

中国少数民族传统医学课程丛书

中国少数民族遗传学概论

◎ 李树春 / 编著



*Genetics for the
Ethnic Groups of China*

中央民族大学出版社
China Minzu University Press



李树春 河北人，满族，中央民族大学副教授，硕士生导师。2008年毕业于中国协和医科大学生物化学与分子生物学系，获得理学博士学位。现任中央民族大学中国少数民族公共卫生与遗传学研究所副所长，分子遗传学实验室主任。主要研究方向为少数民族遗传学。参与“973”、“863”重大、国家自然科学基金等多项课题。发表SCI、中文核心期刊论文20余篇。

中国少数民族遗传学概论

Zhongguo Shaoshuminzu Yichuanxue Gailun

李树春 / 编著



新疆维吾尔自治区图书馆 XT0-1078432

中央民族大学出版社
China Minzu University Press

图书在版编目 (CIP) 数据

中国少数民族遗传学/李树春编著.

—北京: 中央民族大学出版社, 2010. 8

ISBN 978-7-81108-895-3

I. ①中… II. ①李… III. ①少数民族—人类遗传学

—中国 IV. ①Q987

中国版本图书馆 CIP 数据核字 (2010) 第 179315 号

中国少数民族遗传学概论

作 者 李树春

责任编辑 常远岐

封面设计 布拉格

出 版 者 中央民族大学出版社

北京市海淀区中关村南大街 27 号 邮编: 100081

电话: 68472815(发行部) 传真: 68932751(发行部)

68932218(总编室)

68932447(办公室)

发 行 者 全国各地新华书店

印 刷 者 北京宏伟双华印刷有限公司

开 本 787×1092(毫米) 1/16 印张: 21

字 数 340 千字

印 数 1000 册

版 次 2010 年 8 月第 1 版 2010 年 8 月第 1 次印刷

书 号 ISBN 978-7-81108-895-3

定 价 53.00 元

版权所有 翻印必究

中央民族大学

少数民族传统医学研究中心

“985 工程”学术出版物编审委员会

主任委员：崔 箭

委 员：徐斯凡

朴香兰

庞宗然 黄秀兰

申刚义

前 言

我国是一个多民族国家，在辽阔的中华大地上生活着 56 个民族，其中 55 个少数民族分布于我国 64% 的土地上，他们和汉族一起共同创造了中华民族悠久的历史和文化。我国 56 个民族在民族源流、遗传表型等方面均有其突出特点，在遗传病发病上也有较大不同。这些遗传资源对于人类基因组多样性研究以及群体遗传学研究具有得天独厚的条件。

中国少数民族遗传学是应用遗传学的技术和方法，以少数民族遗传资源为研究对象，研究少数民族遗传结构及其发展规律的科学，主要是用遗传学、数学及其统计学方法研究和探讨少数民族起源、亲缘关系、基因频率和基因型频率及其影响因素。尤其是后基因组时代，特别是国际遗传多样性计划的重要组成部分，对提高少数民族人群遗传素质，促进现代遗传学的发展以及疾病防治具有重要作用。

中国少数民族遗传学主要研究对象和范围包括：

1. 中国少数民族体质特征

人体的某种体质性状的表现型是由特定基因型和特定环境共同作用的产物。在人体体质形成过程中，遗传起着非常重要的作用。因此，少数民族体质特征研究也是中国少数民族遗传学研究的主要内容之一。该项研究一般包括观察指征和测量指征。少数民族体质特征研究，对于探讨民族起源、进化和发展以及遗传与环境的相互作用均具有重要意义。

2. 少数民族起源和亲缘性

针对我国少数民族人群以不同指标通过遗传距离的聚类分析、主成分分析及遗传拓扑学分析，探讨各民族的遗传结构和彼此间的亲缘关系，进而揭示中华民族的源流。

3. 中国少数民族血型分布

血型不仅能够反映各民族之间的亲缘关系，还与很多疾病的易感性相关，因

此研究包括 ABO、Rh、MNSs、Kidd、Duffy、Kell、Diego、P、Xg 以及 Lutheran 血型系统在各少数民族的频率分布，也是中国少数民族遗传学的重要内容之一。

4. 中国少数民族遗传多态性

保护生物遗传多样性已经引起世界各国政府的高度关注，人类遗传多样性是生物多样性的重要组成部分，而中国拥有 56 个民族和一些未识别群体，其遗传资源和遗传多样性资源更加宝贵。开展少数民族遗传多样性研究对人类进化、遗传与环境的相互作用、人类健康具有相当重要的作用。

5. 药物基因组学

药物基因组学 (pharmacogenomics) 是 20 世纪 90 年代末发展起来的基于功能基因组学 (functionalgenomics) 与分子药理学的一门科学。它从基因水平研究基因序列的多态性与药物效应多样性之间的关系，即研究基因本身及其突变体对不同个体药物作用效应差异的影响，以此为平台开发药物，指导合理用药，提高用药的安全性和有效性，避免不良反应，减少药物治疗的费用和风险。少数民族药物基因组学是基于药物反应在少数民族遗传多态性而提出来的，表现为药物代谢酶的多态性、药物受体的多态性和药物靶标的多态性等。这些民族遗传背景上的多态性可能导致许多药物治疗中药效和不良反应的民族差异。

6. 少数民族遗传病和高发病研究

在分层多阶段整群随机抽样调查的基础上，了解我国各少数民族的遗传病和高发病的发病情况。在此基础上，根据不同民族的特异遗传特征，研究遗传病和高发病的发病机制，找到致病基因，为少数民族疾病防治提供理论基础。

本书编写过程中，得到了中央民族大学崔箭教授和徐斯凡教授的鼎力相助，对此深表谢意。同时，少数民族遗传学研究所研究生王桂萍、满序聪、朱美娇、胡华刚和赵莹等同学也对本书书稿做了大量的校对工作，在此一并致谢。本书在编写过程中，查阅了大量专家学者的著作和研究成果，虽然书后附有参考文献书目，但是难免挂一漏万，恳请各位专家和读者谅解。由于编者水平有限，时间仓促，呈现在读者面前的《中国少数民族遗传学概论》肯定存在一些缺点和不足，敬请广大同仁和读者提出批评建议，以便我们进一步修订和完善。

作者

目 录

第一章 中国少数民族简况及其体质特征	(1)
第一节 中国少数民族简况	(1)
第二节 中国少数民族体质特征	(32)
第二章 中国少数民族遗传学的分子基础	(122)
第一节 遗传物质	(122)
第二节 DNA 的双螺旋结构	(125)
第三节 遗传密码	(128)
第四节 基因	(131)
第五节 基因突变	(140)
第三章 中国少数民族遗传学的细胞基础	(148)
第一节 染色质(chromatin)和染色体(chromosome)	(148)
第二节 有丝分裂	(161)
第三节 减数分裂	(166)
第四章 中国少数民族遗传学基本原理	(173)
第一节 孟德尔定律	(173)
第二节 连锁与互换定律	(200)
第三节 Hardy—Weinberg 定律	(209)

中国少数民族遗传学概论

第四节	突变、自然选择、迁移和遗传漂变对基因频率的影响	(218)
第五节	近亲繁殖和杂种优势	(227)
第五章	中国少数民族遗传学的数据处理	(240)
第一节	概率原理	(240)
第二节	卡方检验	(246)
第三节	遗传度、杂合度、多态信息量和吻合度测验	(248)
第四节	主成分分析	(250)
第五节	聚类分析	(258)
第六章	中国少数民族遗传多态性	(282)
第一节	遗传多态性概述	(282)
第二节	ABO 血型系统	(291)
第三节	HLA 系统	(294)
第四节	线粒体 DNA 系统	(298)
第五节	药物代谢酶系统	(301)
第七章	中华民族的源与流	(310)
第一节	东亚人的起源	(310)
第二节	中华民族的源流	(314)
参考文献	(319)

第一章 中国少数民族简况 及其体质特征

第一节 中国少数民族简况

壮 族

壮族是我国少数民族中人口最多的一个民族，2000年人口普查资料显示，壮族总人口为1617.88万人。其中分布在广西壮族自治区的有1415.4万人，主要聚居在南宁、百色、河池、柳州四个地区及南宁市、柳州市；分布在云南省文山壮族苗族自治州及其他各县的有101.09万人；分布在广东省连山壮族瑶族自治县等地的有4.8万人；分布在贵州省黔东南苗族侗族自治州的有1万多人；还有4000多人散居在湖南省江华瑶族自治县。

广西壮族地区是一个山水秀丽、物产富饶的地方。境内山峦起伏，石灰岩分布很广。由于长期的雨水侵蚀，形成了壮丽的石林、岩洞、暗流等奇特的岩溶地貌。壮族有自己的语言，属汉藏语系壮侗语族，20世纪50年代创制了拉丁拼音文字。壮族信仰多神，崇拜祖先，佛教、道教对壮族影响也很大。壮族历史悠久。在秦朝势力进入岭南以前，中国东南沿海地区就居住着“百越”族群，壮族就是由“百越”的一支发展形成的。壮族喜欢依山傍水而居，在青山绿水之间，点缀着一栋栋干栏式木楼，这就是壮族人的传统民居。壮族多用自织的土布作衣料，款式多种多样。节庆的饮食最能反映壮族饮食习惯的特色。三月三吃的五色饭是用红兰草、黄饭花、枫叶、紫蕃藤等植物的汁液分别浸泡糯米，然后煮熟制成的。壮乡素有“歌海”之称，被誉为“歌仙”的刘三姐，就是壮族民间歌手的典型代表。壮锦是壮族的传统纺织工艺品，与湘绣、蜀锦齐名，图案精美，织工精细，享誉海内外。

满 族

满族是人口较多的中国少数民族之一。目前全国满族总人口已超过 1000 万（2000 年人口普查资料显示，总人口为 1068.23 万人），仅次于壮族。满族主要集中分布在辽宁省，现在设有 6 个满族自治县。在河北、北京、吉林、内蒙古、黑龙江、新疆、宁夏、甘肃、福建、山东等省、市、自治区，也有很多散居的满族。

满族有自己的语言和文字。满语属阿尔泰语系满—通古斯语族满语支。满文是 16 世纪末以蒙古字母为基础创制的。满族曾信仰萨满教，崇拜祖先，有祭天、祭祖的习俗。后来，受佛教、道教和儒家思想影响，崇拜对象变得多元化。满族的传统服饰有鲜明的民族特色。“旗头”、“旗鞋”和“旗袍”是满族妇女的代表性服饰装束。满族男子旗袍袖口接有袖头，半圆形如马蹄，俗称“马蹄袖”，平时挽起，放下时可手背御寒。满族人的主食一直到现今，除食用大米和白面之外仍然保留了传统的食品和制作方法，主要有玉米、高粱、小米、大黄米、大豆、小豆等。满族早期多依山傍水而居。房屋多为三间或五间，中间开门，东、西两屋为卧室。卧室内设火炕，取暖烧饭一举两得。满族传统的婚姻仪式较为复杂，大致经过通媒、放定、纳彩、过箱、迎亲、拜堂、拜祖、分大小、回门等程序。满族早期的体育与娱乐是结合在一起的，主要有骑射、摔跤、雪地走（即妇女穿盆底鞋着旗装在雪地中竞走）、踢熊头（类似于足球）、滑冰、跑冰车、跑马城、采珍珠、蹴球等等。

回 族

回族是以中亚细亚各民族为主要族源，并以移民迁徙方式和商业交流活动在中国境内逐渐形成的一个保持伊斯兰文化传统的民族共同体。回族族体的最早来源可追溯到公元 7 世纪（唐宋时期）来华经商留居的穆斯林“蕃客”后裔；回族主要来源是公元 13 世纪蒙古人西征以及元朝时期从波斯、中亚细亚和阿拉伯等地大批签发或自愿东来的各族穆斯林（他们首先成为元朝社会地位较高的色目人之主要组成部分，人口达到数十万），回回民族在形成、发展的过程中，分别融合了所在地区的一些汉、维吾尔、蒙古等民族的部分人口。

回族，自称“回回”或“回民”。回族现有人口已超过 980 万（2000 年），在我国 55 个少数民族中，人口数量仅次于壮族、满族人口。回族聚集区宁夏地处黄土高原与内蒙古高原的过渡地带，物产丰富，有红、黄、蓝、白、黑“宁夏五宝”：红指枸杞，黄指甘草，蓝指贺兰石，白指滩羊皮，黑指煤炭。回族在千百年的发展过程中，受伊斯兰文化、中国传统文化以及生活环境的影响，逐步形成了独具特色的民族风俗习惯。

回族绝大多数成员，把阿拉伯——伊斯兰文化看做是回族文化的“来源”。回族普遍信仰伊斯兰教。回族有严格的饮食方面的禁忌。回族人大都十分熟悉“清真”一词。清真在宗教意义上是指回族虔诚的伊斯兰教信仰及其相关的宗教行为；在个人生活行为意义上是指讲求心净、身净、居处净和饮食净。回族讲求食物的可食性、清洁性及节制性，民间概括为“饮食净”。在动物的可食性方面，受伊斯兰传统文化及中国传统文化“食可养性”观念影响，通过“审物之形象、察物之义理”，一般选择“禽食谷、兽食草”，且貌不丑陋、性不贪婪懒惰、蹄分两瓣能反刍的牛、羊、驼、兔、鹿、獐、鸡、鸭、鹅、雁、雀、鱼、虾等为食，并且除水产品外须“台思米”，即断喉宰杀方可食用。狼虫虎豹熊、驴马骡猪狗、狐猫鼠蟒蛇、鹰鹞鸢鲨鲸等与酒、动物之血属禁食之物。

回族与其他穆斯林民族共同创造和发展了中国清真饮食文化。

苗 族

苗族是我国一个古老的人口众多的民族。2000 年人口普查资料显示，苗族总人口为 894.01 万人，在少数民族中仅次于壮、满、回三个民族，居第四位。历史上苗族分布地域变化很大，根据历史文献记载和苗族口碑资料，苗族先民最先居住于黄河中下游地区，“三苗”时代又迁移至江汉平原，后又因战争等原因，逐渐向南、向西大迁徙，进入西南山区和云贵高原。自明清以后，有一部分苗族移居东南亚各国，近代又从这些地方远徙欧美。

苗语属汉藏语系苗瑶语族。1956 年制定了苗文的拉丁字母拼音文字方案。苗族的宗教信仰主要是自然崇拜和祖先崇拜，也有部分人信仰天主教和基督教。苗族主要从事农业，兼营油茶、油桐、漆树等经济林木业。苗族地区盛产木材，因此，过去的房屋多为木结构。建筑在山区的吊脚楼，一般分为两

层，上层住人，下层圈养牲畜或堆放杂物。苗族服饰式样繁多，色彩艳丽。苗族人民能歌善舞。苗族的“飞歌”高亢嘹亮，极富感染力。舞蹈有芦笙舞、板凳舞、铜鼓舞等，以芦笙舞最为普遍。

维 吾 尔 族

维吾尔族人口有 8399393 人（2000 年），其人口在中国少数民族中仅次于壮族、满族、回族，接近或超过苗族。维吾尔族基本分布在新疆维吾尔自治区，而且尤以喀什、和田和阿克苏地区为主要聚集地。另外，在湖南（桃源县和常德市）、河南和北京等省市也有部分维吾尔族居住。

维吾尔族是一个农业民族，有经营农业的悠久历史，根据新疆的地理环境，发展了绿洲灌溉型农业。新疆的维吾尔族有 11% 的人口居于北疆。北疆的维吾尔族主要经营畜牧业，出产的细毛羊、伊犁马、焉耆马等非常有名。维吾尔族有着悠久的手工业传统，在冶金、雕玉、纺织和皮革加工等方面有一定成就。维吾尔族喜欢吃面粉制作的烤馕、拉面、炒面等。烤馕和抓饭是他们最常见的主食，喜欢吃牛、羊肉。另外还有烤羊肉串、烤包子、薄皮包子等民族风味食品。维吾尔族的传统建筑具有明显的干旱地区建筑特点。房顶多为穹隆顶或平顶，墙壁较厚，拱形门窗，窗口少而且小，天窗较大，用来采光。地毯、花帽、艾德来斯绸、民间印花布和英吉沙小刀等是维吾尔族最盛名的传统工艺制品。维吾尔族是一个能歌善舞的民族。民族乐器有几十种之多，其中都他和热瓦甫是最常用的独奏、合奏乐器。维吾尔族有自己的语言文字。维吾尔语属阿尔泰语系突厥语族。

土 家 族

土家族是一个古老而又年轻的民族，主要居住在我国湘、鄂、渝、黔交界的广大地区。说它古老，早在先秦时期其先民就居住在这里，而且许多地方至今仍保留着土家族特有的风习；说它年轻，是因为新中国成立后的 1957 年才被国务院正式确认为一个单一民族。1957 年 9 月湘西土家族苗族自治州建立时，全国土家族人口有 50 余万人，其中湖南 34 万多人，湖北近 20 万人。党的十一届三中全会后，由于党的民族政策的全面贯彻落实，湘、鄂、渝、黔边的部分群众纷纷要求恢复土家族成分，“还本归原”。经过各级政府

鉴别认定和国家民委和公安部批准，有相当一部分群众恢复了土家族成分。

2000年第五次全国人口普查统计，土家族人口为8028133人。土家族有自己的语言，属汉藏语系藏缅语族中比较接近彝语的一种独立语言，没有本民族文字。土家族居住的湘、鄂、渝、黔比邻地区，以武陵东脉和清江流域为中心，西抵贵州梵净山和乌江，东接彝陵和江汉，北界巫山长江，南控兰澧芷沅，方圆约10万平方公里。值得称道的是，土家族地区有神奇的山水风光，这是得天独厚的旅游资源。土家族地区有不少名胜古迹，如永顺的溪州铜柱、永顺老土司城、咸丰唐崖土司城、五峰土汉疆界碑、长阳武落钟离山、永顺不二门、酉阳桃花源等等。

彝 族

彝族现有人口776.23万。主要分布在云南、四川、贵州省和广西壮族自治区。中国最著名的卫星发射基地西昌，就是凉山彝族自治州州府所在地。彝族有自己的语言文字。彝语属汉藏语系藏缅语族，彝族文字是一种音节文字，形成于13世纪。彝族的宗教信仰主要是自然崇拜与祖先崇拜，云贵地区部分彝族受道教和佛教影响。彝族历史悠久，早在2000多年前，彝族的祖先就生息繁衍在云贵高原和金沙江、澜沧江流域。公元8世纪，以彝族为主体建立了南诏国政权。

彝族的传统民居被称为“土掌房”。这种“土掌房”一般以石块为房基，土坯砌墙，形成平台屋顶。彝族服饰古朴、独特。生活在不同地区的彝族人民的服饰有所不同，大致可分为凉山型、乌蒙山型、红河型等。凉山的彝族成年男子往往在脑后留一绺长发，象征男性尊严不可侵犯，俗称“天菩萨”。彝族的食品主要是由玉米、荞麦的面粉制作的。彝族人喜欢饮酒、吃“坨坨肉”。每年的农历六月二十四日是彝族最隆重的火把节。彝族的婚俗也很独特，大凉山彝族群众中至今仍普遍保留着“抢婚”的遗风。彝族人民能歌善舞，民间流传着各种曲调，如爬山调、迎客调、娶亲调等等。

蒙 古 族

蒙古族是一个富有传奇色彩的民族，对亚欧历史进程产生过巨大影响。全球蒙古族人口约有820万，主要分布在中国、蒙古国（约240万）、俄罗斯

(约 70 万) 三个国家。我国境内的蒙古族总人口目前已接近 600 万，主要居住于内蒙古自治区，全区蒙古族人口约有 380 万，其余多分布于新疆维吾尔自治区巴音郭楞蒙古自治州、博尔塔拉蒙古自治州和布赛尔蒙古自治县，辽宁省喀喇沁左翼蒙古族自治县、阜新蒙古族自治县，吉林省前郭尔罗斯蒙古族自治县，黑龙江省杜尔伯特蒙古族自治县，甘肃省肃北蒙古族自治县，青海省海西蒙古族藏族自治州和河南蒙古族自治县，也有一部分聚居或散居于云南、河北、北京等省市。蒙古族自称“蒙古”(mongol)，其意可解为“永恒之火”；“蒙兀”是“蒙古”一词最早的汉文译名，始见于唐代。蒙古族一般还被称为“马背民族”。蒙古族有自己的语言和文字，蒙古语属阿尔泰语系蒙古语族，分内蒙古、卫拉特、巴尔虎—布里亚特三种方言(按地区划分为中部方言、西部方言和东北部方言)。蒙古文是一种很便于书写的拼音文字，初创于成吉思汗时代。蒙古族文化艺术生活非常丰富。蒙古族素有“音乐民族”之称，其民歌体裁有长、短调之分。蒙古族服饰主要包括首饰、袍子、腰带、靴子四个部分。首饰大致可分为头饰、项饰、胸饰、腰饰、手饰等五大类。蒙古族的饮食非常讲究，品种多样，炒米、奶茶以及奶油、奶皮、奶酪和酸奶子等是蒙古族牧民最有特色的日常食品。勒勒车和配有精美马鞍的骏马是蒙古民族最重要的传统交通工具。热情好客，待人诚恳，是蒙古族人民的传统美德。献哈达、敬鼻烟是他们比较正规的见面礼俗。

