因组多态性遗传结构与变异的影响作用；从遗传学（分子人类学）角度阐明中华民族不同群体的遗传结构和变异规律；论述中华民族健康与疾病基因型、单倍型和临床表型的相互关系；介绍了中华民族群体遗传多态性数据在法医学中的应用。

《中华民族遗传结构与亲缘关系》论述了中华民族遗传变异与亲缘关系的系统理论，并采用大量的数据列表和图表，运用基因组学和生物信息学成果，具体、形象地阐明了中华民族的起源、迁徙以及民族之间在遗传特征上的区别和联系，以此勾勒出中华民族遗传结构的总体轮廓。希望中华民族遗传变异与亲缘关系研究可以为民族学、社会学、人类学以及生命科学领域的创新发展提供一定的思路和启示。

《法医基因组学》综合运用基因组学、生物信息学、计算机科学和数学等多方面知识与方法，阐明和理解大量的基因组数据、信息所包含的法医学意义，并应用于解决法医学研究和司法鉴定相关的各种问题。法医基因组学（forensic genomics）研究使得法医DNA分析技术的发展日新月异，获得广泛的应用，并推动人类遗传学、生物医学、动物学、考古学等其他学科的进步。在实际案例中，法医基因组学不仅可以用DNA遗传标记开展个体识别和亲权鉴定，而且可以有效利用全基因组数据。比如lobSTR分析技术，它能够剖析全基因组STRs，为个体识别和个体医疗开辟了新的途径，还能为生物群体进化、重塑生物群体的演绎历史，以及认识人类健康与疾病提供新的视角。

《成瘾基因组学》系统探索了精神活性物质长期反复作用对中脑腹侧被盖区-伏隔核多巴胺神经元功能的重塑作用及分子机制，采用包括基因组学、分子生物学、组织学和行为学的理论与技术，从不同角度来梳理、整理、提炼成瘾的理论研究成果和实践方法。近20年来对于成瘾机制的探索无论是从宏观

到微观，还是从器官组织到分子水平都有了飞速的发展。同时，越来越多的证据提示：精神活性物质成瘾记忆诱导大脑的基因调控机制发生改变，这些数据对于系统理解成瘾记忆的分子基础和致瘾机制、预测预防易感人群以及防治成瘾复吸都具有重要的科学价值。

《基因组拷贝数变异与基因组病》所论述的基因组拷贝数变异与基因组病是临床遗传学的重要内容之一。该书围绕中华民族群体基因组多态性和生物标记，全面系统地论述了基因组拷贝数变异与基因组病、基因组拷贝数变异与临床表型的相互关系。书中的主要内容包括：基因组拷贝数变异、基因组病、遗传诊断与咨询、基因组病的临床表现与诊断标准等。

《人类单基因遗传疾病》针对60余种单基因遗传疾病，系统介绍了疾病的临床表现、遗传学机制、诊断流程和相关实验操作方法，同时对产前诊断、治疗和预后、遗传咨询也有详实的描述。书中还以典型病例的形式再现单基因遗传病患者“就诊—病史采集—临床诊断—基因诊断—基因检测报告解读—遗传咨询”等全过程，使读者身临其境，加深对单基因病的认识。

“中华民族基因组多态现象研究”丛书历时三年的辛苦采编，由中国科学院、西安交通大学、四川大学、中南大学及国外相关机构等的一线学者共同完成，是一次集体智慧的展示。本丛书是站在巨人的肩膀上，对既往人类基因组学研究的成果与结晶进行了一次系统而科学的归纳梳理。我们期盼以人类基因组研究前沿的“盛筵”，以飨读者，在人类不断探索自身的里程碑上留下浓墨重彩的一笔，也对广大读者尤其是相关研究领域的科技工作者们有所裨益。

“中华民族基因组多态现象研究”丛书的问世，要感谢国家出版基金的资助、西安交通大学出版社给予的重视和支持；感谢所有关心和帮助过本丛书的同仁，特别致谢项目实施过程中数以百计的编撰者和编辑，数以千计的实验人员和辅助人员，数以万计的样本贡献者和组织协作者，以及我们的亲人、



好友的精神支持和理解，没有他们的给予，就没有今天的结果。人类基因组计划的精神贡献“共有、共为、共享”已经成为人类科学活动的楷模，成为本丛书写作的动力，对政治、经济、社会、哲学、安全等方面产生越来越重要的作用，这是我们最为推崇的科学精神。

未来，基于基因组结构和序列变化的基因组学研究无疑将成为生物学和医学的核心命题研究。基因组学技术的快速迭代和规模化使大数据挖掘、复杂信息分析等新概念、新技术变为现实，成为催生新思维、新境界和新作为的圣地。从基因组以DNA序列为研究主体到基因组生物学以生物学命题为研究主体，再到以生物谱系如哺乳动物为研究主体，这符合生物学的发展规律，生物医学研究与临床医学实践正朝着“精准化”高速发展。