藏 族

藏族是我国民族大家庭中的重要一员，全国的藏族总人口目前超过 550 万，主要聚居在西藏自治区(约占一半)，其余分布于青海省的玉树、海南、黄南、海北、果洛藏族自治州和海西蒙古族藏族自治州，甘肃省的甘南藏族自治州和天祝藏族自治县，四川省的甘孜、阿坝藏族羌族自治州及木里藏族自治县，云南省的迪庆藏族自治州等地。

藏族有自己的语言文字。藏语属汉藏语系藏缅语族藏语支。藏族经济以高原畜牧业和高原农业为主，牲畜与农作物主要有绵羊、山羊、牦牛和青稞、小麦等。藏族各地的婚俗不尽相同，但一般都要请活佛打卦求签，选定结婚吉日。而藏族的丧葬习俗则比较特别，以天葬(鸟葬)为主，还有塔葬和水葬等。藏族节日很多，既有生产性、纪念性节日，也有社交、游乐性节日，

另外还有很多宗教性节日。藏族人民能歌善舞，歌舞种类很多，舞蹈的宗教特性较为明显。藏族的作家文学和民间文学，在数量上居中国少数民族前列。此外，藏区还有大量的野生动物资源，其中藏羚羊、野牦牛、野驴、盘羊等系青藏高原特有的珍稀物种，属国家保护动物，白唇鹿为中国特有，是世界珍稀动物之一。在能源方面，藏族地区主要有水能、地热能和太阳能等。西藏的地热蕴藏量居全国第一位，其自然风光和宗教与民族文化旅游资源更是得天独厚。

布 依 族

布依族是我国民族大家庭中一个历史悠久的民族，主要分布在贵州、云南、四川等省，总人口 297.15 万人（2000 年），其中以贵州的布依族人口最多，占全国布依族人口的 97%。

布依族地区属我国南方隆起于四川盆地和广西丘陵之间的一个亚热带岩溶化高原，环境独特，自然资源丰富。布依族地区属中、北亚热带湿润气候区，气候基本特点是热量充足，雨量充沛，雨热同季，终年温暖湿润，“冬无严寒，夏无酷暑”。经济作物主要有棉花、油菜、甘蔗、烤烟、茶叶、土靛、土麻、黄果（橙）、柚子、芭蕉、香蕉、木耳、板栗等等。河、湖产中鱼类也很多，都匀、贵定、平塘、龙里、花溪出产的娃娃鱼（大鲵）是稀有的国家保护动物。布依族地区地下矿藏丰富，尤以煤矿、金矿、锑矿、磷矿著名。布依族地区风光秀丽、景色宜人，已成为中外友人和旅游者考察和观光的胜地。

侗 族

侗族是中华民族大家庭中具有悠久历史的一个民族，主要分布在黔、湘、桂、鄂四省（区）毗邻地方。2000 年人口普查资料显示，侗族总人口为 296.03 万人。其中，贵州省侗族人口为 1400344 人，占全国侗族总人口的 55.7%；湖南省有 749026 人，广西壮族自治区有 286900 人，湖北省有 63912 人，散居全国其他地方的有 13832 人。

侗族主要从事农业，兼营林业。侗语属汉藏语系壮侗语族。1958 年制定了拉丁字母拼音文字方案。侗族的宗教信仰多为自然崇拜和多神信仰。侗族村寨依山傍水。侗寨的房屋一般是用杉木建造的木楼，多为二至三层，在河边或山坡上的

房屋大多为吊脚楼。侗族服饰朴素，色调以青、蓝、白、紫为主。男子多穿对襟短衣，或右衽无领短衣，包大头巾。女子上穿大襟无领无扣上衣，下穿百褶裙或长裤，束腰带、裹腿，包头帕或戴银冠及各种银饰等。侗族的主食是大米，山区则多食糯米。侗族人还喜食酸味，有荤酸、素酸、煮酸、腌酸等，故有“侗不离酸”的俗语。油茶是侗族人民喜爱的饮料，也是待客佳品。侗族传统节日有侗年、吃新节、祭牛神节等。吃新节时，早稻成熟，家家用新米做饭，供奉祖先，并唱侗歌、侗戏，举行斗牛活动。侗族有“饭养身、歌养心”的谚语。

瑶

族

瑶族，现有人口 263.74 万。瑶族主要分布在广西、湖南、云南、广东、贵州、江西等省区。瑶族以其历史悠久、迁徙频繁和文化独特而为世人所瞩目。瑶族是一个山居民族，大部分散居在海拔 1000 米以上的高山和密林之中，少部分居住石山或半石山地区和丘陵、河谷地带。瑶族分布的区域较大，东起广东南雄，西至云南勐腊，南达广西防城，北至湖南辰溪，都是瑶族人民长期活动的地区。五岭、十万大山、都阳山、雪峰山、罗霄山、六韶山、哀牢山等山脉横亘其间，山峦起伏，千溪万涧，纵横交错。大多数的瑶族集散在两广与湖南接壤的五岭南北地区，形成“南岭无山不有瑶”的分布局面。

瑶族地区山清水秀，物产丰富。各种自然资源尤其是森林资源、水力资源、矿产资源、土特产资源和旅游资源十分丰富。瑶族有自己的语言，属汉藏语系苗瑶语族。各地语言差别较大，分别使用勉、布努、拉珈三种语言。瑶族地区通用汉文或壮文，没有本民族文字。瑶族的宗教信仰因地而异，有自然崇拜、祖先崇拜，也有图腾崇拜；还有些地区信仰道教。瑶族房屋建筑风格多样，其中最具有特色的是木楼。瑶族服装常饰以五色丝线的绣花，古有“瑶好五色衣裳”之称。瑶族人以大米、红薯、芋头为主食，喜喝甜酒。凡有宾客临门，主人先敬上一杯甜酒。瑶族的民间文学十分丰富。《密洛陀》是著名的创世史诗，《盘王歌》展现了民族历史和文化生活的广阔画面，汉文手抄本《评皇券牒》则记述了民族起源、迁徙、生产等内容，均具有重要历史价值。

朝 鲜 族

中国朝鲜族人口目前接近 200 万（2000 年人口普查资料显示，总人口为

192.38万人),主要分布在东北三省。吉林省朝鲜族占全国朝鲜族总人口的60%以上,而吉林省朝鲜族约70%又聚居在延边朝鲜族自治州。黑龙江省朝鲜族人口接近50万,辽宁省也有23万以上。东北三省朝鲜族人口约占全国朝鲜族总人口的97%。还有一些朝鲜族居住于内蒙古、北京、河北等省、市、自治区。中国朝鲜族居住地区现有1个自治州和1个自治县。

在全国56个民族中,计划生育工作进行得最为自觉、效果最为明显的就是朝鲜族。朝鲜族没有全民族性的统一的宗教。他们的先人中曾有儒教、道教、佛教信徒,18世纪末以后又曾有基督教、天主教信徒。朝鲜族是能歌善舞的民族,延边朝鲜族自治州以“歌舞之乡”著称于世。著名的民间舞蹈有农乐舞、长鼓舞、扇舞、顶水舞等。朝鲜族主要从事农业,以善种水稻闻名。新中国成立以前,中国朝鲜族人口的90%左右都是农民,以耕种水稻为主,手工业多为半农半工的小手工业,私营工商业更少。朝鲜族人民热爱体育运动。跳板和荡秋千是朝鲜族妇女喜爱的传统体育运动,摔跤和踢足球则是男人们擅长的运动。延边素有“足球之乡”的美称。中国朝鲜族有自己的语言和文字。朝鲜族很重视教育,1949年成立了中国第一所少数民族大学——延边大学,并自筹资金成立了百余所小学。

白 族

白族,总人口为1858063人(2000年)。白族绝大部分居住在云南省大理白族自治州,少部分散居在云南省丽江、碧江、保山、南华、元江、昆明、安宁等地;四川省西昌、贵州省毕节、湖南省桑植等地也有少数白族散居。白族聚居的大理地区地处云贵高原的西北部,大部分属西南峡谷区,为横断山脉的南端,西高东低,地形、地貌、气候特色截然不同,大部分地区属低纬度高原季风气候。

白族自称“白子”、“白尼”、“白伙”,汉语意为“白人”;他称有“民家”、“那马”、“勒墨”。在贵州威宁的白族,在新中国成立前,因有九姓被称为“九姓族”。白族先民,两汉史籍称为“昆明”;三国两晋时称为“叟”、“爨”;唐宋时称“白蛮”、“河蛮”、“下方夷”;元明时称为“白人”;明清以后称“民家”等。新中国成立后,根据民族意愿,统称白族。其民族来源在学术界有土著说、汉人迁来说、哀牢九隆族说、多种族融合说、氏羌族源说、

西爨白蛮说等。白族共同体的形成是在大理国时期。

哈 尼 族

哈尼族现有人口 143.97 万人（2000 年），主要分布于云南省红河哈尼族彝族自治州的红河、元阳、绿春、金平四县，思茅地区的墨江哈尼族自治县，普洱哈尼族彝族自治县、镇沅彝族哈尼族自治县、景东县、江城哈尼族彝族自治县、澜沧县，西双版纳傣族自治州的景洪、勐腊、勐海县，以及玉溪地区元江哈尼族彝族自治县、新平县等地。以上这些地区居住着哈尼族 95% 以上的人口，其余部分散居于建水、石屏、思茅、孟连、景谷、玉溪、峨山、易门、通海、双柏、晋宁、禄劝等县市。哈尼族与汉、白、彝、傣、拉祜、苗、瑶、回、壮、布朗等民族交错而居。在哈尼族居住地区栖息着孔雀、鹦鹉、竹鸡、白鹇、虎、豺、熊、鹿、飞鼠、风猴等珍禽异兽。不少地方被列为国家级自然保护区。

哈尼族地区农作物主要有稻谷、玉米、豆类、荞麦等。经济作物主要有茶叶、咖啡、可可、柠檬、橡胶、竹麻、甘蔗、烤烟、蓝靛等。哈尼族有自己的语言，哈尼语属汉藏语系藏缅语族彝语支，分为哈雅、豪白、碧卡三个方言。哈尼族宗教信仰主要是多神崇拜和祖先崇拜。哈尼族村寨大多建在半山腰，一般都是有血缘关系的几个家庭聚居而成。哈尼族的传统住房是由土基墙、竹木架和茅草顶形成的楼房。哈尼族人喜欢用藏青色土布做衣服。哈尼族擅种梯田，元江两岸到处都是层层叠叠的梯田，有的高达数百级，从河谷一直延伸到山顶，像彩色天梯伸向天边，十分壮观。

哈 萨 克 族

中国哈萨克族总人口为 125.05 万人（2000 年）。哈萨克族主要分布在新疆维吾尔自治区伊犁哈萨克自治州及新疆巴里坤哈萨克自治县、木垒哈萨克自治县，少数分布于甘肃阿克塞哈萨克族自治县和青海海西蒙古族藏族自治州。在历史上，哈萨克族绝大多数人过着逐水草而居的游牧生活，因而其服饰带有较浓郁的草原畜牧生活的特点。其饮食习惯也与赖以生存的畜牧业紧密相关，以肉、奶、面等食物为主，而奶制品在哈萨克族饮食中的分量很大。奶制品的主要有鲜奶、酸奶、奶皮子、奶豆腐、奶疙瘩、酥油、酥酪、奶糕

等。其中马奶酒是哈萨克牧民在各种聚会、盛宴中必不可少的饮品。

哈萨克族及其先民饲养的伊犁马和巴里坤马闻名全国。由于长期从事游牧畜牧业，需要不断移动，马就成了乘骑的主要交通工具。哈萨克族信仰伊斯兰教。哈萨克族历史上曾信仰原始宗教，相信万物有灵，因而天、地、日、月、星宿、水火都成为崇拜的对象。哈萨克族民间乐曲十分丰富，古朴雄浑，优美动听。哈萨克族最善于用舞蹈来表达自己的思想感情和抒发自己内心的喜怒哀乐。哈萨克族人民喜爱的传统娱乐活动主要是叼羊、赛马、姑娘追和摔跤等，其中男女骑手娱乐活动“姑娘追”更富有情趣和浪漫色彩。哈萨克的传统节日是“诺吾鲁孜”节，类似于春节，在阴历春分日。哈萨克族游牧的草原土特产、名贵药材都很丰富。地下资源也相当丰富，有金、铜、铁、铅、锌等。如阿尔泰山即为（蒙古语）“金山”之意，富蕴县的铍、锂、钽、铌、铯、铷、钫等 82 种稀有金属矿藏储量居全国之冠，该县可可托海三号矿被称为“天然矿物博物馆”。

黎 族

黎族人民居住在海南岛的中南部，黎族居住区处于北回归线以南，具有得天独厚的气候条件，光照充足，长夏无冬，四季常春。

黎族总人口为 124.78 万人（2000 年），其中 84.7% 分布在实行民族区域自治的琼中、保亭、白沙、乐东、昌江、陵水 6 个自治县和享受民族自治地方政策待遇的三亚、通什、东方 3 市，少数散居于万宁、儋州、琼海 3 市和屯昌县等。黎族源于古代“百越”的一支，早在秦汉以前就从两广地区陆续来到海南岛定居。“黎”这一专有族称始于唐末，沿用至今。黎族有自己的语言，属汉藏语系壮侗语族，无文字，通用汉文。黎族的宗教信仰以祖先崇拜为主，杂以自然崇拜，有些地区还保留着图腾崇拜的痕迹。因状似倒扣船只而得名的船形屋，是黎族的传统住房。黎族称其为“布隆亭竿”，意为“竹架棚房子”。“布隆亭竿”有辅地形、高架形和向金字形过渡三种。黎族服装以棉麻为原料，黎锦是黎族人最为著名的纺织工艺品。嚼槟榔是黎族人的一种嗜好，槟榔是待客、订婚不可缺少的佳品。在节日或劳动间歇，黎族人喜欢跳“竹竿舞”，通常是在庭院或打谷场上跳这种舞。

傣 族

傣族人口为 1158989 人（2000 年），主要聚居在云南南部的西双版纳傣族自治州、西部的德宏傣族景颇族自治州、西南部的孟连傣族拉祜族佤族自治县和耿马傣族佤族自治县，这些地区的傣族人口占全省傣族总人口的 60% 以上。其余散居在澜沧江以东的景谷、景东、普洱等地，澜沧江以西的风庆、云县、临沧、双江、澜沧等县；红河沿岸的新平、元江、元阳、红河、金平等县，金沙江畔的华坪、大姚、禄劝等地及与越南交界的马关县也有分布。从地理位置上看，90% 以上的傣族都分布在北纬 25 度以南，海拔 1300 米以内的低纬度低海拔的河谷平坝地区。这个区域江河纵横，山川秀丽，怒江、澜沧江、把边江、元江由北向南流经傣族的地区，为热带、亚热带气候带类型，气候条件优越，年平均气温在 21℃ 左右，终年无雪，轻霜期 30 余天，全年无四季之分，只有明显的干季和湿季。雨量充沛，年均降雨量 1260—1965 毫米，80% 的雨量集中在 5—10 月，年蒸发量 1409—1815 毫米，雾天有 100 余天，年日照时数为 1790—2340 小时。在德宏州一带，由于低纬度高原地带太阳入射的高度较大，空气透明度好，是全国的光照高质区。

傣族自称“傣仂”、“傣雅”，民族识别时定名傣族。傣语属汉藏语系壮侗语族，傣文为拼音式文字。傣族信仰小乘佛教。傣族是较早种植水稻的民族之一，大部分傣族地区盛产稻谷，许多地区的稻谷可二熟到三熟，西双版纳有“滇南谷仓”之称，德宏州的“芒市谷子遮放米”也享誉省内外。

畲 族

畲族是中国的一个古老民族。“畲”字有“垦田而居者”之意。畲族自称“山哈”，畲语意为“山客”。1956 年 12 月畲族作为单一民族被确认。2000 年人口普查资料显示，畲族总人口为 70.96 万人，主要分布于福建、浙江、广东、江西、安徽、贵州、湖南等省。畲族的主要姓氏为“蓝”、“雷”、“钟”，人数较多的姓氏还有“李”、“吴”等。畲族有自己的语言，部分畲语属汉藏语系苗瑶语族苗语支，接近瑶族布努语的炯奈话。畲族的宗教信仰主要是祖先崇拜。

几十年前，畲族住房多为茅草房和木结构瓦房，一般都有厅堂和左右厢

房。盖房往往是先立柱、上梁、盖瓦，然后再筑墙和整修室内。畚族妇女多穿大襟小袖衫，衣领、袖口和右襟多镶花边。每年农历三月三是畚族的传统节日，各家照例要蒸乌米饭，用来聚餐、赠友、祭祀祖先。祖图，又称“盘瓠图”，是畚族信仰的主要标志之一。二月二“会亲节”是畚族传统节日之一。畚族地区多属亚热带湿润性季风气候，农产品以稻谷、玉米、豆类、烟叶、土豆为主，盛产林木及毛竹。畚族地区还盛产茶叶、油桐、香菇及名贵药材，浙江景宁的“金奖惠明茶”颇受人们的喜爱。此外，畚族地区还有丰富的动植物及矿产资源。

傈 傈 族

傈傈族是我国云南世居民族之一，2000年人口普查资料显示，傈傈族总人口为63.49万人。傈傈族生活于我国东起川滇边界的西昌、木里、盐边等县，西至云南西北部的广阔地域，云南怒江、澜沧江、金沙江三江并流地区是其分布的中心区。怒江傈傈族自治州的维西、泸水、福贡、贡山、兰坪四县是其主要聚居区。

傈傈族有自己的语言，属汉藏语系藏缅语族。1957年创制了以拉丁字母为基础的新文字。

过去傈傈族信奉多神，崇拜自然，还有图腾崇拜的残余。19世纪中叶后，基督教和天主教传入傈傈族地区。傈傈族有自己独特的历法——自然历，他们把一年分成开花月、鸟叫月、烧山火月、采集月、收获月、煮酒月、狩猎月、过年月和盖房月等十个月。

傈傈族的粮食是玉米、荞麦。傈傈族人认为动物苦胆是一种良药，也是一种调味品，因此总把胆汁掺入酒中饮用，在煮肉时则把苦胆放入锅中一块煮，用来去腥开胃。傈傈族少女到了十三四岁，家人要为她举行穿裙礼的仪式，以此承认其已步入成年期，获得了社交权。傈傈族的主要节日有澡堂会、收获节、过年节等。

仡 佬 族

距今2100多年前，在今天中国的贵州、云南一带，生活着一个古老的农耕民族——僚，今天生活在这里的仡佬族就是他们的后裔之一。仡佬族主要居住于贵

贵州省境内，人口有 57.94 万人，占全国仡佬族总人口的 70% 以上。大体而言，黔北的务川、道真两个仡佬族苗族自治县为仡佬族的主要聚居区域。仡佬语属汉藏语系，有的学者主张把仡佬语群看成是汉藏语系中一个独立的语族。仡佬族的宗教信仰以信奉道教和自然崇拜为主，有的也信奉佛教。

根据文献记载，古代仡佬族的住房多为干栏式建筑，贵州北部地区的仡佬族至今仍保留着传统的住宅样式。仡佬族人生活的地区，大多数在云贵高原向四川盆地过渡的斜坡地带，那里地形复杂，气候潮湿；仡佬族人的经济生活以农业为主，山地种植玉米，平地种植水稻，因此，这两种作物也是仡佬族人最主要的粮食。仡佬族人习惯把玉米粉放在蒸笼里蒸熟，叫作玉米干饭，是仡佬族人一年四季餐桌上的主食，如果在节日里或是有远客临门，他们就在玉米面里加上相等的白米蒸熟，称为“混合饭”。

东 乡 族

东乡族人口超过 50 万（2000 年人口普查资料显示，总人口为 51.38 万人）。东乡族主要聚居在甘肃省临夏回族自治州东乡族自治县以及积石山保安族东乡族撒拉族自治县，一部分东乡族散聚在甘肃临夏县、和政县、广河县、康乐县、临夏市、兰州市、定西地区和甘南藏族自治州等地，还有一小部分散居在新疆维吾尔自治区、青海省和宁夏回族自治区等省区。东乡族语言属阿尔泰语系蒙古语族，东乡族的词汇中，汉语借词较多，也有不少突厥语、阿拉伯和波斯语借词。东乡族有本民族的语言而没有文字，现在使用的是汉文，绝大部分群众懂汉语。

东乡族信仰伊斯兰教。东乡族的节日与其他信仰伊斯兰教的民族相同，主要有三大宗教节日，即尔德节、古尔邦节和圣纪节。

拉 祜 族

拉祜族自称拉祜、拉祜纳、拉祜西和拉祜普。他称有“锅锉”、“果葱”、“苦聪”、“黄古宗”、“傥黑”、“黄傥黑”、“缅甸”、“目舍”等。2000 年人口普查资料显示，拉祜族总人口为 45.37 万人，其人口在中国 56 个民族中排列在第 23 位，分布在 24 个省、市、自治区，其中云南省的拉祜族占总人口的 99.2%。拉祜语属汉藏语系藏缅语族。1957 年在原有文字基础上制定了统一

标准的拼音文字。

拉祜族信仰大乘佛教和万物有灵的原始宗教，少部分人信仰基督教和天主教。拉祜族以黑为美。拉祜族习惯于用竹筒烧饭菜，用竹筒烧出的饭菜不仅保持了原料的原汁原味，又有青竹的清香，十分可口。拉祜族有丰富的口头文学，诗歌中有一种叫做“陀普科”（谜语），深为群众所喜爱。“扩扎”是拉祜族最重要的节日，一般在农历正月，择吉日举行。“扩扎”共九天，分三段。

水 族

2000年人口普查资料显示，水族总人口为40.69万人，其中93.22%居住在贵州省境内，广西南丹、宜山、融水、环江、都安、来宾、河池等县（市）共有水族12799人，占水族总人口的3.7%，云南省富源县古敢水族乡有水族7688人，占总人口的2.2%。三都水族自治县是全国唯一的水族自治县，贵州的水族主要聚居于该县，其余分布在与三都毗邻的荔波、独山、都匀、丹寨、雷山、榕江、从江等县（市），少数散居于黎平、凯里、麻江、福泉等地。三都水族自治县境内气候类型属中亚热带湿润季风气候，夏长冬短，春秋分明，年平均气温18℃，无霜期为320天，年均降雨量1350毫米，雨热同季，农作物可一年两熟到三熟。地下矿藏有汞、锑、金、铝、锌、铜、铁、硫黄、重晶石、金刚石、煤等，地表水年径流量为34亿立方米，水利资源理论储能为15万千瓦。

水族有自己的语言，属汉藏语系壮侗语族。水族原有一种古老的文字，称为“水书”，造字方法有象形、会意、谐音和假借，通用单词100多个，现在则通用汉文。水族信仰多神，崇拜自然物。相传水族的先民是岭南“百越”之一骆越人的一支。水族的名称最早出现在中国明朝末年的文献中。水族传统民居为干栏式建筑，多用杉木和松木建造，屋顶用杉树皮或瓦片覆盖。水族服装多为青、蓝色。男子穿大襟长衫，青布包头；妇女穿蓝色大襟上衣，青布长裤，衣裤都镶有花边，系青绿色花腰带，节日穿裙子，戴各式耳环、项圈、手镯等银制饰品。水族有许多传统节日，最隆重的当推“端节”。端节主要活动为祭祀和赛马。“除夕”与“初一”相连的两顿饭忌荤食素，但不忌鱼虾。水族妇女泡制的“鱼包韭菜”是端节酒席上必备的佳肴。相传此菜原用于治疗，如今已失

去了用来治病的意义，保留了祝愿大家身体健康、生活平安的美好心愿。

佤 族

佤族是居住在我国西南边疆的一个少数民族。2000年人口普查资料显示，佤族总人口为39.66万人，主要分布在云南省西盟、沧源、孟连、耿马、澜沧、双江、镇康、永德以及西双版纳傣族自治州和腾冲、昌宁、景东、普洱等县。其主要居住区位于澜沧江与怒江之间，怒山山脉南段的舒展地带。

佤语属南亚语系孟高棉语族佤德昂语支，大致可分为佤、勒佤和布饶三个方言。阿佤山土壤有砖红壤、赤红壤、红壤、黄壤、黄棕壤、棕壤、亚高山灌丛草甸土、紫色土、水稻土等49个土类，12个亚类，28个土属，44个土种。佤族居住地区动植物资源非常丰富，有野生亚洲象、孟加拉虎、金钱豹、长臂猿、灰叶猴、蜂猴、羚羊、蟒蛇、眼镜蛇、巨蜥、孔雀、犀鸟、翡胸鹦鹉、白鹇、太阳鸟等上百种鸟兽。阿佤山区地下矿藏也极为丰富，这里的银矿在明末清初就已开采，其中著名的茂隆银矿，是石屏汉人吴尚贤从内地带来一批汉族民众与当地佤族联合开采的，在当时不仅促进国家和民族地区的经济发展，也促进了佤族人民与内地各族人民的友好交往。

纳 西 族

纳西族是一个历史悠久、文化灿烂的民族，现有人口30.88万人（2000年），主要分布在祖国西南部的云南、四川以及西藏三省区交界处的金沙江上游与横断山脉地区，即云南省的丽江、宁蒗、永胜、中甸、德钦、鹤庆、剑川、兰坪、华坪、贡山，四川省的盐边、盐源、木里、巴塘、攀枝花，以及西藏自治区的芒康、察隅等县市。其中，云南省的丽江纳西族自治县聚居着三分之二以上的纳西族人口，是纳西族最集中的聚居区。纳西语属汉藏语系藏缅语族。早在公元7世纪，纳西族人民创造了象形表意文字“东巴文”和音乐文字“哥巴文”，但均不通用。被称为“活着的象形文字”的东巴文是目前世界上唯一保留完整的象形文字。由于同汉族的密切关系，纳西族通用汉文。

纳西族信仰东巴教和藏传佛教。在漫长的历史发展与民族交往中，纳西族创造了自己灿烂的历史文化，融诗歌、音乐、舞蹈为一体是纳西族民间文艺的一大特色。“喂热热”、“阿丽丽”、“京盘”、“纳西族古典舞蹈”等一批具

有浓郁民族特色的歌舞在国内外享有盛名。纳西族的建筑、雕刻和绘画融合纳西族、汉族与藏族三个民族的传统风格，具有浓郁的地方特色。丽江纳西族的房屋建筑多为“三方一照壁”、“四合五天井”格局的瓦房，即由较高的正房和两侧略低的偏房，以及正房对面的一座照壁构成，安静宽敞，自成单元，布局合理。纳西族的“三叠水”（又称雪山宴）是招待贵宾的方式，因使用大碗、小碗、盘子三种不同高度的餐具而得名。纳西族有许多传统节日，如尝新节、朝山节、三月会、火把节等，最具特色的是“棒棒会”。纳西族地区名胜古迹很多，有直插云天的玉龙雪山、号称“长江第一弯”的石鼓、大研镇的黑龙潭、丽江的虎跳峡。丽江古城大研镇内街巷幽深、清泉密布、家家流水、户户垂柳，因其优美的环境、保存完好的古城风貌及深厚的文化底蕴，于1997年12月3日被联合国列为“世界文化遗产”。

羌 族

羌族，现有人口30.61万人（2000年）。羌族主要分布在四川省北川羌族自治县和阿坝藏族羌族自治州的茂县、汶川、理县、黑水等地。羌语属汉藏语系藏缅语族，分南北两大方言。羌族没有本民族文字，通用汉文。羌族信仰原始宗教，也有一部分羌族人信仰藏传佛教。羌族历史可追溯到殷商时代，早在3000多年前，殷代甲骨文中就有关于羌人的记载。今天的羌族是古代一部分羌人的后裔。

羌族人居住在独具特色的羌寨中。羌寨中的碉房是羌族的石砌民居，一般高达三四层。底层圈养牲畜，中层住人，上层储藏粮食，屋顶为平台，既可以用来摊晒粮食，又可以作为老人休息、妇女针黹、孩子游戏的场所。今天的羌族人，大多还穿着传统的民族服装。羌族男子的服装一般是蓝布长衫，外套羊皮褂子，包青色头帕。羌族妇女的服饰比较鲜艳，她们多穿蓝色或绿色的花边长衫，腰系绣花围裙和飘带，戴黑色头帕。但在茂汶地区黑虎乡一带，妇女们一律都戴白色头帕。

玉米是羌族地区最主要的粮食作物，小麦、青稞、土豆、荞麦、油菜等农作物的种植也较为普遍。羌族人有把猪肉吊在房梁上风干做成“猪膘”的习惯。陈年的猪膘，肉色嫩黄、晶莹剔透，吃起来油而不腻、十分可口。在喜庆而隆重的节日里，羌族人还要端出自家用玉米或青稞酿造的酒与众人共饮。

羌族地区资源丰富，经济林木有花椒、核桃、茶树、漆树等，还盛产虫草、贝母、鹿茸、麝香等名贵药材。

土 族

土族是我国人口较少的民族之一，人口有 24.12 万人。主要分布在青海省互助土族自治县、民和回族土族自治县、大通回族土族自治县以及黄南藏族自治州的同仁县和乐都县，部分散居于海北藏族自治州的门源县和海西蒙古族藏族自治州等地，还有 2 万多人分布于甘肃省天祝藏族自治县、肃南裕固族自治县、兰州市永登县、临夏回族自治州积石山保安族东乡族撒拉族自治县和甘南藏族自治州卓尼县等地区。其中青海省境内的土族约占全国土族总人口的 85%。

土族有自己的语言，土语属阿尔泰语系蒙古语族，分互助、民和、同仁三个方言区。土族基本上全民信仰藏传佛教。土族每户人家几乎都有一个独立的庭院，靠山向阳，多为土木结构。土族服饰比较独特，土族男女的上衣大都配有绣花高领，妇女服饰以五彩花袖衫最具代表性。土族粮食以青稞、小麦、土豆为主，至今仍保留着牧业时期的痕迹，如喜喝奶茶、吃手把肉和酥油炒面等。忌食马、骡、驴、狗、猫等动物肉。土族人喜欢饮酒，并形成了土族特有的酒文化。热情好客是土族传统风尚，迎送客人敬三杯酒就是这种风尚最突出的表现。土族婚俗别具一格，具有鲜明的民族特点。土族的婚俗大致要经过请媒、定亲、送礼、聚亲、送亲、结婚仪式、谢宴等程序。土族家庭对生儿育女非常重视。孩子满月举行极为隆重的仪式，按土族的习俗，男孩子提前一天满月，女孩子则必须满一个月。土族的丧葬习俗比较独特，分火葬、土葬、天葬和水葬四种方式。土族忌讳有人在牲畜圈棚附近解手。土族人忌讳当客人的面吵架或打骂孩子。春节和“纳顿”节是土族最热烈、最隆重的两大节日。土族是一个能歌善舞的民族。“家曲”与“野曲”是流传于土家地区的民间说唱艺术。

仡 佬 族

在我国广西西北部，有一条风景如画的绿色长廊，传说那里是凤凰的故乡，在这里居住着古代骆越民族后裔中的一支——仡佬族。仡佬族，现有人

口 20.74 万人，主要聚居在广西罗城仫佬族自治县。

仫佬族语言属汉藏语系壮侗语族，多数人通汉语和壮语，没有本民族文字，普遍使用汉族文字。仫佬族的宗教信仰主要是多神崇拜。仫佬族的社会生产以农业为主，善种水稻，采煤是仫佬族社会生产的一个组成部分。仫佬族居住比较集中，同宗同姓的人居住在同一村寨，在同一村寨里，同姓但不共祖的仫佬族人一般也分开居住。聚族而居是仫佬族的一个重要民族特点。

仫佬族民居，多为砖瓦顶的矮楼建筑，最突出的特点是以地炉取暖做饭。仫佬族人粮食为大米、玉米和薯类，喜食酸辣食品和糯米饭。民族风味食品有白馍、水圆、粽子、牛舌粑，红、黄、黑三色糯米饭和鸭肉蘸酱是逢年过节必不可少的食品。仫佬族崇尚青色，多着青色服饰，其服饰素雅大方，一般只在衣袖、裤脚上有小条的刺绣点缀。仫佬族的节日文化丰富多彩，有三月三的婆王节（又称花婆节）、四月八的牛神节、五月初五的端午节、八月十五的中秋节（也叫走坡节），最具特色的是三年一大庆、一年一小庆的“依饭节”。

锡 伯 族

锡伯族现有人口 18.88 万人，主要分布在辽宁、吉林、黑龙江、新疆等地。新疆察布查尔锡伯自治县是锡伯族最大的聚居区。锡伯族有自己的语言和文字。锡伯语属于阿尔泰语系满—通古斯语族，锡伯文是 1947 年在满文的基础上略加改变而成的。锡伯族人基本通晓汉语，新疆地区的锡伯族人有的还兼通维吾尔语和哈萨克语。

锡伯族的宗教信仰早期属原始宗教中的自然崇拜，后来信仰萨满教和藏传佛教。锡伯族原为游牧民族，弓箭在他们的生活中占有重要地位。锡伯族人，特别是迁移到新疆的锡伯族人，善于农业生产。锡伯族人在日常生活中喜欢面食，如有尊贵客人来访，则要杀羊款待，做“全羊席”（锡伯族人称之为“莫尔雪克”）。锡伯族的主要节日有春节、清明节、抹黑节、“杜因拜扎坤”节、端午节、中秋节等。

柯 尔 克 孜 族

我国柯尔克孜族人口为 160823 人（2000 年），主要分布在新疆维吾尔自治区的克孜勒苏柯尔克孜自治州，部分散布在伊犁、塔城、阿克苏和喀什等

地区。此外，在黑龙江省富裕县也居住着数百柯尔克孜族人，他们的祖辈是18世纪从新疆迁去的。

柯尔克孜族有着悠久的历史 and 灿烂的文化，以热情好客闻名于世。目前，柯尔克孜族大部分人信仰伊斯兰教，属逊尼派的哈乃菲教派；居住在新疆额敏县和黑龙江省富裕县的柯尔克孜族受蒙古族影响，信仰藏传佛教。柯尔克孜语属阿勒泰语系突厥语族，有“皆盖”、“特斯开”两种方言。词汇的多源性是柯尔克孜语的一个特点。

达 斡 尔 族

达斡尔族是我国北方具有悠久历史和灿烂文化的少数民族之一。2000年人口普查资料显示，达斡尔族总人口为13.24万人，主要分布在内蒙古、黑龙江和新疆等省区。其中内蒙古莫力达瓦达斡尔族自治县、鄂温克族自治县、黑龙江省齐齐哈尔市梅里斯达斡尔族区，是达斡尔族的主要聚居区。另外，吉林、辽宁、北京、天津、上海、陕西、广西等全国二十几个省、市和自治区均有少量达斡尔族人口分布。

达斡尔族有自己的语言，属阿尔泰语系蒙古语族，无本民族文字，大多数人使用汉文。达斡尔族主要信仰萨满教，部分人信仰藏传佛教。“达斡尔”的意思是“开拓者”。达斡尔族的生产以农业为主，兼营牧业和狩猎。达斡尔族人的村落大多是依山傍水，风景秀丽。他们的村落、房屋都修建得十分整齐。在居室中建有三面相连的大炕，俗称“蔓子炕”，这种炕保温性能好，是达斡尔族人民不可缺少的取暖设施。达斡尔族人的男装多为长袍，两侧开口，长袍一般是蓝色、灰色，长袍上的衣扣大多用皮条编成。达斡尔族人以善于造车而闻名，被称为“草上飞”的北国名车——大轱辘车即出自达斡尔族人之手。达斡尔族人最为隆重的节日是“阿聂节”（相当于汉族的春节），这是一年一度的传统节日。由于冬季漫长寒冷，达斡尔族人民对被称之为“波依阔”（与现代曲棍球运动极其相似）的体育运动情有独钟。崇尚科学文化教育的达斡尔族，非常重视对后代的智力投资，不仅在自治旗建立健全了中小学教育体系，而且还纷纷送子女到全国各地学习深造。目前，达斡尔族每万人中拥有的大、中、小学生的平均数远远高于全国的平均水平，居各民族的第二位。

景 颇 族

景颇族是我国云南世居民族之一，主要聚居于德宏傣族景颇族自治州的山区，有一部分人居住在怒江傈僳族自治州的片马、古浪、岗房以及耿马、澜沧等县。现有人口 132143 人（2000 年）。景颇语属汉藏语系藏缅语族，1957 年在原有拉丁字母拼音文字的基础上创制了景颇文。

景颇族普遍信仰万物有灵的原始多神教，少数人信仰基督教，也有人信仰小乘佛教。景颇族住房多为竹木结构，分上下两层，上层住人，下层养家畜。景颇族以米为主食，吃法有烂米饭、糯米饭、糯米粑粑或竹筒饭。蔬菜除部分是自家种植的外，多靠采集的野菜来制作，吃法除煮、焖、炒或凉拌外，最有特色的要算是春菜了。景颇族口头文学十分丰富，有创世纪神话、历史传说、民间故事等，多与音乐相结合，又说又唱，词曲优美动听。

毛 南 族

毛南族，现有 10.72 万人。毛南族有自己的语言，属汉藏语系壮侗语族，通用汉语；主要聚居在中国云贵高原的茅南山、九万大山、凤凰山和大石山一带，而广西环江毛南族自治县的上南、中南、下南一带山区被称为“三南”，素有“毛南之乡”之称。毛南族的宗教信仰以道教为主，兼有一些佛教成分。毛南族主要从事农业生产，兼营各种副业。毛南族饲养的菜牛远销上海、香港等地，颇有声誉。他们编织的竹器，工艺精湛，著名的花竹帽，精致美观而又实用，既是毛南族的手工艺品，又是姑娘们珍爱的装饰品。

毛南族人的居室为干栏式建筑。毛南族石雕远近闻名，如今存留在凤凰山上的古墓群就是毛南石雕的典型代表。毛南族人的主食为大米、玉米，蔬菜有豌豆、白菜、南瓜、豆角、萝卜等。在节日的宴席上，豆腐是主要的食品。毛南族人民十分好客，客人到家都待为上宾，他们热情地把家里腌好的酸肉、酸螺蛳、酸菜都端出来款待客人，而且还要请客人品尝最有民族风味的“毛南饭”。毛南族居住的地方重峦叠嶂，耕地不足，他们在石山岩缝中把每一寸土地都开发出来，垒石保土，可以说惜土如金。

撒 拉 族

撒拉族人口目前已经超过 10 万（2000 年人口普查时为 104503 人）。撒拉族主要聚居在青海省循化撒拉族自治县和化隆回族自治县黄河谷地，以及甘肃省积石山保安族东乡族撒拉族自治县大河家乡一带。撒拉族自称“撒拉尔”，在汉文史籍中还有“撒拉儿”、“沙刺”、“撒刺”等多种写法，因信仰伊斯兰教，又被称为“撒拉回”，当地其他民族称之为“撒拉”。

撒拉族有自己的民族语言，属阿尔泰语系突厥语族西匈语支乌古斯语组。撒拉族在历史上曾使用以阿拉伯文字母为基础的撒拉文，本民族称之为“土尔克文”。撒拉族是我国信仰伊斯兰教的少数民族之一，宗教对其历史发展和政治、经济、文化等方面都有较深的影响。撒拉族的主要粮食为小麦、青稞、荞麦，制作的方法颇为讲究，有馒头、面片、拉面、散饭、搅团等许多品种。撒拉族人的服饰与回族人大致相同。撒拉族过伊斯兰教的主要节日圣纪节、开斋节和古尔邦节。大小清真寺遍布撒拉族各个村落，是穆斯林进行宗教活动的主要场所。

布 朗 族

布朗族是一个历史悠久的民族，现有人口 9.19 万人（2000 年）。布朗族大部分分布在云南省西部和西南部澜沧江中下游两岸的山岳地带，主要聚居于西双版纳傣族自治州勐海县边境线一带的布朗山、巴达、西定、勐满、打洛的山区和景洪县的小勐养和大勐笼，勐腊县的勐捧镇、芒果树乡，一小部分人散居在思茅地区、临沧地区和保山地区等地。

布朗族主要从事山地农业，粮食作物以旱稻和水稻为主，兼种玉米、芝麻、瓜果、豆类和薯类。经济作物有茶叶、棉花、棕片、大麻等。村寨周围是大片的茶林，这里是著名的“普洱茶”原料产地之一。优越的自然环境为布朗族人民生活提供了良好条件。

塔 吉 克 族

塔吉克族历史悠久。我国塔吉克族总人口目前已经超过 4 万人（2000 年人口普查资料显示，总人口为 4.10 万人），主要分布在新疆维吾尔自治区西南部的塔什库尔干塔吉克自治县。塔吉克族属欧罗巴人种印度地中海类型。

“塔吉克”是本民族的自称，意即“王冠”。塔吉克族有自己的语言，属于印欧语系伊朗语族帕米尔语支，分为色勒库尔塔吉克语和瓦罕塔吉克语两种方言。由于民族交往频繁，新疆许多塔吉克族人兼通维吾尔语和柯尔克孜语，普遍使用维吾尔文。

塔吉克族主要从事畜牧业，兼营农业，许多世纪以来他们在海拔 3000 米左右的山谷里安家落户，春天播种一些耐寒作物，初夏赶着畜群到高山草原放牧，秋后回村收获过冬，周而复始，过着半游牧半定居的生活。塔吉克族男子穿无领对襟长外套，系腰带，寒冷时外加羊皮大氅，戴羊皮高筒帽。女子穿连衣裙，戴圆顶绣花棉帽，缀有后帘，可以保暖。男女都喜欢穿长筒野羊皮软靴，用牦牛皮做靴底，轻柔坚实，适于在山路上行走。他们都喜食酥油、酸奶、奶疙瘩、奶皮子等乳制品和奶茶，并以肉食为上好食品，他们的住房大都为方形平顶，以木石构筑而成。由于高原多风雪，住房比较低矮，室内四周筑土炕。夏季上山放牧时，多住毡房。塔吉克族能歌善舞，舞蹈多模拟雄鹰展翅、飞翔回旋的姿势。

阿 昌 族

阿昌族是我国云南省特有的少数民族之一，人口有 33936 人（2000 年），其中德宏傣族景颇族自治州有 23060 人，占阿昌族总人口的 85.38%，主要分布在德宏州的陇川县户撒和梁河县九保、囊宋三个阿昌族乡，潞西县江东乡高埂田和盈江、瑞丽有少量分布。此外，还有一些阿昌族散居于大理白族自治州云龙县的漕涧、旧州等地和保山地区腾冲县和龙陵县。