当然，想要完整阐释中华民族遗传研究的脉络并非易事，尤其是面对浩如烟海的资料和快速更新的知识，限于编撰者的时间和精力，丛书中必有不尽如人意之处，且丛书中提到的一些研究正在进行中，尚未定论，争议在所难免，但这正是本丛书出版的意义。我们认为，对以往研究中的问题进行总结和分折，对正在研究、有争论的问题进行交流和讨论，必将推动本领域的科学发展，这也正是我们希望看到的。

2015年10月31日



《中华民族遗传变异与亲缘关系》全面展现出一幅随着中华民族迁徙史而带来的民族遗传多态性改变的民族基因谱画卷。民族遗传基因的演化过程比史书更加精准地记录了中华民族的历史变迁。

以往生物学家利用部分DNA测序、解剖学或行为学、人类学、语言学、地理学、历史记录的特征均无法构建出清晰的、精准的人类亲缘关系演化树。由于现代人类在早期快速形成物种，扩张时间很短，群体间在单个基因的水平上没有演化出足够多的序列差异，因此仅用少量的DNA序列很难提供足够的信号确定群间的亲缘关系。另外，由于许多群体可以演化出类似的形态或行为，这为利用解剖学和行为学特征进行群体树构建制造了许多障碍。随着基因组测序技术的日渐成熟和测序成本下降，以及构建演化树计算方法和比较基因组学等的发展，使得科学家们能比过去更好地解决这些科学难题。

为了解决现代人类群体分化时间及人类群体之间亲缘关系的问题，本书采用全基因组DNA序列来推断人类亲缘关系树。通过全基因组数据的方法推断出的人类进化树发展史与单纯使用编



码蛋白基因来构建演化树相比存在较大的差异，因此本书的编撰者在构建民族亲缘关系树时还利用了非编码序列和基因间区序列。

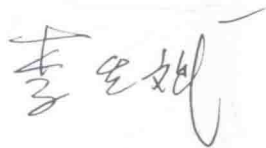
解读民族的遗传变异、构建民族基因组遗传信息图谱，是研究民族亲缘关系的重要工具和手段，是人类遗传研究的重要课题。在近百年来的研究中，科学家们利用遗传标记对全世界范围内的上千个民族进行了遗传变异解读，特别是对其中的几十个民族群体做了全基因组遗传信息图谱的构建。科学家们不但找到了民族共有的、特有的遗传变异，而且推演了民族变迁轨迹及民族之间的交流和融合等的历史事件。中华民族遗传变异与亲缘关系全基因组遗传信息的系统解读为地域性、民族群体性疾病的诊疗研究注入了新的活力。民族特有的遗传信息的解读，将疾病所涉及的遗传易感性和外界诱导因素的综合分析变成现实，使得民族群体性疾病更容易解决。

本书结合中国各民族的历史、文化、语言、体质和地理特征，系统地梳理总结了十几年来针对中国不同区域和民族群体的遗传多态性研究的阶段性成果，包括对大量非编码区系统的研究和应用，及其所积累的大量基础科学数据，为人类学、医学遗传学、法医学等领域创新应用铺平道路。本书的主体内容分为三大部分：一，对中华民族核基因组、线粒体基因组的遗传变异进行了综述；二，介绍了中华民族遗传多态性研究在个人基因组时代和千人基因组时代得到的研究成果，即汉族人、蒙古人的个人基因组图谱和中国多民族群体的基因组遗传多态

性；三，细致描述了利用短串联重复序列构建民族亲缘关系的研究。本书全方位展示了民族的多态性位点及其数据库。

本书全面系统地论述了中华民族遗传结构与亲缘关系的理论，并采用了大量的数据列表和图表，更具体化、形象化地运用基因组学和生物信息学成果来说明中华民族的起源、迁徙以及民族之间在遗传特征上的区别和联系，以此勾勒出中华民族遗传结构的总体轮廓。希望中华民族遗传结构与亲缘关系的研究可以为民族学、社会学、人类学以及生命科学领域的创新发展带来新的启示。

本书的编写主要以西安交通大学法医学院的科研工作为基础，大家分工合作。第1章和第2章由刘清波负责；第3章和第4章由李波负责；第5章由常辽和付东科负责；第6章由官方霖负责；第7章由徐月红负责；第8章由张喆负责。李生斌担任本书主编，负责组织和统稿。国家出版基金的资助，西安交通大学出版社的积极支持，编辑的热情帮助，使本书得以出版，同时本书的撰写对同行的研究成果多有参考，在此一并致谢。



2015年12月31日

C

003

前言