阿昌族是中国云南境内最早的世居民族之一。阿昌族居住于滇西高山峡谷之中，其间气势磅礴的高黎贡山支脉自腾冲逶迤纵贯向南，形成绵延起伏的丘陵山地，地势北高南低。这一地区资源丰富，地下埋着宝藏，地上森林茂密，水利资源、森林资源、矿产资源等正被勘查或有待于开发利用。丰富的自然资源为阿昌族长期在这里定居提供了优厚的生存条件，为阿昌族社会经济文化的发展奠定了基础。阿昌族手工业发达，尤其擅长刀具的锻造。

普 米 族

普米族是我国具有悠久历史和古老文化的民族之一。2000 年人口普查资

料显示，普米族人口为 3.36 万人。云南省怒江州的兰坪县、丽江地区的宁蒗县、丽江县和迪庆藏族自治州的维西县是主要聚居地。

普米族信仰多种神，崇拜祖先，也有信仰藏传佛教的。普米族有自己的语言，普米语属汉藏语系藏缅语族，现在大多数人使用汉语。普米族十分喜爱“四弦琴”，无论在田间地头还是他们居住的木楞房中，到处都会听到抒情、纯朴的四弦琴声。按照普米族的古老传统，缔结婚姻、建立家庭是一个漫长而复杂的过程，因而请一个能说会道、精明能干的媒人是十分必要的。

鄂 温 克 族

鄂温克族是我国人口较少的少数民族之一，2000 年第五次全国人口普查资料显示，人口总数为 30505 人。鄂温克族主要分布在内蒙古自治区和黑龙江省。其中在内蒙古自治区有 23367 人，占全国鄂温克族总人口的 88.8%，黑龙江省有 2571 人，占 9.8%，其余的人散居在全国各省、市和自治区。

鄂温克族有自己的语言，鄂温克语属阿尔泰语系满—通古斯语族，分为海拉尔、陈巴尔虎和敖鲁古雅三种方言，无本民族文字，牧区通用蒙古文，农区和山区通用汉文。鄂温克族大部分人信仰萨满教，牧区有些人信仰藏传佛教。鄂温克是民族自称，意为“住在大山中的人”。历史上鄂温克族曾被称为“索伦”、“通古斯”、“雅库特”等，后统称鄂温克族。由于居住地不同，鄂温克族人生产、生活方式也有较大的差异，有的从事畜牧业，有的从事农业，还有的从事狩猎。驯鹿是鄂温克族人狩猎时必不可少的交通工具。他们饲养驯鹿已有很长的历史，他们称驯鹿为“鄂伦”。鄂温克族人长幼之间恪守尊卑有序的传统，长辈和老年人享有很高的威望。鄂温克族人极其好客，认为家里来客人是喜事，无论认识与否，他们都把来者视为贵客，总是拿出最好的东西款待客人。鄂温克族人崇尚天鹅，以天鹅为图腾。天鹅舞是鄂温克族的民间舞蹈，鄂温克语叫作“斡日切”。

怒 族

怒族是我国人口较少、使用方言较多的民族之一。现有人口 28759 人（2000 年），主要分布在云南怒江傈僳族自治州（简称怒江州）的贡山独龙族怒族自治县（简称贡山县）、福贡县、泸水县及兰坪白族普米族自治县（简称

兰坪县)。怒族有自己的语言,属汉藏语系藏缅语族,但各地方言差异很大,无民族文字。怒族主要信仰原始宗教,与藏族相邻的部分怒族人信仰藏传佛教,部分人信仰天主教、基督教。怒族人很早就经营农业。

怒族人住房有木板房和竹篾房两种。怒族妇女从小就要学习捻羊毛线、织羊毛袜子,姑娘长大后,要把自己织的羊毛袜子送给心上人,小伙子若收下羊毛袜子,就表示接受了姑娘的爱情。所以,看一个怒族小伙子有没有对象,只要看他脚下穿没穿上羊毛袜子就行了。怒族人喜欢自己酿酒,每当贵客光临,便以美酒相待。如将客人视为知己,便要与客户人脸对脸、嘴对嘴,一手搂肩,一手端酒杯,仰面将酒一饮而尽,这叫做饮“同心酒”。每年的农历三月十五日怒族的“朝山节”(也称为“鲜花节”)。节日当天,全寨人聚在一起前往当地的钟乳洞,接洞中钟乳石滴下的水,这水被称之为“仙奶”。回家后,“仙奶”被洒入粮食种子中,以祈愿来年的粮食获得丰收;也有把“仙奶”倒入醋、酒中的,以求身体健康不生病。

京 族

京族是主要从事沿海渔业的人口较少的一个少数民族,共有 2.25 万人(2000 年),分布在广西壮族自治区防城港市下辖的东兴市境内,主要聚居在江平镇的巫头、沱尾、山心三个海岛上,俗称“京族三岛”。京族有本民族的语言,系属尚未确定。京族人的服装朴素美观,独具风格。在京族人的村子里,最具有民族特色的建筑是“哈亭”,它不仅供奉全村人最为崇拜的神祇,以及每个家族的祖先牌位,而且京族人最为隆重的传统节日——哈节也在这里举行。近年来,京族地区的生产力水平和产业结构发生了巨大变化,除传统的捕捞、水产品加工和浅海滩涂养殖外,还发展起了边境贸易和旅游业,实现了渔、农、工、副、商全面发展。

基 诺 族

基诺族自称“基诺”,“基”为舅舅,“诺”为后边,意为尊敬舅舅的民族,他称为“攸乐”。现有 2.09 万人(2000 年),主要分布在云南省西双版纳傣族自治州景洪市的基诺山基诺族乡和景洪市勐旺乡补远村,其余的人散居在附近勐养镇、允景洪镇三达山、大渡岗乡及勐腊县勐仑乡等地。

基诺族有自己的语言，基诺语属汉藏语系藏缅语族，没有本民族文字，过去多靠刻木记事。基诺族普遍喜欢住在平缓的小山坡上，住房多为干栏式建筑。基诺族人聚居的基诺山，属热带地区，这里各种物产十分丰富。除种植水稻、玉米等农作物外，基诺山里的各种山珍野味也给基诺人提供了丰富的营养食品。基诺山竹子种类繁多，基诺族男子普遍会竹编工艺。细细的竹片，经他们的巧手编出花纹美观精巧的各种工艺品和生活用品。用竹筒做的打击乐器“奇可”是基诺族节庆和歌舞活动中不可缺少的，演奏起来音色优美和谐。基诺族有丰富的史诗、故事和传说。基诺族有自己特有的民族习俗、祭祀礼仪和节日。每个基诺族少年长到15岁左右就要举行成年礼，并要改装易服穿戴成年人的衣饰。“打铁节”，基诺语称“特毛切”，这是基诺族最隆重的节日，一般于每年农历十二月举行，历时三天。

德 昂 族

德昂族是一个历史悠久的民族。解放前，德昂族在我国境内仅有几千人，现有人口1.79万人（2000年），是云南省特有的、人口较少的少数民族之一。德昂族主要居住在云南省路西县与镇康县，少数人散居于盈江、瑞丽、陇川、保山、梁河、耿马等地，与傣族、景颇族、佤族等民族杂居在一起。德昂族有自己的语言，属南亚语系孟高棉语族，没有本民族文字，部分地区通用傣语、汉语、景颇语。

德昂族信仰佛教。德昂族人以农业生产为主，也善于种茶，素有“古老茶农”的美称。茶叶是德昂族重要的经济来源。德昂族还擅长编织竹器，文化上多受傣族影响。与南方许多民族一样，德昂族喜欢住干栏式竹楼。德昂族人好饮茶，主食以大米为主，玉米、小麦、豆类次之。德昂族的服饰以深色为主，男子多穿蓝、黑大襟上衣及宽而短的裤子，裹黑色或白色头巾，头巾两端饰以彩色绒球；妇女多穿黑色、藏青色的对襟上衣和长裙，配上银项圈、耳筒、耳坠、红绒球等首饰，别具特色。由于德昂族信仰佛教，人们的日常生活与佛教有密切关系。一般男孩儿到10岁便要入寺为僧，学习佛经，除少数人升为佛爷外，大多数人几年后便可还俗。德昂族人的传统节日也多与佛教有关，最隆重的节日是泼水节，届时人们要用“水龙”为佛像洗尘，排成长队，祝福吉祥，互相泼水共贺新年。

保安族

保安族是甘肃省三个特有的少数民族之一。保安族目前约有 1.65 万人(2000 年)。保安族主要聚居在甘肃省积石山保安族东乡族撒拉族自治县,其余的人散居在甘肃省临夏回族自治州、甘南藏族自治州及兰州市等地。保安族有本民族的语言,属阿尔泰语系蒙古语族,是该语族唯一完全丧失元音和谐的语言。没有本民族文字,通用汉文。

保安族的先民是信仰伊斯兰教的穆斯林,早在青海同仁居住时,伊斯兰教就在保安族中广泛传播。保安族农业以种植小麦、玉米、豆子为主,食品通常做成馒头、面条、油香、馓子或搅团、散饭。食用肉类必须要由阿訇屠宰,忌吃猪、马、驴、骡及一切凶禽猛兽的肉,一切自死之物和动物的血均不能食用。牛羊和家禽屠宰后不能说“死了”,否则即不能食用。舀水、舀饭均不能用反手。“油香”和“指甲面片”是保安族中很有特点的食品。

俄罗斯族

俄罗斯族,现有人口 1.56 万人。俄罗斯族主要居住在新疆的伊犁、塔城、阿勒泰、乌鲁木齐等地,其中伊犁地区人口较多,其余的人散居在黑龙江省与内蒙古自治区等地。俄罗斯族语言属印欧语系斯拉夫语族,使用俄文。俄罗斯族人大多信仰东正教,一小部分人信仰基督教。俄罗斯族人迁居新疆后,带来了先进的农、牧业生产器具,引进并培育了农牧业优良品种,并使用先进的畜牧业生产技术。

俄罗斯族人不仅善于经营农业,而且心灵手巧,很多人擅长手工业和各种修理业。俄罗斯族人性情开朗幽默,待人接客讲究礼节。他们最隆重的礼节是用面包和盐迎接客人,来客须用刀子切下一块面包沾少许盐吃下后方可进屋,这种礼节多用于迎接新娘和尊贵的客人,象征友谊和善意。在上个世纪,俄罗斯族人的住房多为土木建筑,屋顶用麦秸覆盖,住房分为卧室、客厅和储藏室。俄罗斯族重视教育,他们走到哪里就把学校办到哪里。在中国的俄罗斯族人中知识分子比例较高,不少人在大、中学校教授俄语,或在国家机关从事翻译工作。他们在新疆又学会了其他民族的语言,因此深受周围各民族人民的尊重与爱戴。

裕 固 族

裕固族是我国人口较少的民族之一。目前全国裕固族总人口接近 1.5 万人，主要分布在甘肃省张掖地区肃南裕固族自治县和酒泉市黄泥堡裕固族乡，其余的人散居在兰州和新疆哈密、昌吉等地。裕固族文化基本上属于北方游牧文化的范畴，除汉族文化之外，裕固族文化中包含有明显的蒙古族和藏族文化的成分。

裕固族起源于唐代游牧在鄂尔浑河流域的回鹘。9 世纪中叶，其中一支迁至甘肃河西走廊一带，史称“河西回鹘”。明初，陆续迁至祁连山一带，逐渐形成了裕固族。1953 年经本民族确定称“裕固”族。

裕固族是以畜牧业为主的民族。过去，他们不仅住的是毛毡帐篷，吃的肉、奶是畜产品，而且生产和生活上用的大部分用具也都是毛皮制成的。现在，已有一部分人主要从事农业生产。

裕固族大体上使用三种语言：东部裕固语（又称恩格尔语，属蒙古语族）、西部裕固语（又称尧乎尔语，属突厥语族）和汉语。新中国成立后，没有本民族作家文学的局面已被打破，裕固人开始用汉语来进行民族文学的创作。

乌 孜 别 克 族

我国乌孜别克族总人口有 1.24 万人，主要分布在北疆的乌鲁木齐、伊宁、木垒、奇台、塔城和南疆的喀什、和田、莎车、叶城等县市，但居住在伊宁的人较多。

乌孜别克族族名源于 14 世纪蒙古金帐汗国乌孜别克汗。15 世纪，金帐汗国解体。居住在撒马尔罕、花刺子模、安集延、布哈拉等地区的乌孜别克商人，沿着古代“丝绸之路”，经新疆到内地经商，往来频繁，络绎不绝，其中一部分商人逐渐在新疆一些城镇定居下来，繁衍生息，形成了中国的乌孜别克族。

乌孜别克族有自己的语言文字，其语言属于阿尔泰语系突厥语族西匈语支。由于长期与维吾尔族、哈萨克族杂居，所以大部分乌孜别克族人都用维吾尔文或哈萨克文。乌孜别克族的文化教育水平较高，很多人从事文化教育工作。

乌孜别克族信仰伊斯兰教。风俗习惯、衣食起居等和维吾尔族大致相同，

他们和维吾尔族、塔塔尔族有联姻的传统。乌孜别克族人的饮食丰富多彩，多吃羊、牛、马肉及其乳制品。一日三餐都离不开馕和奶茶，喜欢吃“库尔达克”（土豆炖肉）和蜂蜜、糖浆。乌孜别克族对子女的出生特别重视，孩子出生时，要举行隆重仪式，请知识渊博有名望的长者起名，并在孩子出生第11天大宴宾客以示祝福，在出生40天时还要举行“洗礼”。

门 巴 族

门巴族是我国人口较少的民族之一，主要分布在西藏自治区门隅和墨脱等地。2000年人口普查资料显示，门巴族总人口为8923人，其中墨脱县6069人，林芝县542人，错那县549人，此外，米林、乃东、拉萨等县市及在内地工作学习的有338人。

门巴族有自己的语言，属汉藏语系藏缅语族，通用藏文。门巴族人主要信仰藏传佛教，也有部分人信仰原始宗教。门巴族人主要从事农业生产，种植水稻，也兼营畜牧业和狩猎。门巴族与藏族关系密切，在生活习俗方面受藏族影响较大。在经济生活中，门巴族人享誉四方的是他们的家庭手工业，特别是加工木碗和编织竹器的技艺，更是长盛不衰。

鄂 伦 春 族

鄂伦春族是我国民族大家庭中人口较少的民族之一，他们自古以来一直从事狩猎生产。其性情淳朴、坚强，以勇敢强悍而著称，被誉为“兴安岭之王”。“鄂伦春”的族名，包含两种意思，即“山岭上的人们”和“使用驯鹿的人”。

鄂伦春族的人口发展，只有百年的历史可查。据记载，1895年鄂伦春人口约有18000人，1917年约有4111人，1938年约有3000人，新中国建立前鄂伦春人口的急剧下降，与游猎民族不断迁徙的生活方式和近现代社会战争频繁密切相关。新中国建立后，由于实施了正确的民族政策，国家积极扶持和帮助鄂伦春族经济社会的发展，鄂伦春族人口数量增长较快，素质有了极大的提高。2000年人口普查资料显示，鄂伦春族总人口为0.82万人。目前，我国的鄂伦春族人口已近万人。鄂伦春语属阿尔泰语系满一通古斯语族的北语支，没有文字，曾经使用过满文，现在主要使用汉语。新中国成立前，鄂

伦春族信仰萨满教，崇拜各种自然物，相信万物有灵，但以崇拜祖先为主。鄂伦春人以食兽肉为主，主要捕猎狍、鹿、野猪等。他们特别喜欢生吃狍肝、喝烧酒和马奶酒。狩猎是鄂伦春族人生存的需要，一年四季他们都游猎在茫茫的林海中。鄂伦春族无论男女老少都能歌善舞，他们常常将诗歌、音乐、舞蹈结合在一起，形成自己独特的民族风格。

独 龙 族

独龙族是我国少数民族中人口较少的民族之一。全国独龙族人口共有7426人（2000年），云南省有5536人，其中71%聚居在贡山独龙族自治县独龙江流域的河谷地带，22.6%分布在怒江峡谷中游河谷地带，其余的散居于澜沧江上游等地。独龙族聚居区山高谷深，群峰环抱，沟壑纵横，交通极为不便。独龙族以农业为主，但也保留着渔猎的传统。他们相信万物有灵，崇拜山川、河流、大树、巨石等自然物。

讲究信用、遵守诺言是他们的道德传统。历史上，独龙族被誉为不用锁门的民族，他们始终保持着“路不拾遗，夜不闭户”的古老而纯朴的社会风尚。世代生活在独龙江两岸的独龙族人，已把他们的生命与峡谷中的一切融为一体。其房屋多建在山坡上，依山傍水构筑成大大小小的村寨。一座座竹楼轻巧地建在陡峻的山坡上，称之为“长竹楼”。独龙族人的食物丰富多彩。他们虽然以谷物为主食，但对各种山珍野味情有独钟，特别喜欢吃含有丰富淀粉的野生植物。从一种被称作“四维”的从董棕树芯里挖出的董棕粉制作的食物是独龙族最喜爱的食物，也是招待来宾以示热情友好的佳肴。独龙毯作为装饰品在独龙族人家里随处可见。独龙族地区的资源十分丰富，有多种珍贵植物被列为国家重点保护对象。这里的大理石矿是富集矿区，彩色大理石“贡翠”是天然的大理石珍品。

塔 塔 尔 族

我国塔塔尔族人口目前已超过0.5万人（2000年人口普查资料显示，总人口为4890人），是我国人口较少的民族之一。塔塔尔族信仰伊斯兰教，属伊斯兰教中的逊尼派。现在主要从事畜牧业生产，一部分从事农业、手工业。塔塔尔族人多在秋冬季节结婚，婚礼在新娘家举行。塔塔尔族的葬礼按伊斯

兰教规定进行，但没有为死者戴孝的习俗。

塔塔尔族过“肉孜节”、“古尔邦节”，本民族传统的民族节日是“萨邦节”。“萨邦节”多在春天举行，并开展摔跤、穿麻袋赛跑、叼匙竞走、爬滑竿、赛马等各种娱乐活动。塔塔尔语是在保加尔语和克卜察克语的基础上发展形成的，属阿尔泰语系突厥语族克卜察克语。塔塔尔族的文化水平较高。他们很早就接受了欧洲文化，并将其与本民族文化结合在一起，形成了一种新的文化。塔塔尔族的文学艺术具有悠久的历史。塔塔尔族的教育事业发展较早，因而其人口中知识分子所占的比例很高。

赫 哲 族

赫哲族，现有人口 4640 人（2000 年），大部分居住在我国东北部的黑龙江省的同江、抚远、饶河等市县，另一部分散居在佳木斯、富锦、集贤、桦川、依兰等县。赫哲族有本民族的语言，赫哲语属阿尔泰语系满—通古斯语族，没有本民族文字，大多数人使用汉文。赫哲族信仰萨满教，相信万物有灵。赫哲族是中国北方唯一的以捕鱼为生、用狗拉雪橇的民族。以渔猎生产为主的赫哲族人餐桌上的主要菜肴离不开鱼。他们吃鱼的方法很多，也很讲究。赫哲族人贮藏鱼的地方叫鱼楼，一般都建在居所庭院中，几乎每家一座，平时还用来放置渔具等物品。

狗拉雪橇是赫哲人主要的冬季交通工具。在长期的渔业生产中，他们积累了丰富的捕鱼经验，练就了高超的捕鱼技术，方法多种多样，工具也五花八门。从前，捕鱼和狩猎是赫哲人衣食的主要来源。他们不仅以鱼肉、兽肉为食，穿的衣服也多半是用鱼皮、狍皮和鹿皮制成。“乌日贡节”是赫哲族人在 1985 年新设立的节日。赫哲族人民喜爱音乐，善于唱歌，流传着许多民间歌曲，歌曲优美舒展，旋律奔放。

高 山 族

高山族，现有人口约 40 万。99% 以上高山族人民居住在台湾，台湾省高山族有 30 多万人，另有居住在平原地区、汉化程度颇高的高山族即平埔人约 10 万。据 2000 年第五次全国人口普查，大陆有高山族 4461 人，在全国各省、市、自治区均有分布，主要分布在福建、山东、北京等省市。高山族有自己

的语言，属南岛语系印度尼西亚语族，没有文字，通用汉文。

基督教传入后，一部分高山族人信奉基督教。高山族的传统房屋一般用竹子做墙壁，用木头做立柱与横梁，以茅草盖顶。高山族的食物比较简单，以大米、粟米和甘薯为主食，烤鹿肉和酸鹿肉是高山族的风味食品。以前，高山族常用的衣料是用苧麻自织的“番布”。高山族喜欢用鸡尾、鸟羽做头饰。高山族是一个能歌善舞的民族，被誉为“歌舞繁荣的社会”。其杵舞和长发舞都具有很高的艺术水平。高山族的节日往往与农事活动有关，比如播种节和丰收节，主要活动是祭祖、举行农耕仪式、会餐、歌舞娱乐等。

珞巴族

珞巴族是我国人口较少的民族之一，2000年人口普查资料显示，珞巴族总人口为2965人。主要分布在西藏东南部的珞渝地区，以墨脱、米林、隆子、林芝、察隅等几个边境县最集中，另有少数散居于拉萨、山南等地。珞巴族的语言属汉藏语系藏缅语族，各地方言差别大，没有本民族文字。新中国成立前，只有极少数人通晓藏语和藏文，通常以刻木、结绳记事。珞巴族主要从事农业。采集和狩猎在生产中也占相当的比重。

珞巴族的传统住房是石木结构的碉房，坚固耐久且具有很好的防御功能。珞巴族的服饰充分体现着粗犷豪放的风格，野生植物纤维和兽皮是他们服装的主要原料。珞巴族人传统的主食是玉米和鸡爪谷。烧烤是他们最常见的加工食物的方法。每年庄稼收获后，珞巴族都要举行近两天的丰收节，叫作“昂德林节”。节日期间全村男女身穿盛装，各家端出准备好的美酒，共享节日食品。也许因为没有本民族文字，也许是神奇美丽的喜马拉雅山给予的无穷艺术灵感，珞巴族人的口头文学十分丰富。

第二节 中国少数民族体质特征

壮族

亚中等身材。面型较宽，属阔面型。头型较宽，男性多为中头型，女性

第一章 中国少数民族简况及其体质特征

多为圆头型。直发黑色，褐色眼，眼裂开度中等，眼裂斜度内外平行，男性多无蒙古褶，女性多有蒙古褶。鼻根低，男性鼻梁多为直型和波型，女性多为凸型和波型，鼻尖上翘，鼻翼宽大。唇型多为凸型唇和正型唇，红唇中等偏厚。耳壳多无达尔文结节，耳垂多为圆形、方形和三角形。

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	测量特征	头长	187.40	178.80
	头宽	150.20	144.40	
	头围	560.80	545.40	
	头矢状弧	366.60	352.60	
	头冠状弧	355.50	343.60	
	面高	184.40	175.80	
	面宽	142.70	134.70	
	额宽	100.40	97.80	
	口宽	49.50	46.60	
	鼻高	53.80	51.10	
	鼻宽	40.40	37.40	
	鼻长	—	—	
	指数特征	头指数(头宽/头长×100)	80.30	80.80
	面指数(面高/面宽×100)	84.90	85.50	
	鼻指数(鼻宽/鼻高×100)	75.40	73.50	
观察特征	面部	卵圆		
眼部	多无蒙古褶，有上眼睑皱褶			
鼻部	直鼻梁，鼻孔卵圆			
耳部	方形、圆形和三角耳垂， 无达尔文结节			
口部	正唇			
毛发	直发，眉毛稀少			

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1635.00	1551.00	
	肩宽	364.50	328.20	
	上肢全长	—	—	
	上臂长	—	—	
	手长	—	—	
	腰围	—	—	
	骨盆宽	—	—	
	下肢全长	—	—	
	大腿长	—	—	
	小腿长	—	—	
	足长	—	—	
	足高	—	—	
	体重(kg)	52.50	44.80	
	指 数 特 征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
足长宽指数		—	—	
上肢全长/身高		—	—	
上下肢长度指数		—	—	
满 族				

身材中等。头型多圆头型、高头型和阔头型。面型多为狭面型和中面型。发旋，男性胡须少。眼裂开度中等，眼裂斜度外高内低，无蒙古褶，上眼睑皱褶少。鼻根中等偏低，鼻梁直，鼻尖及鼻基部水平，鼻翼微突，鼻宽大于两眼内角宽。耳垂多三角形，耳壳大多无达尔文结节。口宽中等，正唇型，红唇中等偏薄。

第一章 中国少数民族简况及其体质特征

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征		头长	182.98	170.48
		头宽	157.04	148.74
		头围	556.97	544.25
	测量特征	头矢状弧	346.23	340.60
		头冠状弧	364.18	356.06
		面高	193.12	187.16
		面宽	144.85	140.11
		额宽	108.10	104.25
		口宽	50.94	47.62
		鼻高	53.13	48.44
		鼻宽	37.36	35.82
		鼻长	—	—
		指数特征	头指数(头宽/头长×100)	84.89
		面指数(面高/面宽×100)	86.03	87.11
		鼻指数(鼻宽/鼻高×100)	70.88	73.33
	观察特征	面部	狭或中面型	
		眼部	蒙古褶微显或无, 有上眼睑皱褶	
鼻部		直或凸鼻梁		
耳部		三角形耳垂, 无达尔文结节		
口部		正唇		
	毛发	卷发, 男性胡须少		
体 部 特 征	测量特征	身高	1645.42	1534.23
		肩宽	—	—
		上肢全长	—	—
		上臂长	—	—
		手长	—	—
		腰围	—	—
		骨盆宽	—	—

续表

		参 数	男(mm)	女(mm)
体 部 特 征	测量特征	下肢全长	—	—
		大腿长	—	—
		小腿长	—	—
	指数特征	足长	—	—
		足高	—	—
		体重(kg)	58.32	51.37
		身高坐高指数(坐高/身高×100)	54.01	54.52
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
		足长宽指数	—	—
	上肢全长/身高	—	—	
	上下肢长度指数	—	—	

回 族

身材中等偏高，黑发直形，男性胡须较少。头型多高头型、圆头型和狭头型。面型以中面型、阔面型和狭面型为多。眼裂开度中等，半数以下有蒙古褶，上眼睑皱褶发育好。鼻根中等高，鼻梁直形，鼻基水平方向，鼻尖以尖小和中间型为主，鼻孔多三角形和圆形，属狭鼻型。耳垂多圆形和三角形，多数无达尔文结节。红唇中等偏厚。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	180.00	170.21
		头宽	151.26	144.83
		头围	555.18	538.59
		头矢状弧	—	—
		头冠状弧	—	—
		面高	190.85	180.44

续表

		参 数	男(mm)	女(mm)	
头 面 部 特 征	测量特征	面宽	141.65	134.53	
		额宽	103.72	102.48	
		口宽	49.41	45.51	
		鼻高	50.66	46.13	
		鼻宽	37.30	33.94	
			鼻长	—	—
	指数特征	头指数(头宽/头长×100)	84.21	85.10	
		面指数(面高/面宽×100)	85.63	83.60	
		鼻指数(鼻宽/鼻高×100)	74.16	74.08	
	观察特征	面部	中面型、阔面型和狭面型		
眼部		多无蒙古褶, 有上眼睑皱褶			
鼻部		鼻梁直型			
耳部		圆形或三角形耳垂, 多无达尔文结节			
口部		正唇, 上唇皮肤高度中等			
	毛发	黑发直形			
体 部 特 征	测量特征	身高	1671.50	1569.16	
		肩宽	381.77	347.12	
		上肢全长	744.71	698.44	
		上臂长	295.83	270.80	
		手长	184.25	171.25	
		腰围	—	—	
		骨盆宽	—	—	
		下肢全长	887.62	850.62	
		大腿长	428.21	408.36	
		小腿长	420.08	387.23	
		足长	243.70	223.30	
	足高	62.69	56.89		
	体重(kg)	59.90	51.73		

续表

		参 数	男(mm)	女(mm)
体 部 特 征	指数特征	身高坐高指数(坐高/身高×100)	53.75	54.21
		身高肩宽指数(肩宽/身高×100)	22.85	22.05
		手长宽指数	46.10	44.16
		足长宽指数	40.99	41.16
		上肢全长/身高	44.58	44.50
		上下肢长度指数	83.98	82.80
苗 族				
身材矮小, 肤色深。男性多为中面型和狭面型, 女性多为狭面型。体毛稀少, 发黑, 直型发。眼裂开度中等, 眼裂斜度内外平行, 多无蒙古褶。直型鼻梁, 鼻翼微突, 多属狭鼻型和中鼻型, 鼻孔多呈卵圆。红唇, 正唇形。耳垂形状多为方形、圆形和三角形, 多数无达尔文结节。腿型多为短腿或中腿型。				
体质特征参数				
		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	180.02	173.92
		头宽	147.33	141.20
		头围	544.31	526.59
		头矢状弧	337.35	332.34
		头冠状弧	340.07	329.82
		面高	189.42	181.93
		面宽	116.08	108.38
		口宽	52.14	50.78
		额宽	100.30	96.09
		鼻高	53.62	50.52
	鼻宽	36.88	32.98	
	鼻长	—	—	

续表

		参 数	男(mm)	女(mm)
头	指数特征	头指数(头宽/头长 $\times 100$)	81.90	81.23
		面指数(面高/面宽 $\times 100$)	101.56	100.65
		鼻指数(鼻宽/鼻高 $\times 100$)	76.46	77.33
面部特征	观察特征	面部	卵圆	
		眼部	多无蒙古褶, 有上眼睑皱褶	
		鼻部	直鼻梁, 鼻孔卵圆	
		耳部	方形、圆形和三角耳垂, 无达尔文结节	
		口部	正唇	
		毛发	直发, 眉毛稀少	
体部特征	测量特征	身高	1534.66	1434.78
		肩宽	355.70	323.66
		上肢全长	669.84	627.86
		上臂长	280.00	273.30
		手长	176.21	169.71
		腰围	712.47	754.47
		骨盆宽	275.25	276.22
		下肢全长	853.41	791.54
		大腿长	452.58	406.77
		小腿长	338.40	325.62
		足长	236.31	223.41
		足高	62.41	58.69
		体重(kg)	51.00	42.11
		指数特征	指数特征	身高坐高指数(坐高/身高 $\times 100$)
身高肩宽指数(肩宽/身高 $\times 100$)	20.10			21.60
手长宽指数	45.50			43.10
足长宽指数	41.40			41.20
上肢全长/身高	45.30			44.80
上下肢长度指数	82.70			82.60

维 吾 尔 族

身材较高，黑色直形发，眉毛发育好，黑褶色眼，眼裂中等，眼裂斜度内外平行，上眼睑皱褶发育好，约半数人有蒙古褶。鼻根较高，直型鼻梁，鼻尖向前，鼻孔卵圆，狭型鼻。红唇，唇薄，多为凸唇型。面型多卵圆，头型多阔头型、圆头型和高头型。

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	186.20	175.60	
	头宽	160.60	153.20	
	头围	557.70	539.50	
	头矢状弧	340.00	323.60	
	头冠状弧	357.70	345.30	
	测量特征	面高	192.20	182.20
	面宽	145.00	134.80	
	额宽	114.90	111.00	
	口宽	49.60	47.40	
	鼻高	56.30	53.60	
	鼻宽	35.00	33.30	
	鼻长	55.50	52.10	
	指数特征	头指数(头宽/头长×100)	88.62	88.78
面指数(面高/面宽×100)	87.05	85.65		
鼻指数(鼻宽/鼻高×100)	62.39	62.56		
观察特征	面部	卵圆		
	眼部	无蒙古褶，有上眼睑皱褶		
	鼻部	直鼻梁，鼻孔卵圆		
	耳部	圆形耳垂，有达尔文结节		
	口部	凸唇		
	毛发	直发，眉毛浓密中等		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1694.60	1578.80	
	肩宽	396.80	359.34	
	上肢全长	737.30	683.60	
	上臂长	317.70	296.70	
	测量特征	手长	191.80	177.10
		腰围	760.30	667.80
		骨盆宽	277.20	271.80
		下肢全长	899.00	855.70
		大腿长	—	—
		小腿长	384.10	393.50
		足长	251.60	231.60
		足高	72.20	66.90
		体重(kg)	61.00	52.13
		指数特征	身高坐高指数(坐高/身高×100)	52.22
	身高肩宽指数(肩宽/身高×100)		23.56	22.77
	手长宽指数		46.23	44.76
	足长宽指数		39.75	39.32
上肢全长/身高	43.76		43.29	
上下肢长度指数	—	—		

土 家 族

身材较矮小，黑发直形。头型多圆头型、高头型和中头型。面型多为阔面型。眼裂斜度外高内低，多数有蒙古褶和上眼睑皱褶。鼻根中等高，鼻梁直微凹，鼻基底水平稍下垂，属中等偏狭鼻型，女性属中鼻型和窄鼻型。耳垂多圆形，多数有达尔文结节。正唇型，红唇较厚，下颏多直型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征		头长	183.90	174.50
		头宽	151.90	146.00
		头围	548.60	539.70
	测量特征	头矢状弧	344.70	333.10
		头冠状弧	352.10	340.40
		面高	186.40	177.70
		面宽	141.50	133.80
		额宽	105.40	102.70
		口宽	48.50	45.60
		鼻高	48.90	46.00
		鼻宽	35.40	32.40
		鼻长	48.50	44.70
		指数特征	头指数(头宽/头长×100)	82.61
	面指数(面高/面宽×100)		84.30	82.95
	鼻指数(鼻宽/鼻高×100)		72.88	70.92
	观察特征	面部	阔面型	
		眼部	无蒙古褶或微显	
鼻部		鼻梁直微凹		
耳部		圆形或三角形耳垂, 达尔文结节微显		
口部		正唇		
	毛发	黑发直形		
体 部 特 征	测量特征	身高	1592.60	1487.60
		肩宽	364.00	327.60
		上肢全长	709.00	657.20
		上臂长	307.60	283.40
		手长	176.30	165.70
		腰围	—	—
		骨盆宽	—	—

续表

		参 数	男(mm)	女(mm)
体 部 特 征	测量特征	下肢全长	845.90	799.00
		大腿长	445.10	423.20
		小腿长	336.20	315.20
		足长	236.80	220.00
		足高	64.70	60.50
		体重(kg)	52.70	47.80
	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	22.85	22.02
		手长宽指数	—	—
		足长宽指数	—	—
上肢全长/身高		44.43	44.06	
	上下肢长度指数	—	—	

彝 族

身材中等偏矮。头型属于中头型。皮肤颜色较浅。多为中—阔面型。发直而黑，男性眉毛较浓。眼裂开度中等，眼裂斜度外高内低，多有蒙古褶。鼻根高度中等，鼻梁平直，男性鼻基部水平，女性鼻尖上翘，鼻孔多圆形。耳垂呈方形。多为凸唇型，下颏直型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	191.72	183.74
		头宽	150.44	145.49
		头围	556.36	535.03
		头矢状弧	348.49	334.66
		头冠状弧	337.67	326.14
		面高	190.55	179.46

续表

		参 数	男(mm)	女(mm)	
头 面 部 特 征	测量特征	面宽	142.45	135.56	
		额宽	106.46	105.20	
		口宽	53.17	50.81	
		鼻高	51.45	47.64	
		鼻宽	36.22	33.33	
			鼻长	—	—
	指数特征	头指数(头宽/头长×100)	78.52	79.26	
		面指数(面高/面宽×100)	84.55	83.24	
		鼻指数(鼻宽/鼻高×100)	70.96	70.62	
	观察特征	面部	中—阔面型		
眼部		约半数有蒙古褶，有上眼睑皱褶			
鼻部		多直鼻梁，鼻孔多卵圆			
耳部		方形耳垂，多无达尔文结节或微显			
口部		凸唇			
	毛发	直发，眉毛密度中等			
体 部 特 征	测量特征	身高	1625.56	1514.13	
		肩宽	365.66	337.39	
		上肢全长	716.60	664.31	
		上臂长	304.69	277.90	
		手长	183.95	176.67	
		腰围	—	—	
		骨盆宽	263.81	262.21	
		下肢全长	839.95	793.03	
		大腿长	422.64	402.15	
		小腿长	349.72	330.13	
		足长	240.97	223.56	
		足高	67.39	61.70	
	体重(kg)	55.55	49.32		

续表

		参 数	男(mm)	女(mm)
体 部 特 征	指数特征	身高坐高指数(坐高/身高×100)	53.50	53.50
		身高肩宽指数(肩宽/身高×100)	23.30	23.00
		手长宽指数	45.50	43.00
		足长宽指数	42.70	42.50
		上肢全长/身高	45.60	45.10
		上下肢长度指数	85.40	84.10

蒙 古 族

身材高大,指距大于身高。眼裂开度中等,眼裂斜度多为内外平行,多数有上眼睑皱褶和蒙古褶。鼻根中等,鼻被平直,鼻翼微突,多属中鼻型。耳垂男性多为圆形,女性多为圆形和三角形。头型属圆头型和高头型。面型属狭面型。男性属长躯干型,女性属中躯干型,男女均属中胸型、中肩型和中骨盆型。四肢分型属于中腿型和中臂型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	186.2	175.6
		头宽	156.1	148.8
		头围	570.00	549.50
		头矢状弧	327.70	312.40
		头冠状弧	355.10	340.90
		面高	135.22	127.72
		面宽	146.6	136.7
		额宽	110.8	106.4
		口宽	54.9	50.7
		鼻高	51.9	45.30
		鼻宽	38.9	33.9
		鼻长	49.16	45.30

续表

		参 数	男(mm)	女(mm)
头	指数特征	头指数(头宽/头长×100)	83.96	84.98
		面指数(面高/面宽×100)	82.95	79.78
		鼻指数(鼻宽/鼻高×100)	75.06	75.17
面部特征	观察特征	面部	椭圆	卵圆
		眼部	蒙古褶, 上眼睑皱褶	
		鼻部	直鼻梁, 鼻孔卵圆	
		耳部	圆形耳垂, 无达尔文结节	
		口部	凸唇	
		毛发	直发, 眉毛浓密中等	
体部特征	测量特征	身高	1712.00	1574.00
		肩宽	389.00	350.00
		上肢全长	740.00	678.00
		上臂长	316.93	294.80
		手长	186.65	176.45
		腰围	833.00	768.00
		骨盆宽	295.00	289.00
		下肢全长	893.32	841.64
		大腿长	448.80	425.69
		小腿长	374.47	344.91
		足长	251.10	228.70
		足高	72.00	65.60
		体重(kg)	68.53	56.93
指数特征	指数特征	身高坐高指数(坐高/身高×100)	53.75	54.16
		身高肩宽指数(肩宽/身高×100)	22.74	22.24
		手长宽指数	46.42	44.69
		足长宽指数	40.06	40.00
		上肢全长/身高	43.67	43.12
		上下肢长度指数	77.99	76.98

藏 族

身材中等，黑发直形，发质较硬，体毛稀少。头型多阔头型。面型阔面型为多。褐色眼，眼裂开度中等偏狭，眼裂斜度外高内低，有蒙古褶。鼻根中等高，鼻梁直形，鼻孔卵圆形，属狭鼻型或中鼻型。耳垂多圆形和方形，多数无达尔文结节。唇中等厚，唇型稍突，颧骨突出。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	193.30	183.35
		头宽	156.10	148.18
		头围	569.23	576.49
		头矢状弧	—	—
		头冠状弧	—	—
		面高	192.20	182.18
		面宽	147.50	139.06
		额宽	107.00	103.52
		口宽	50.50	46.29
		鼻高	59.10	47.62
		鼻宽	38.60	35.32
			鼻长	49.80
	指数特征		头指数(头宽/头长×100)	80.37
		面指数(面高/面宽×100)	85.64	83.27
		鼻指数(鼻宽/鼻高×100)	74.95	74.73

续表

		参 数	男(mm)	女(mm)
头 面 部 特 征	观察特征	面部	阔面型	
		眼部	有蒙古褶	
		鼻部	鼻梁较直	
		耳部	圆形或方形耳垂, 多无达尔文结节	
		口部	多凸唇, 其次正唇	
		毛发	黑发直形	
体 部 特 征	测量特征	身高	1629.34	1526.91
		肩宽	366.07	333.94
		上肢全长	726.18	666.12
		上臂长	296.74	286.64
		手长	176.42	165.88
		腰围	794.96	810.90
		骨盆宽	266.25	279.56
		下肢全长	844.84	737.72
		大腿长	446.79	426.06
		小腿长	357.11	335.99
		足长	238.70	220.67
		足高	68.37	61.07
		体重 (kg)	55.53	53.18
指数特征	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
		足长宽指数	—	—
		上肢全长/身高	—	—
		上下肢长度指数	—	—

布依族

身材矮小，躯干较宽，皮肤黄色或浅黄。黑发直形。头型多圆头型和中头型。面型多阔面型。褐色眼，眼裂开度中等偏宽，眼裂斜度外高内低，多缺失蒙古褶，上眼睑皱褶发育好。鼻根中等或低型，鼻翼高度中等，属中鼻型。耳垂多圆形，其次为三角形。上唇高度中等。

体质特征参数

参 数		男(mm)	女(mm)
头 面 部 特 征	测量特征		
	头长	185.30	177.00
	头宽	151.90	147.60
	头围	556.20	539.60
	头矢状弧	—	—
	头冠状弧	—	—
	面高	185.40	175.00
	面宽	140.80	134.00
	额宽	110.2	106.70
	口宽	50.50	48.50
指数特征	鼻高	50.80	46.90
	鼻宽	38.20	35.00
	鼻长	—	—
	头指数(头宽/头长×100)	82.10	83.50
	面指数(面高/面宽×100)	81.90	80.10
	鼻指数(鼻宽/鼻高×100)	80.10	79.80
观察特征	面部	阔面型	
	眼部	多无蒙古褶，上眼睑皱褶发育好	
	鼻部	中鼻型	
	耳部	耳垂多圆形，其次为三角形	
	口部	上唇高度中等	
	毛发	黑发直形	

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1586.00	1494.00	
	肩宽	374.00	341.00	
	上肢全长	700.00	649.00	
	上臂长	302.00	281.00	
	手长	—	—	
	测量特征	腰围	—	—
		骨盆宽	280.00	279.00
		下肢全长	867.00	824.00
		大腿长	432.00	409.00
		小腿长	371.00	352.00
	足长	—	—	
	足高	—	—	
	体重(kg)	53.20	47.40	
	指数特征	身高坐高指数(坐高/身高×100)	52.70	52.80
		身高肩宽指数(肩宽/身高×100)	23.60	22.80
手长宽指数		—	—	
足长宽指数		—	—	
上肢全长/身高		—	—	
上下肢长度指数		81.00	78.90	
侗 族				

身材较矮小。头型多阔头型、圆头型、高头型。面型多为中面型和狭面型。眼裂开度中等，眼裂斜度外高内低，大多有蒙古褶，上眼睑皱褶发育好。鼻根低平，鼻梁直，鼻尖向前，鼻基部水平和上翘，鼻翼高度和突度中等，鼻孔多倾斜型。耳垂多圆形和方形，半数人有达尔文结节。口裂较宽，正唇或凸唇型，下颏直型。

第一章 中国少数民族简况及其体质特征

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	185.33	179.02
		头宽	149.82	145.90
		头围	543.52	522.88
		头矢状弧	336.37	326.47
		头冠状弧	344.34	338.60
	观察特征	面部	中或狭面型	
		眼部	蒙古褶中显, 有上眼睑皱褶	
		鼻部	直鼻梁	
		耳部	圆形, 方形、三角形耳垂, 多无达尔文结节	
		口部	凸唇	
		毛发	发浓密、黑直	
		指数特征	头指数(头宽/头长×100)	80.89
	面指数(面高/面宽×100)		81.17	80.00
	鼻指数(鼻宽/鼻高×100)		78.08	79.45
	体 部 特 征	测量特征	身高	1607.60
肩宽			382.70	345.10
上肢全长			710.20	655.80
上臂长			299.64	275.16
手长			177.85	165.57
腰围			—	—
骨盆宽			265.04	277.93
观察特征		下肢全长	850.14	799.97
		大腿长	431.82	417.52
		小腿长	351.87	320.95
		足长	236.63	218.62
		足高	66.89	61.16
		体重(kg)	52.08	46.97
指数特征		身高坐高指数(坐高/身高×100)	52.60	52.90
		身高肩宽指数(肩宽/身高×100)	23.10	22.20
	手长宽指数	47.20	45.90	
	足长宽指数	42.30	42.00	
	上肢全长/身高	45.00	44.30	
	上下肢长度指数	83.70	81.90	

瑶 族

身材矮小，体型属中间型。男性头型属于圆头型、高头型和中头型，女性多数属狭头型。皮肤浅棕色。男性多为阔面型，女性多为超阔面型。发直而黑，眉毛男性中等，女性稀少。眼裂开度中等，眼裂斜度外高内低，多有蒙古褶。鼻根高度中等，鼻梁平直，鼻尖上翘。少数有达尔文结节，耳垂呈方形和三角形。多为凸唇型，红唇厚度中等。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	180.73	173.71
		头宽	148.84	143.80
		头围	543.81	534.70
		头矢状弧	341.11	326.77
		头冠状弧	345.99	338.62
		面高	189.38	182.39
		面宽	138.25	131.32
		额宽	100.35	97.94
		口宽	56.76	52.97
		鼻高	52.44	50.24
	鼻宽	37.71	35.01	
	鼻长	—	—	
	指数特征	头指数(头宽/头长×100)	82.40	82.90
面指数(面高/面宽×100)		86.30	86.20	
鼻指数(鼻宽/鼻高×100)		72.20	70.00	
观察特征	面部	阔面		
	眼部	多无蒙古褶，无上眼睑皱褶		
	鼻部	多直鼻梁或凹鼻梁，鼻孔三角或卵圆		
	耳部	方形或圆形耳垂，多无达尔文结节		
	口部	凸唇		
	毛发	直发，眉毛稍稀少		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1568.30	1472.60	
	肩宽	362.26	307.39	
	上肢全长	697.14	655.90	
	上臂长	294.74	273.24	
	手长	173.60	159.99	
	测量特征	腰围	—	—
	骨盆宽	262.02	264.94	
	下肢全长	835.88	793.02	
	大腿长	416.07	389.73	
	小腿长	353.69	344.65	
	足长	240.36	225.63	
	足高	66.36	59.44	
	体重(kg)	48.77	44.66	
	指数特征	身高坐高指数(坐高/身高×100)	53.70	53.60
	身高肩宽指数(肩宽/身高×100)	23.10	20.90	
	手长宽指数	—	—	
足长宽指数	—	—		
上肢全长/身高	44.40	44.50		
上下肢长度指数	—	—		
朝 鲜 族				

身材中等偏矮，黑发直形。头型多圆头型和阔头型。面型多为卵圆型。眼裂开度狭窄，眼裂斜度外高内低，蒙古褶显著，上眼睑皱褶发育好。鼻根中等偏低，男性鼻背直凸，女性鼻梁多凸型，鼻尖和鼻基底方向，男性多水平，女性多上翘。鼻翼突度不显著，鼻孔多卵圆型，男性属高鼻型和中鼻型，女性属中鼻型和窄鼻型。男性耳垂多圆形，女性耳垂多三角形，多数有达尔

中国少数民族遗传学概论

文结节。红唇中等厚，下颏多直型或微突。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	183.90	174.70
		头宽	157.70	151.70
		头围	—	—
		头矢状弧	—	—
		头冠状弧	—	—
	测量特征	面高	190.70	179.50
		面宽	146.10	140.70
		额宽	101.60	99.90
		口宽	49.90	46.90
		鼻高	59.40	55.60
		鼻宽	37.30	34.60
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	86.41	86.91
		面指数(面高/面宽×100)	88.77	86.45
		鼻指数(鼻宽/鼻高×100)	63.12	63.16
观察特征	面部	多卵圆型		
	眼部	蒙古褶明显，上眼睑皱褶发育好		
	鼻部	男性属高鼻型和中鼻型， 女性属中鼻型和窄鼻型		
	耳部	男性耳垂多圆形，女性耳垂 多三角形，多有达尔文结节		
	口部	正唇		
	毛发	黑发，直形		
体 部 特 征	测量特征	身高	1643.40	1547.30
		肩宽	—	—
		上肢全长	—	—
		上臂长	—	—
		手长	—	—
		腰围	—	—
		骨盆宽	—	—
			下肢全长	—
	大腿长		—	—
	小腿长		—	—
	足长		—	—
	足高		—	—
	体重(kg)		—	—
	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
手长宽指数		—	—	
足长宽指数		—	—	
上肢全长/身高		—	—	
上下肢长度指数		—	—	

白 族

身材中等，肤色较浅，深褐色眼，黑色直形发。男性多中头型，女性多圆头型。面型属中面型。眼裂开度中等，眼裂斜度外高内低，蒙古褶微显。鼻根高度中等，鼻梁较直，鼻尖方向水平，鼻基部略上翘，属狭鼻型。耳垂多方形或三角形。红唇中等厚。

体质特征参数			
参 数		男(mm)	女(mm)
头 面 部 特 征	头长	191.83	180.72
	头宽	151.79	144.67
	头围	—	—
	头矢状弧	—	—
	头冠状弧	—	—
	面高	185.43	184.59
	面宽	141.93	135.38
	额宽	107.29	105.82
	口宽	51.33	49.01
	鼻高	51.82	49.88
	鼻宽	35.74	32.66
	鼻长	—	—
	指数特征	头指数(头宽/头长×100)	79.12
面指数(面高/面宽×100)		83.77	84.49
鼻指数(鼻宽/鼻高×100)		69.32	65.92
观察特征	面部	中面型	
	眼部	蒙古褶微显	
	鼻部	狭鼻型	
	耳部	耳垂多方形或三角形	
	口部	红唇中等厚	
	毛发	黑色直发	

续表

参 数		男(mm)	女(mm)
体 部 特 征	身高	1637.90	1532.80
	肩宽	—	—
	上肢全长	—	—
	上臂长	—	—
	手长	—	—
	测量特征		
	腰围	773.30	746.80
	骨盆宽	—	—
	下肢全长	—	—
	大腿长	—	—
	小腿长	—	—
	足长	—	—
	足高	—	—
	体重(kg)	54.09	49.71
	指数特征		
身高坐高指数(坐高/身高×100)	—	—	
身高肩宽指数(肩宽/身高×100)	—	—	
手长宽指数	—	—	
足长宽指数	—	—	
上肢全长/身高	—	—	
上下肢长度指数	—	—	
哈 尼 族			

身材较矮小，肤色较浅，黑褐色眼，黑色直形发，眉毛稀少。头型属中头型偏圆头型。面型属阔面型。眼裂开度中等，眼裂斜度外高内低，有蒙古褶，大多有上眼睑皱褶。鼻根高度中等，鼻梁较直，鼻尖和鼻基底方向水平，鼻翼高度中等，鼻孔多为椭圆形，属阔鼻型。耳垂多圆形。红唇较薄，属凸唇型。

第一章 中国少数民族简况及其体质特征

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	185.00	178.27	
	头宽	145.56	140.39	
	头围	—	—	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	181.63	172.39
		面宽	139.68	133.68
		额宽	113.83	111.54
		口宽	53.30	50.24
		鼻高	51.82	48.02
		鼻宽	37.38	34.33
		鼻长	43.84	40.06
	指数特征	头指数(头宽/头长×100)	78.80	78.89
		面指数(面高/面宽×100)	83.32	81.47
		鼻指数(鼻宽/鼻高×100)	72.43	71.77
观察特征	面部	中面型		
	眼部	蒙古褶微显		
	鼻部	狭鼻型		
	耳部	耳垂多方形或三角形		
	口部	红唇中等厚		
	毛发	黑色直发		
体 部 特 征	测量特征	身高	1594.10	1497.80
		肩宽	329.96	304.62
		上肢全长	704.12	658.11
		上臂长	285.71	263.80
		手长	179.98	173.05
		腰围	—	—
		骨盆宽	265.93	269.40
	指数特征	下肢全长	839.45	793.79
		大腿长	400.82	384.42
		小腿长	345.80	322.78
		足长	235.07	219.11
		足高	69.83	63.63
		体重(kg)	52.71	49.63
		身高坐高指数(坐高/身高×100)	53.05	53.01
	观察特征	身高肩宽指数(肩宽/身高×100)	—	—
手长宽指数		—	—	
足长宽指数		—	—	
上肢全长/身高		—	—	
上下肢长度指数		—	—	

哈萨克族

身材较高大，黑色直形发，眉毛中等。头型属特圆头型、高头型和阔头型。面型属阔面型和中面型。眼裂开度中等，眼裂斜度内外平行，蒙古褶微显，上眼睑皱褶发育好。鼻翼高度和突度中等，鼻尖和鼻基部方向向前，鼻根较高，属狭鼻型和中鼻型。耳垂多圆形。上唇高度中等，红唇中等厚。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	185.40	177.00
		头宽	162.20	153.60
		头围	575.60	558.30
		头矢状弧	344.40	324.90
		头冠状弧	361.00	343.70
	面高	191.80	184.40	
	面宽	150.40	141.50	
	额宽	125.20	121.30	
	口宽	50.30	47.10	
	鼻高	56.20	55.10	
	鼻宽	35.50	32.60	
	鼻长	48.80	47.10	
	指数特征	头指数(头宽/头长×100)	87.19	86.90
面指数(面高/面宽×100)		83.49	84.95	
鼻指数(鼻宽/鼻高×100)		62.99	59.64	
观察特征	面部	阔面型或中面型		
	眼部	蒙古褶微显，上眼睑皱褶发育好		
	鼻部	狭鼻型和中鼻型		
	耳部	耳垂多圆形		
	口部	红唇中等厚		
	毛发	黑色直发		

续表

参 数		男(mm)	女(mm)
体 部 特 征	身高	1692.90	1562.00
	肩宽	377.60	345.60
	上肢全长	746.30	694.40
	上臂长	318.20	295.90
	手长	190.50	180.30
	腰围	773.30	746.80
	骨盆宽	276.80	274.50
	下肢全长	903.60	843.20
	大腿长	471.10	424.40
	小腿长	364.00	357.10
	足长	251.00	231.90
	足高	71.40	65.00
	体重(kg)	62.98	54.69
	身高坐高指数(坐高/身高×100)	53.09	53.26
	身高肩宽指数(肩宽/身高×100)	22.28	22.18
	手长宽指数	45.22	43.79
	足长宽指数	40.00	40.78
上肢全长/身高	44.13	44.43	
上下肢长度指数	82.55	82.55	
黎 族			

身材亚中等，黑发直形，发旋多为单旋顺时针方向。头型多中头型偏圆头型、高头型和狭头型。面型男性为中面型，女性多阔面型。浅褐色眼，眼裂开度中等偏宽，眼裂斜度外高内低，多缺失蒙古褶，上眼睑皱褶发育好。鼻根中等偏低，鼻梁直凸型，鼻尖和鼻基底上翘，属中鼻型。耳垂多圆形，多数无达尔文结节。红唇中等偏厚，唇型前突，属突唇型。

中国少数民族遗传学概论

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	测量特征	头长	183.27	174.55
		头宽	145.94	140.12
		头围	546.25	536.00
		头矢状弧	350.25	340.80
		头冠状弧	342.95	335.40
	指数特征	面高	186.56	178.70
		面宽	139.66	133.41
		额宽	101.86	99.76
		口宽	50.61	47.90
		鼻高	52.45	48.24
		鼻宽	39.51	35.92
		鼻长	—	—
	观察特征	面部	中面型、阔面型	
		眼部	多无蒙古褶，上眼睑皱褶发育好	
		鼻部	鼻梁直凸，鼻尖上翘	
		耳部	圆形耳垂，多无达尔文结节	
		口部	多凸唇	
	毛发	黑发直形		
体 部 特 征	测量特征	身高	1636.73	1536.35
		肩宽	382.08	346.69
		上肢全长	734.10	684.73
		上臂长	299.95	280.72
		手长	190.72	179.70
		腰围	723.61	764.80
		骨盆宽	268.59	279.42
		下肢全长	882.30	837.04
		大腿长	462.80	441.96
		小腿长	340.90	316.48
	指数特征	足长	246.58	229.20
		足高	76.36	67.54
		体重(kg)	53.00	47.84
		身高坐高指数(坐高/身高×100)	53.30	53.30
		身高肩宽指数(肩宽/身高×100)	23.35	22.57
指数特征	手长宽指数	44.77	43.70	
	足长宽指数	39.40	39.61	
	上肢全长/身高	—	—	
	上下肢长度指数	—	—	

傣族

身材亚中等偏矮。头型属于中头型。皮肤颜色为棕黄色。男性多为中面型，女性多为阔面型。发直而黑，男性眉毛较浓，女性眉毛中等。眼色深褐，眼裂开度中等，眼裂斜度外高内低，蒙古褶微显。鼻根高度中等，鼻梁平直，鼻尖水平，鼻孔男多三角形，女多卵圆。耳垂呈圆形。多为凸唇型，下颌直型。红唇中等厚。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	189.19	180.80
		头宽	149.28	145.01
		头围	—	—
		头矢状弧	—	—
		头冠状弧	—	—
		面高	188.63	181.79
		面宽	138.42	133.56
		额宽	106.90	104.74
		口宽	53.86	50.72
		鼻高	49.88	47.11
	鼻宽	36.79	34.25	
	鼻长	—	—	
	指数特征	头指数(头宽/头长×100)	78.89	80.25
面指数(面高/面宽×100)		86.20	83.32	
鼻指数(鼻宽/鼻高×100)		74.10	73.03	
观察特征	面部	中—阔面型		
	眼部	蒙古褶微显		
	鼻部	多直鼻梁，鼻孔男多三角形， 女多卵圆		
	耳部	圆形或方形耳垂		
	口部	凸唇		
	毛发	直发，眉毛密度中等偏密		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1617.44	1522.19	
	肩宽	—	—	
	上肢全长	—	—	
	上臂长	—	—	
	手长	—	—	
	测量特征	腰围	—	—
		骨盆宽	—	—
		下肢全长	—	—
		大腿长	—	—
		小腿长	—	—
		足长	—	—
		足高	—	—
	体重(kg)	54.33	49.77	
	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
		足长宽指数	—	—
上肢全长/身高		—	—	
上下肢长度指数		—	—	

畲 族

身材中等偏矮，肤色浅黄，发色黑褐。头型多中头型和高头型。眼裂斜度外高内低，无蒙古褶。鼻梁较直，鼻翼不突，属于狭鼻型。耳垂多圆形或方形。唇薄。胸部发育好，腿长。骨盆向外突出明显。

第一章 中国少数民族简况及其体质特征

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征		头长	187.17	177.86
		头宽	146.35	139.97
		头围	557.07	536.86
		头矢状弧	—	—
	测量特征	头冠状弧	—	—
		面高	188.87	178.90
		面宽	137.95	131.28
		额宽	105.11	102.08
		口宽	52.21	49.11
		鼻高	55.53	52.71
		鼻宽	37.75	34.84
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	78.26	79.04
		面指数(面高/面宽×100)	90.62	89.95
鼻指数(鼻宽/鼻高×100)		68.25	66.69	
观察特征	面部	中头型或高头型		
	眼部	多无蒙古褶		
	鼻部	狭鼻型		
	耳部	耳垂多圆形或方形		
	口部	唇薄		
	毛发	黑褐色		
体 部 特 征		身高	1632.23	1527.25
		肩宽	375.34	340.94
		上肢全长	713.59	655.47
		上臂长	—	—
		手长	—	—
	测量特征	腰围	—	—
		骨盆宽	277.15	278.30
		下肢全长	856.34	811.45
		大腿长	—	—
		小腿长	—	—
		足长	—	—
		足高	—	—
	体重(kg)	56.18	49.66	

续表

		参 数	男(mm)	女(mm)
体 部 特 征	指数特征	身高坐高指数(坐高/身高×100)	53.38	53.05
		身高肩宽指数(肩宽/身高×100)	23.14	22.33
		手长宽指数	—	—
		足长宽指数	—	—
		上肢全长/身高	43.98	42.92
		上下肢长度指数	83.42	80.82

倮 倮 族

身材矮小。眼裂开度多为中等偏窄，眼裂斜度内外平行，大多数人无蒙古褶，上眼睑皱褶发育较好。鼻梁多为直型，鼻根较高，鼻尖和鼻基底方向大多为水平向前，鼻孔形状为卵圆形和椭圆形，鼻翼微突，大多属狭鼻型。上唇皮肤部多数较直，为正唇型，红唇中等厚，多数人耳壳无达尔文结节，眉脊微显，颧骨多数较突出，整个面部扁平而宽，大多呈椭圆形或卵圆形，脸型属狭脸型，过狭脸型，头型大多属于中头型、高头型和狭头型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	186.37	180.10
		头宽	149.33	143.44
		头围	553.47	542.67
		头矢状弧	—	—
		头冠状弧	—	—
		面高	126.72	116.56
		面宽	140.33	132.76

第一章 中国少数民族简况及其体质特征

续表

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	额宽	103.26	98.55
		口宽	54.16	51.94
		鼻高	57.45	52.76
		鼻宽	38.56	35.18
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	80.18	79.67
		面指数(面高/面宽×100)	90.32	87.81
		鼻指数(鼻宽/鼻高×100)	67.34	66.92
	观察特征	面部	中头型、高头型和狭头型	
		眼部	无蒙古褶, 上眼睑皱褶发育好	
鼻部		狭鼻型		
耳部		无达尔文结节		
口部		正唇型		
	毛发	黑发直形		
体 部 特 征	测量特征	身高	1586.68	1488.86
		肩宽	—	—
		上肢全长	—	—
		上臂长	—	—
		手长	—	—
		腰围	—	—
		骨盆宽	277.03	282.64
		下肢全长	—	—
		大腿长	—	—
		小腿长	—	—
	指数特征	足长	—	—
		足高	—	—
		体重(kg)	51.96	45.78
指数特征	身高坐高指数(坐高/身高×100)	—	—	
	身高肩宽指数(肩宽/身高×100)	—	—	
	手长宽指数	—	—	
	足长宽指数	—	—	
	上肢全长/身高	—	—	
	上下肢长度指数	—	—	

仡佬族

身材亚中等偏矮。眼裂斜度男性外高内低，女性内外平行，多数人有蒙古褶。男性鼻梁多为直型和波浪型，女性鼻梁多为凹型和直型，男性鼻根中等高，女性鼻根多为低平，鼻基底呈水平向前，鼻翼甚突，属中鼻型。大多数耳壳无达尔文结节。上唇皮肤高度中等，多为正唇型。男性红唇中等厚，女性红唇中等偏薄。额部直立和微斜，面部男性狭窄多为过狭面型，女性以狭面型为主，头型多属圆头型、高头型和狭头型。

体质特征参数

参 数		男(mm)	女(mm)
头 面 部 特 征	头长	186.80	179.70
	头宽	148.00	142.60
	头围	553.30	540.10
	头矢状弧	319.80	310.70
	头冠状弧	345.70	335.80
	面高	121.70	114.40
	面宽	140.60	135.00
	额宽	102.20	104.00
	口宽	54.70	49.70
	鼻高	54.80	52.30
	鼻宽	37.30	34.70
	鼻长	—	—
	指数特征	头指数(头宽/头长×100)	79.26
	面指数(面高/面宽×100)	86.84	84.08
	鼻指数(鼻宽/鼻高×100)	68.45	67.04
观察特征	面部	过狭面型、狭面型	
	眼部	多数有蒙古褶	
	鼻部	中鼻型	
	耳部	耳壳无达尔文结节	
	口部	正唇型	
	毛发	黑发直形	

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1619.00	1497.90	
	肩宽	372.00	344.70	
	上肢全长	735.00	671.90	
	上臂长	298.00	275.00	
	手长	188.10	173.90	
	腰围	699.50	700.00	
	测量特征	骨盆宽	273.90	282.70
	下肢全长	865.50	808.80	
	大腿长	437.20	418.50	
	小腿长	364.90	333.50	
	足长	248.20	226.00	
	足高	63.30	56.40	
	体重(kg)	53.20	48.00	
	指 数 特 征	身高坐高指数(坐高/身高×100)	53.68	53.62
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	45.19	44.14
足长宽指数		40.96	40.71	
上肢全长/身高		45.41	44.86	
上下肢长度指数		—	—	

东 乡 族

身材中等，头长而窄，面型以过狭面型和狭面型为主。鼻型为狭鼻型，头型为高头型、狭头型和中头型。

体质特征参数				
参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	188.60	180.69	
	头宽	146.92	139.39	
	头围	558.82	—	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	128.66	120.80
		面宽	140.14	132.90
		额宽	102.30	100.34
		口宽	52.08	49.31
		鼻高	56.36	52.10
	指数特征	鼻宽	35.95	32.14
		鼻长	—	—
		头指数(头宽/头长×100)	77.95	77.20
指数特征	面指数(面高/面宽×100)	91.89	90.96	
	鼻指数(鼻宽/鼻高×100)	64.60	61.98	
观察特征	面部	过狭面型和狭面型		
	眼部	—		
	鼻部	狭鼻型		
	耳部	—		
	口部	—		
	毛发	—		

续表

		参 数	男(mm)	女(mm)	
体 部 特 征		身高	1667.44	1542.37	
		肩宽	372.95	339.83	
		上肢全长	722.95	666.37	
		上臂长	292.55	273.68	
		手长	183.46	172.09	
	测量特征	腰围	729.82	720.84	
		骨盆宽	277.23	278.65	
		下肢全长	886.96	820.05	
		大腿长	445.92	411.43	
		小腿长	372.01	347.35	
		足长	250.34	222.93	
		足高	65.47	58.17	
		体重(kg)	58.41	50.16	
		指数特征	身高坐高指数(坐高/身高×100)	53.59	54.06
			身高肩宽指数(肩宽/身高×100)	22.37	22.04
	手长宽指数		46.73	45.22	
	足长宽指数		41.34	39.81	
	上肢全长/身高		43.37	43.21	
			上下肢长度指数	81.56	81.32

拉 祜 族

身材矮小，皮肤浅黄褐色，发黑直，体毛和胡须极少。眼裂开度中等，眼裂斜度内外平行，上眼睑褶皱发育好，多数人有蒙古褶。鼻梁凹型，鼻根中等高，鼻尖上翘，鼻翼高度中等，鼻孔形状呈卵圆形，属狭鼻型。上唇皮肤平直，多为正唇型。耳壳大多缺失达尔文结节，耳垂圆形。眉脊较突出，颧骨突出，面部扁平，多属狭面型，头型属圆头型、高头型和中头型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	184.18	176.88
		头宽	148.70	142.79
		头围	542.18	523.94
		头矢状弧	—	—
		头冠状弧	—	—
		面高	124.21	116.90

续表

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	面宽	140.63	131.53
		额宽	101.35	98.49
		口宽	53.67	51.25
		鼻高	56.72	52.42
		鼻宽	38.60	35.30
	指数特征	鼻长	—	—
		头指数(头宽/头长×100)	80.81	80.80
		面指数(面高/面宽×100)	88.40	88.97
	观察特征	鼻指数(鼻宽/鼻高×100)	68.28	67.63
		面部	狭脸型	
眼部		多数人有蒙古褶		
鼻部		狭鼻型		
耳部		耳壳大多缺失达尔文结节, 耳垂圆形		
体 部 特 征	测量特征	口部	正唇型	
		毛发	发黑直, 体毛和胡须极少	
		身高	1575.96	1472.56
		肩宽	359.63	322.63
		上肢全长	708.32	644.08
		上臂长	—	—
		手长	—	—
		腰围	—	—
		骨盆宽	264.36	264.60
		下肢全长	796.52	746.38
	指数特征	大腿长	—	—
		小腿长	—	—
		足长	—	—
		足高	—	—
		体重(kg)	50.13	42.45
指数特征	身高坐高指数(坐高/身高×100)	—	—	
	身高肩宽指数(肩宽/身高×100)	—	—	
	手长宽指数	—	—	
	足长宽指数	—	—	
	上肢全长/身高	—	—	
		上下肢长度指数	—	—

水 族

身材亚中等偏矮，男性胡须少，黑发直形。头型多过圆头型、高头型和狭头型。面型多中面型。眼裂开度中等，眼裂斜度内外平行，40%的人缺失蒙古褶，上眼睑皱褶发育好。男性鼻根中等，女性鼻根多低平。男性鼻基底水平，女性上翘，鼻翼突度明显，多属阔鼻型。耳垂多方形或三角形，耳壳缺失达尔文结节。男性为正唇型，女性为凸唇型，红唇厚度中等。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	183.44	176.33
		头宽	151.72	144.60
		头围	550.51	529.05
		头矢状弧	326.86	320.92
		头冠状弧	354.27	338.76
		面高	189.24	179.37
		面宽	142.05	135.20
		额宽	102.74	100.14
		口宽	55.00	52.63
		鼻高	52.44	49.52
	鼻宽	38.39	35.69	
		鼻长	—	—
	指数特征		头指数(头宽/头长×100)	82.79
		面指数(面高/面宽×100)	84.11	82.85
		鼻指数(鼻宽/鼻高×100)	73.61	72.53
观察特征		面部	中面型	
		眼部	上眼睑皱褶发育好	
		鼻部	阔鼻型	
		耳部	耳垂多方形或三角形	
		口部	正唇型	凸唇型
		毛发	黑发直形，少胡须	

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1601.38	1477.23	
	肩宽	371.25	334.64	
	上肢全长	713.71	655.31	
	上臂长	285.46	264.10	
	手长	185.10	171.27	
	测量特征	腰围	699.49	678.52
	骨盆宽	275.24	278.28	
	下肢全长	854.27	796.02	
	大腿长	427.88	413.75	
	小腿长	367.12	327.15	
	足长	244.25	226.19	
	足高	61.18	55.24	
	体重(kg)	52.08	44.59	
	指数特征	身高坐高指数(坐高/身高×100)	53.26	53.22
	身高肩宽指数(肩宽/身高×100)	—	—	
	手长宽指数	45.83	44.99	
足长宽指数	42.98	42.86		
上肢全长/身高	44.57	44.36		
上下肢长度指数	—	—		
佤 族				

身材亚中等，肤色黄。发黑直型，头型属圆头型、高头型和阔头型。男女均性属长躯干型。男性多宽肩型，中骨盆型。女性多中肩型、窄骨盆型。面型属狭面型。眼色多褐色，眼裂斜度外高内低，大多有上眼睑皱褶。鼻根高度中等，鼻梁多直型，属狭鼻型。耳垂多圆形。

第一章 中国少数民族简况及其体质特征

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	184.00	175.40	
	头宽	152.40	146.60	
	头围	556.50	537.20	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	127.40	116.50
		面宽	142.30	134.80
		额宽	106.00	103.80
		口宽	52.80	50.00
		鼻高	56.30	52.20
		鼻宽	38.80	35.80
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	82.90	83.70
		面指数(面高/面宽×100)	89.60	86.50
		鼻指数(鼻宽/鼻高×100)	63.90	69.00
	观察特征	面部	多狭面型	
眼部		眼色褐色, 多有上眼睑皱褶		
鼻部		狭鼻型		
耳部		耳垂圆形		
口部		红唇中等厚		
体 部 特 征	毛发	黑色直型		
	测量特征	身高	1604.00	1507.00
		肩宽	337.10	331.90
		上肢全长	727.50	673.70
		上臂长	310.40	285.60
		手长	—	—
		腰围	852.10	808.40
		骨盆宽	266.10	261.00
		下肢全长	862.90	822.90
		大腿长	407.10	391.50
		小腿长	377.50	375.00
		足长	—	—
		足高	—	—
		体重(kg)	54.90	49.10
	指数特征	身高坐高指数(坐高/身高×100)	53.80	53.90
身高肩宽指数(肩宽/身高×100)		23.20	22.00	
手长宽指数		—	—	
足长宽指数		—	—	
上肢全长/身高		45.40	44.70	
上下肢长度指数		84.30	81.90	

纳西族

身材中等。眼裂开度中等，眼裂斜度外高内低，上眼睑皱褶发育较好，大多无蒙古褶。鼻梁多为直型，次为凹形，鼻根较高，鼻尖和鼻基底方向多呈水平向前，鼻孔形状多为椭圆形，次为卵圆形，鼻翼微突，属狭鼻型。上唇皮肤多数较直，属正唇型。红唇中等厚。多数人耳壳无达尔文结节，耳垂多为方形，次为圆形。眉峰突度，男性多为中显，女性多为微显。颧骨较突出，整个面部较扁平而宽，大多呈椭圆形或卵圆形。男性多属中面型，女性多属狭面型。头型多属中头型、高头型和狭头型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	189.18	180.91
		头宽	150.15	146.44
		头围	564.80	550.76
		头矢状弧	—	—
		头冠状弧	—	—
		面高	124.97	118.13
		面宽	141.18	135.82
		额宽	103.54	100.33
		口宽	53.35	50.17
		鼻高	56.90	53.81
		鼻宽	38.57	35.03
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	79.50	81.10
		面指数(面高/面宽×100)	—	—
鼻指数(鼻宽/鼻高×100)		67.97	65.37	
观察特征	面部	中面型	狭面型	
	眼部	上眼睑皱褶发育较好,大多无蒙古褶		
	鼻部	狭鼻型		
	耳部	耳垂多为方形,耳壳无达尔文结节		
	口部	正唇型		
	毛发	—		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1659.23	1554.30	
	肩宽	372.89	343.23	
	上肢全长	737.62	681.53	
	上臂长	—	—	
	手长	—	—	
	腰围	—	—	
	测量特征	骨盆宽	287.99	293.60
		下肢全长	915.14	871.90
		大腿长	—	—
		小腿长	—	—
		足长	—	—
		足高	—	—
		体重(kg)	58.20	52.80
		身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	22.47	22.08
指 数 特 征	手长宽指数	—	—	
	足长宽指数	—	—	
	上肢全长/身高	44.46	43.85	
	上下肢长度指数	—	—	
羌 族				
身材亚中等偏矮, 体型偏狭长, 腿较长, 面部狭窄, 鼻狭而高。				
体质特征参数				
参 数		男(mm)	女(mm)	
头 面 部 特 征	测量特征	头长	188.75	181.81
		头宽	149.73	143.45
		头围	—	—
		头矢状弧	—	—
		头冠状弧	—	—
		面高	192.01	186.81

续表

参 数		男(mm)	女(mm)	
头 面 部 特 征	测量特征	面宽	142.15	136.26
		额宽	104.02	103.26
		口宽	49.69	48.38
	指数特征	鼻高	57.62	53.83
		鼻宽	36.61	34.03
		鼻长	—	—
		头指数(头宽/头长×100)	79.40	78.96
		面指数(面高/面宽×100)	88.76	88.19
		鼻指数(鼻宽/鼻高×100)	63.91	63.50
		观察特征	面部	狭面型
眼部	—			
鼻部	狭鼻型			
耳部	—			
口部	—			
毛发	—			
体 部 特 征	测量特征	身高	1610.70	1511.80
		肩宽	335.59	329.31
		上肢全长	721.41	669.09
		上臂长	—	—
		手长	176.43	166.78
		腰围	—	—
		骨盆宽	252.50	248.65
		下肢全长	872.31	828.18
		大腿长	—	—
		小腿长	—	—
	足长	235.45	219.92	
	足高	—	—	
	体重(kg)	55.68	52.01	

续表

		参 数	男(mm)	女(mm)
体 部 特 征	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
		足长宽指数	—	—
		上肢全长/身高	—	—
		上下肢长度指数	—	—
土 族				
<p>身材中等偏矮。男性眼裂开度大多为中等，女性眼裂开度大多为较宽，上眼睑皱褶发育良好，约半数人有蒙古褶。鼻根高度中等，直鼻梁较多，鼻基部方向水平，鼻孔以圆形和卵圆形为主。耳廓紧贴颅部者较多，男性耳壳达尔文结节出现率多于女性，男性耳垂多为方形，女性耳垂多为三角形。面型男性以过狭面、狭面型和中面型为主，女性则为狭面型和中面型为主。头型属于中头型、高头型和狭头型。</p>				
体质特征参数				
		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	190.32	182.68
		头宽	147.11	143.11
		头围	555.31	547.14
		头矢状弧	—	—
		头冠状弧	—	—
		面高	124.64	113.66
		面宽	139.04	134.12
		额宽	109.21	109.42
		口宽	48.5	45.76
		鼻高	51.30	46.44
		鼻宽	36.58	32.88
		鼻长	—	—

续表

		参 数	男(mm)	女(mm)
头	指数特征	头指数(头宽/头长×100)	77.35	78.39
		面指数(面高/面宽×100)	89.68	84.80
		鼻指数(鼻宽/鼻高×100)	71.63	71.09
面部特征	观察特征	面部	过狭面、狭面型和中面型 狭面型和中面型	
		眼部	约半数人有蒙古褶	
		鼻部	直鼻梁较多, 鼻根高度中等	
		耳部	耳壳有达尔文结节	
		口部	—	
		毛发	—	
体部特征	测量特征	身高	1634.99	1543.97
		肩宽	369.06	337.75
		上肢全长	723.61	678.76
		上臂长	292.24	273.98
		手长	183.12	173.65
		腰围	760.70	748.17
		骨盆宽	278.51	275.85
		下肢全长	872.21	836.23
		大腿长	426.34	409.53
		小腿长	385.60	366.80
		足长	238.90	222.09
		足高	62.53	57.24
			体重(kg)	55.49
指数特征	指数特征	身高坐高指数(坐高/身高×100)	54.74	54.87
		身高肩宽指数(肩宽/身高×100)	22.58	21.89
		手长宽指数	45.27	43.50
		足长宽指数	37.72	37.23
		上肢全长/身高	44.25	43.97
		上下肢长度指数	82.98	81.22

佤 佬 族

身高中等偏矮，男性多为黄色肤色，女性多为浅黄色肤色。发色多为黑色。眼色多为黑褐色，上眼睑皱褶出现率较高，多有蒙古褶。鼻根高度男低型率与中等型率接近，女多为低型。鼻翼高度多为中等。男以方型耳垂多见，女以圆型耳垂多见。佤佬族男女均为中头型、高头型、狭头型、中鼻型、中腿型、中肩型、中骨盆型。男为狭面型、长躯干型、窄胸型。女为中面型、中躯干型、中胸型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	189.50	179.90
		头宽	148.50	144.50
		头围	554.40	535.00
		头矢状弧	—	—
		头冠状弧	—	—
		面高	123.50	115.90
	测量特征	面宽	139.0	133.20
		额宽	102.20	99.00
		口宽	51.60	48.80
		鼻高	53.20	49.80
	指数特征	鼻宽	39.10	36.10
		鼻长	—	—
		头指数(头宽/头长×100)	78.40	80.40
	观察特征	面指数(面高/面宽×100)	89.10	87.20
		鼻指数(鼻宽/鼻高×100)	74.30	73.20
		面部	狭面型	中面型
眼部		多为黑褐色		
鼻部		中鼻型		
耳部		多为方型	多为圆型	
	口部	上唇皮肤高度以中等多见		
	毛发	多为黑色		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1629.80	1514.30	
	肩宽	375.40	337.00	
	上肢全长	740.00	677.80	
	上臂长	323.50	292.70	
	手长	—	—	
	腰围	—	—	
	测量特征	骨盆宽	274.40	271.10
		下肢全长	875.40	819.70
		大腿长	810.50	757.20
		小腿长	362.10	341.00
		足长	—	—
		足高	—	—
		体重(kg)	55.40	47.70
		身高坐高指数(坐高/身高×100)	53.50	53.70
		身高肩宽指数(肩宽/身高×100)	23.00	22.20
指 数 特 征	手长宽指数	—	—	
	足长宽指数	—	—	
	上肢全长/身高	45.40	44.80	
	上下肢长度指数	84.50	82.70	

锡 伯 族

身材中等偏高。黑色直发，发质较硬，男性胡须中等。眼裂斜度多为内外平行，眼裂开度较窄，大多有蒙古褶。鼻梁直，鼻宽大于两眼内角宽，鼻根高度中等，鼻尖和鼻基部方向均为水平，鼻翼高度中等，微突，属狭鼻型。面型男性多为五角形，女性多为卵圆形，属于狭面型。上唇皮肤高度中等，红唇中等偏薄。耳垂多为三角形、圆形和方形。头型属于圆头型或过圆头型、高头型和阔头型。体型属于中等偏宽短型，女性骨盆较宽。

第一章 中国少数民族简况及其体质特征

体质特征参数

		参 数	男(mm)	女(mm)	
头面部特征		头长	184.45	175.34	
		头宽	160.58	153.20	
		头围	566.60	550.78	
		头矢状弧	347.79	343.14	
	测量特征	头冠状弧	365.56	358.21	
		面高	133.27	121.71	
		面宽	147.68	140.73	
		额宽	119.79	113.02	
		口宽	52.90	48.70	
		鼻高	58.40	53.81	
		鼻宽	38.48	35.43	
		鼻长	52.14	47.31	
	指数特征	头指数(头宽/头长×100)	87.24	87.46	
		面指数(面高/面宽×100)	90.67	86.65	
		鼻指数(鼻宽/鼻高×100)	—	—	
	观察特征	面部	狭面型		
		眼部	大多有蒙古褶		
鼻部		狭鼻型			
耳部		耳垂多为三角形、圆形和方形			
口部		红唇中等偏薄			
毛发		黑色直发, 发质较硬, 男性胡须中等			
体部特征		身高	1697.30	1584.46	
		肩宽	395.76	361.01	
		上肢全长	747.82	700.83	
		上臂长	—	—	
	测量特征	手长	184.79	172.05	
		腰围	806.35	766.52	
		骨盆宽	291.99	286.55	
		下肢全长	909.30	848.16	
		大腿长	458.49	448.21	
		小腿长	374.00	325.24	
		足长	246.41	225.94	
足高	76.83	69.62			
体重(kg)	66.74	57.25			

续表

		参 数	男(mm)	女(mm)
体 部 特 征	指数特征	身高坐高指数(坐高/身高×100)	54.94	53.75
		身高肩宽指数(肩宽/身高×100)	23.42	22.81
		手长宽指数	46.51	45.28
		足长宽指数	40.12	39.39
		上肢全长/身高	43.74	44.21
		上下肢长度指数	82.29	83.19

柯尔克孜族

身材中等偏高，黑色直发。约一半的人内眼角有发育不等的蒙古褶，眼裂开度中等，眼裂斜度内外平行，上眼睑皱褶发育好。鼻根较高，鼻翼发育较弱，鼻尖向前，鼻高等中等，鼻宽较窄，属狭鼻型。面部较高而宽，面型属中面型偏阔面型。红唇较薄。耳垂多为圆形。头长、头宽、头高值都较大，整个头形较圆而高，大多属于圆头型、高头型和阔头型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	189.75	181.51
		头宽	160.63	154.50
		头围	577.10	557.20
		头矢状弧	337.20	338.00
		头冠状弧	358.20	349.20
		面高	126.99	120.26
		面宽	149.37	142.67
		额宽	115.90	114.19
		口宽	52.54	48.89
		鼻高	56.88	53.64
		鼻宽	38.38	34.38
鼻长	52.29	47.74		

续表

		参 数	男(mm)	女(mm)	
头 面 部 特 征	指数特征	头指数(头宽/头长×100)	84.65	85.16	
		面指数(面高/面宽×100)	85.12	84.29	
		鼻指数(鼻宽/鼻高×100)	68.09	64.64	
			面部	中面型偏阔面型	
			眼部	内眼角有发育不等的蒙古褶	
	观察特征		鼻部	狭鼻型	
			耳部	耳垂多为圆形	
			口部	红唇较薄	
			毛发	黑色直发	
体 部 特 征		身高	1682.60	1557.20	
		肩宽	374.70	349.50	
		上肢全长	754.72	690.34	
		上臂长	321.70	289.50	
		手长	181.30	173.90	
	测量特征		腰围	799.90	790.40
			骨盆宽	287.30	291.60
			下肢全长	950.50	884.50
			大腿长	—	—
			小腿长	360.70	360.13
			足长	248.54	227.59
		足高	83.67	70.49	
		体重(kg)	64.04	55.39	
指数特征		身高坐高指数(坐高/身高×100)	—	—	
		身高肩宽指数(肩宽/身高×100)	—	—	
		手长宽指数	—	—	
		足长宽指数	—	—	
		上肢全长/身高	—	—	
		上下肢长度指数	—	—	

达斡尔族

身材中等偏高，男性少胡须。头型属圆头型、中头型和高头型。男性属长躯干型，女性属中躯干型。面型男性属中面型，女性属阔面型。眼裂斜度外高内低，多无蒙古褶，大多有上眼睑皱褶。鼻根低平，鼻梁多直型，属狭鼻型。耳垂多三角形，红唇较薄。

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	测量特征	头长	186.50	178.40
		头宽	154.80	147.90
		头围	558.00	551.00
		头矢状弧	324.40	310.20
		头冠状弧	346.20	338.90
		面高	195.80	183.20
		面宽	143.70	138.20
		额宽	116.30	107.00
		口宽	50.60	49.80
		鼻高	50.40	45.30
	鼻宽	36.20	31.80	
	鼻长	—	—	
	指数特征	头指数 (头宽/头长×100)	83.08	82.99
		面指数 (面高/面宽×100)	84.50	81.10
鼻指数 (鼻宽/鼻高×100)		71.80	70.53	
观察特征	面部	中面型	阔面型	
	眼部	多无蒙古褶，有上眼睑皱褶		
	鼻部	狭鼻型		
	耳部	耳垂多三角形		
	口部	红唇较薄		
	毛发	黑色直发		
体 部 特 征	测量特征	身高	1694.00	1571.00
		肩宽	382.00	347.00
		上肢全长	750.00	686.00
		上臂长	308.90	283.90
		手长	183.00	168.10
		腰围	774.00	716.00
		骨盆宽	287.00	281.00
		下肢全长	937.00	869.00
大腿长	458.80	434.20		

续表

		参 数	男(mm)	女(mm)
体 部 特 征	测量特征	小腿长	365.10	335.90
		足长	249.10	229.10
		足高	70.80	64.70
	指数特征	体重(kg)	64.00	53.50
		身高坐高指数(坐高/身高×100)	53.49	53.81
		身高肩宽指数(肩宽/身高×100)	22.56	22.07
		手长宽指数	57.47	55.80
		足长宽指数	38.38	39.15
	上肢全长/身高	44.29	43.70	
	上下肢长度指数	—	—	

景 颇 族

身材亚中等偏矮。体毛和胡须稀少，皮肤呈浅黄褐色，黑发，平直而硬。眼为深褐色，眼裂开度较宽，眼裂斜度内外平行，大多有蒙古褶。鼻根中等偏高，鼻翼微突，鼻梁直型，鼻尖向前，多属中鼻型。耳壳缺失，无达尔文结节。嘴型稍突，中等唇厚。颧骨突出，面部扁平，多属中面型和狭面型。头型多属中头型、高头型和狭头型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	189.50	181.71
		头宽	148.16	143.41
		头围	562.86	552.34
		头矢状弧	327.43	315.03
		头冠状弧	359.67	350.48
		面高	123.74	114.53

续表

		参 数	男(mm)	女(mm)	
头 面 部 特 征	测量特征	面宽	139.53	133.37	
		额宽	104.26	101.85	
		口宽	51.73	48.94	
		鼻高	51.87	48.16	
		鼻宽	37.78	34.71	
			鼻长	—	—
	指数特征	头指数(头宽/头长×100)	78.26	79.00	
		面指数(面高/面宽×100)	88.78	85.96	
		鼻指数(鼻宽/鼻高×100)	75.61	73.41	
	观察特征	面部	中面型和狭面型		
眼部		深褐色, 大多有蒙古褶			
鼻部		中鼻型			
耳部		耳壳缺失, 无达尔文结节			
口部		中等唇厚			
	毛发	黑发, 平直而硬			
体 部 特 征	测量特征	身高	1607.30	1499.96	
		肩宽	—	—	
		上肢全长	—	—	
		上臂长	—	—	
		手长	—	—	
		腰围	—	—	
		骨盆宽	—	—	
		下肢全长	—	—	
		大腿长	—	—	
		小腿长	—	—	
	指数特征	足长	—	—	
		足高	—	—	
		体重(kg)	55.12	47.43	
		身高坐高指数(坐高/身高×100)	—	—	
		身高肩宽指数(肩宽/身高×100)	—	—	
	手长宽指数	—	—		
	足长宽指数	—	—		
	上肢全长/身高	—	—		
	上下肢长度指数	—	—		

毛南族

眼裂开度中等，方向上斜形，普遍有蒙古褶。直形鼻梁，鼻翼微突。上唇皮肤高度中等，以突唇为多，红唇厚度适中。耳垂以圆形为主。属圆头型，以超阔面型为主。多为中鼻型。男女身高以矮形为主。

体质特征参数				
		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	181.88	171.09
		头宽	147.09	141.63
		头围	550.41	515.36
		头矢状弧	—	—
		头冠状弧	—	—
	指数特征	面高	113.23	102.87
		面宽	130.06	124.07
		额宽	114.77	111.44
		口宽	—	—
		鼻高	46.55	42.79
	观察特征	鼻宽	39.04	36.77
		鼻长	—	—
		头指数(头宽/头长×100)	81.02	82.97
		面指数(面高/面宽×100)	80.83	76.33
鼻指数(鼻宽/鼻高×100)		78.75	81.13	
观察特征	面部	超阔面型		
	眼部	普遍有蒙古褶		
	鼻部	中鼻型		
	耳部	耳垂以圆形为主		
	口部	以突唇为多，红唇厚度适中		
	毛发	—		

续表

参 数		男(mm)	女(mm)
体 部 特 征	身高	1585.31	1493.26
	肩宽	—	—
	上肢全长	—	—
	上臂长	—	—
	手长	—	—
	腰围	—	—
	骨盆宽	—	—
	下肢全长	—	—
	大腿长	—	—
	小腿长	—	—
	足长	—	—
	足高	—	—
	体重(kg)	54.09	48.28
	身高坐高指数(坐高/身高×100)	—	—
	身高肩宽指数(肩宽/身高×100)	—	—
	手长宽指数	—	—
足长宽指数	—	—	
上肢全长/身高	—	—	
上下肢长度指数	—	—	

撒 拉 族

身材中等偏高。头型多高头型、狭头型和中头型。面型为中面型和狭面型。鼻型为狭鼻型。

第一章 中国少数民族简况及其体质特征

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征		头长	190.36	176.08
		头宽	151.94	143.56
		头围	557.41	533.60
		头矢状弧	—	—
		头冠状弧	—	—
	测量特征	面高	190.88	175.71
		面宽	142.70	134.60
		额宽	118.60	116.49
		口宽	47.72	44.15
		鼻高	53.37	49.07
		鼻宽	35.73	31.84
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	80.06	81.69
		面指数(面高/面宽×100)	87.01	84.32
		鼻指数(鼻宽/鼻高×100)	67.22	65.31
观察特征	面部	中面型、狭面型		
	眼部	—		
	鼻部	狭鼻型		
	耳部	—		
	口部	—		
	毛发	—		
体 部 特 征	测量特征	身高	1673.10	1551.70
		肩宽	373.07	334.91
		上肢全长	721.48	663.38
		上臂长	290.03	264.58
		手长	186.98	172.86
		腰围	798.22	754.49
		骨盆宽	282.62	270.12
		下肢全长	935.85	878.79
		大腿长	476.24	453.12
		小腿长	396.59	362.25
	指数特征	足长	244.28	223.90
		足高	64.75	58.15
		体重(kg)	64.22	54.46
		身高坐高指数(坐高/身高×100)	53.98	53.66
	身高肩宽指数(肩宽/身高×100)	22.31	21.59	
	手长宽指数	45.33	43.89	
	足长宽指数	40.22	40.07	
	上肢全长/身高	43.13	42.75	
	上下肢长度指数	77.15	75.57	

布 朗 族

身材矮小。皮肤黄褐色，黑发平直。眉脊发育中等，眉毛中等。眼为褐色到黑褐色，眼裂开度中等，眼裂斜度外高内低，蒙古褶发达，大多有上眼睑皱褶。鼻根中等高，鼻梁多为平直，少数为凹型或凸型，鼻尖和鼻基部多为水平，也有上翘和下垂的，鼻孔多为椭圆形和卵圆形，鼻型属中鼻型。男性耳垂多为圆形、方形和三角形，女性多为方形。面部扁平，下颌微向前突，大多为凸唇形，红唇中等厚。面型多属阔面型或超阔面型，头型属中头型。

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	183.22	176.00	
	头宽	144.23	139.09	
	头围	—	—	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	115.48	108.27
	面宽	138.69	133.85	
	额宽	113.53	109.33	
	口宽	53.33	50.26	
	鼻高	51.30	48.39	
	鼻宽	37.29	34.44	
	鼻长	—	—	
	头指数(头宽/头长×100)	79.02	77.80	
指数特征	面指数(面高/面宽×100)	—	—	
	鼻指数(鼻宽/鼻高×100)	73.01	71.48	
观察特征	面部	多属阔面型或超阔面型		
	眼部	眼为褐色到黑褐色，眼裂开度中等		
	鼻部	中鼻型		
	耳部	耳垂多为圆形、方形和三角形		
	口部	大多为凸唇形，红唇中等厚		
	毛发	黑发平直		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1554.80	1462.20	
	肩宽	329.57	280.20	
	上肢全长	700.03	648.88	
	上臂长	285.74	261.19	
	手长	175.71	169.17	
	测量特征	腰围	—	—
	骨盆宽	261.49	260.94	
	下肢全长	—	—	
	大腿长	—	—	
	小腿长	—	—	
	足长	230.11	214.47	
	足高	65.36	59.57	
	体重(kg)	50.12	45.11	
	指数特征	身高坐高指数(坐高/身高×100)	52.44	52.69
	身高肩宽指数(肩宽/身高×100)	—	—	
	手长宽指数	—	—	
	足长宽指数	—	—	
上肢全长/身高	—	—		
上下肢长度指数	—	—		

塔吉克族

身材中等。黑色波形发，男性胡须多，眉毛浓，再生毛多。眼裂开度较宽，无蒙古褶。鼻根高，鼻尖下垂，鼻孔形状椭圆和卵圆形较多，狭鼻狭面，鼻深度较深。耳垂圆形，红唇中等偏薄。面型多为倒卵圆形和卵圆形。体型中等偏宽短型，女性骨盆较宽，头型属中头型和高头型。

体质特征参数

参 数		男(mm)	女(mm)
头 面 部 特 征	头长	187.79	178.17
	头宽	149.90	143.50
	头围		
	头矢状弧		
	头冠状弧		
	测量特征		
	面高	124.42	113.74
	面宽	139.08	131.33
	额宽	118.96	114.28
	口宽	52.87	47.78
	鼻高	55.07	55.64
	鼻宽	32.24	30.68
	鼻长	55.91	48.64
	指数特征		
	头指数(头宽/头长×100)	79.82	80.54
	面指数(面高/面宽×100)	89.46	86.61
鼻指数(鼻宽/鼻高×100)	58.54	60.58	
面部	狭面, 多为倒卵圆形和卵圆形		
眼部	无蒙古褶		
观察特征	鼻部	狭鼻	
	耳部	耳垂圆形	
	口部	红唇中等偏薄	
	毛发	黑色波形发, 男性胡须多, 眉毛浓, 再生毛多	

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1664.86	1553.44	
	肩宽	371.10	345.13	
	上肢全长	745.92	695.77	
	上臂长	317.72	293.92	
	手长	192.82	179.54	
	测量特征	腰围	773.30	746.80
		骨盆宽	290.34	289.93
		下肢全长	920.93	847.75
		大腿长	—	—
		小腿长	—	—
		足长	249.98	233.36
		足高	66.72	63.28
		体重(kg)	—	—
	指 数 特 征	身高坐高指数(坐高/身高×100)	51.73	52.90
		身高肩宽指数(肩宽/身高×100)	22.29	22.22
		手长宽指数	44.73	43.54
足长宽指数		40.64	40.80	
上肢全长/身高		44.80	44.79	
上下肢长度指数		81.00	82.07	
阿 昌 族				

身材亚中等偏矮。肤色呈浅黄褐色，黑发平直且硬，体毛及胡须少，眉峰微显。眼裂开度中等，眼裂斜度内外平行，上眼睑褶皱发育好，多数人无蒙古褶。鼻根低平到中等，鼻翼微突，鼻孔横向呈卵圆形，鼻宽中等，属狭鼻型。上唇皮肤部平直，唇型稍突，红唇中等厚。面部扁平，颧骨突出，多属狭面型和过狭面型。头型多属中头型、高头型和阔头型。

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	187.56	179.63	
	头宽	148.17	141.27	
	头围	555.50	539.85	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	124.58	118.89
	面宽	139.41	132.51	
	额宽	101.58	99.31	
	口宽	53.96	51.38	
	鼻高	56.38	53.68	
	鼻宽	39.05	35.70	
	鼻长	—	—	
	指数特征	头指数(头宽/头长×100)	79.09	78.70
	面指数(面高/面宽×100)	89.47	89.84	
	鼻指数(鼻宽/鼻高×100)	69.52	66.82	
面部	多属狭面型和过狭面型			
眼部	眼裂开度中等, 上眼睑皱褶发育好			
观察特征	鼻部	属狭鼻型		
	耳部	—		
	口部	唇型稍突, 红唇中等厚		
	毛发	黑发平直且硬, 体毛及胡须少		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1628.92	1522.15	
	肩宽	363.41	326.77	
	上肢全长	737.28	681.19	
	上臂长	—	—	
	手长	—	—	
	腰围	—	—	
	测量特征	骨盆宽	270.29	273.86
	下肢全长	901.92	849.44	
	大腿长	—	—	
	小腿长	—	—	
	足长	—	—	
	足高	—	—	
	体重(kg)	54.27	48.16	
	指 数 特 征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
足长宽指数		—	—	
上肢全长/身高		—	—	
上下肢长度指数		—	—	
普 米 族				

身材中等偏高。皮肤浅黄褐色，发黑直，男性胡须极少。眼裂开度中等，眼裂斜度内外平行，多数人无蒙古褶。鼻梁中等高，鼻翼发育中等，鼻孔倾斜呈卵圆形。红唇中等厚，呈正唇型。颧骨突出，面部扁平。耳垂圆形，多数人耳壳无达尔文结节。头长而窄，属中头型、高头型和狭头型。面高而窄，属过狭面型和狭面型。鼻高而窄，属狭鼻型。体型属瘦。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	193.56	185.42
		头宽	149.33	142.77
		头围	563.63	548.14
		头矢状弧	—	—
	观察特征	头冠状弧	—	—
		面高	130.88	122.78
		面宽	143.48	135.93
		额宽	104.09	101.27
		口宽	53.78	50.94
		鼻高	60.09	56.55
		鼻宽	38.18	34.65
	指数特征	鼻长	—	—
		头指数(头宽/头长×100)	77.19	77.08
面指数(面高/面宽×100)		91.29	90.45	
观察特征	鼻指数(鼻宽/鼻高×100)	63.80	61.56	
	面部	过狭面型和狭面型		
	眼部	多数人无蒙古褶		
	鼻部	狭鼻型		
	耳部	多数人耳壳无达尔文结节		
	口部	正唇型		
体 部 特 征	测量特征	毛发	发黑直, 男性胡须极少	
		身高	1665.23	1540.44
		肩宽	366.58	325.62
		上肢全长	734.69	672.33
		上臂长	—	—
	观察特征	手长	—	—
		腰围	—	—
		骨盆宽	254.67	259.05
		下肢全长	822.38	771.46
		大腿长	—	—
观察特征	小腿长	—	—	
	足长	—	—	
	足高	—	—	
	体重(kg)	57.98	51.40	

续表

		参 数	男(mm)	女(mm)
体 部 特 征	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
		足长宽指数	—	—
		上肢全长/身高	—	—
		上下肢长度指数	—	—
鄂 温 克 族				
<p>身材中等偏矮, 肤色浅褐, 直形褐色头发。头型男性多中头型, 女性多圆头型和超圆头型。面型多中面型和阔面型。眼裂斜度外高内低, 有蒙古褶, 多无上眼睑皱褶。鼻梁男性多直型, 女性多凹型。约半数人有达尔文结节, 耳垂较大, 多呈游离的圆形。</p>				
体质特征参数				
		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	185.02	176.14
		头宽	160.91	153.08
		头围	—	—
		头矢状弧	—	—
		头冠状弧	—	—
		面高	198.53	185.37
		面宽	148.78	140.02
		额宽	107.90	104.47
		口宽	50.18	47.19
		鼻高	—	—
	鼻宽	38.44	34.33	
	鼻长	51.28	46.87	

续表

		参 数	男(mm)	女(mm)
头面部特征	指数特征	头指数(头宽/头长×100)	86.96	87.04
		面指数(面高/面宽×100)	83.04	83.16
		鼻指数(鼻宽/鼻高×100)	75.33	73.80
	观察特征	面部	中面型和阔面型	
		眼部	多无上眼睑皱褶,有蒙古褶	
		鼻部	直型鼻	凹型鼻
耳部		耳垂多圆形		
口部				
	毛发	褐色,直形		
体部特征		身高	1644.20	1520.70
		肩宽	350.00	315.00
		上肢全长	777.10	723.10
		上臂长	—	—
		手长	—	—
	测量特征	腰围	—	—
		骨盆宽	280.80	282.40
		下肢全长	—	—
		大腿长	—	—
		小腿长	—	—
		足长	—	—
		足高	—	—
		体重(kg)	—	—
	指数特征		身高坐高指数(坐高/身高×100)	53.12
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
		足长宽指数	—	—
		上肢全长/身高	47.28	47.55
		上下肢长度指数	—	—

怒 族

女性均为亚中等身高，肤色较深，多为黄色，有一部分人是深黄肤色。大多无蒙古褶，上眼睑大多无皱褶。鼻根低。上唇皮肤高度较低。耳垂三角形率最高。怒族男女均为圆头型、高头型、狭头型、中鼻型、中腿型、中胸型、中肩型、中骨盆。男性为中面型、长躯干型，女性还为阔面型、中躯干型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	185.70	178.30
		头宽	148.20	143.00
		头围	556.10	541.10
		头矢状弧	—	—
		头冠状弧	—	—
		面高	121.00	111.60
		面宽	140.80	133.70
		额宽	107.80	106.30
		口宽	—	—
		鼻高	52.90	48.20
	鼻宽	37.30	34.70	
	鼻长	—	—	
	指数特征	头指数(头宽/头长×100)	80.00	80.20
		面指数(面高/面宽×100)	85.90	83.60
鼻指数(鼻宽/鼻高×100)		71.00	72.60	
观察特征	面部	中面型	阔面型	
	眼部	大多无蒙古褶		
	鼻部	鼻根低		
	耳部	耳垂三角形为主		
	口部	上唇皮肤高度较低		
	毛发	—		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1608.70	1509.40	
	肩宽	367.90	334.70	
	上肢全长	705.90	658.10	
	上臂长	295.80	274.20	
	手长	—	—	
	腰围	—	—	
	测量特征	骨盆宽	276.50	272.90
	下肢全长	872.30	833.60	
	大腿长	404.10	393.40	
	小腿长	370.90	349.90	
	足长	—	—	
	足高	—	—	
	体重(kg)	54.50	49.50	
	指 数 特 征	身高坐高指数(坐高/身高×100)	53.50	53.50
		身高肩宽指数(肩宽/身高×100)	22.90	22.20
手长宽指数		—	—	
足长宽指数		—	—	
上肢全长/身高		43.90	43.60	
上下肢长度指数		81.00	79.00	
基 诺 族				

身材矮小。肤色为浅黄色，眼为深褐色，黑发平直而硬，眉毛中等。眼裂开度中等，眼裂斜度外高内低，大多有从显著到中等显著的蒙古褶，多数有上眼睑皱褶。鼻根中等高，鼻梁多为直型，鼻尖、鼻基部水平向前，鼻翼中等，鼻孔形状多为椭圆形，属阔鼻型。耳垂多为方形。面部扁平度中等，下颌微向前突，上唇皮肤部较高，红唇中等厚。男性属阔面型，女性属过阔面型，头型属中头型。

第一章 中国少数民族简况及其体质特征

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征		头长	184.34	176.51
		头宽	141.21	136.38
		头围	—	—
		头矢状弧	—	—
	测量特征	头冠状弧	—	—
		面高	118.09	110.10
		面宽	140.67	133.34
		额宽	114.88	110.88
		口宽	51.73	48.94
		鼻高	3.93	3.94
		鼻宽	37.78	34.71
		鼻长	43.64	39.54
	指数特征	头指数(头宽/头长×100)	76.76	77.33
		面指数(面高/面宽×100)	84.05	82.64
		鼻指数(鼻宽/鼻高×100)	72.99	72.71
观察特征	面部	阔面型	过阔面型	
	眼部	大多有蒙古褶		
	鼻部	阔鼻型		
	耳部	耳垂多为方形		
	口部	红唇中等厚		
	毛发	黑发平直而硬		
体 部 特 征	测量特征	身高	1565.40	1465.10
		肩宽	330.67	300.94
		上肢全长	703.21	650.65
		上臂长	283.48	261.18
		手长	175.98	166.95
		腰围	—	—
		骨盆宽	265.84	264.84
	指数特征	下肢全长	838.63	795.27
		大腿长	391.83	338.28
		小腿长	356.03	334.90
		足长	231.40	214.52
		足高	68.42	62.17
		体重(kg)	52.09	44.86
		身高坐高指数(坐高/身高×100)	52.22	52.43
		身高肩宽指数(肩宽/身高×100)	—	—
	手长宽指数	—	—	
	足长宽指数	—	—	
	上肢全长/身高	—	—	
	上下肢长度指数	—	—	

德 昂 族

身材较矮小，鼻宽大于两眼内角宽，属中鼻型，面型属过狭面型，头型属中头型、高头型和狭头型。

体质特征参数				
参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	183.53	177.40	
	头宽	144.64	138.31	
	头围	553.83	525.83	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	112.35	102.63
		面宽	137.11	129.08
		额宽	108.55	104.63
		口宽	55.34	54.56
		鼻高	48.04	44.00
		鼻宽	39.53	36.31
	指数特征	鼻长	—	—
		头指数(头宽/头长×100)	78.87	78.02
		面指数(面高/面宽×100)	81.97	79.58
观察特征	鼻指数(鼻宽/鼻高×100)	83.48	82.72	
	面部	过狭面型		
	眼部	—		
	鼻部	中鼻型		
	耳部	—		
	口部	—		
毛发	—			

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1599.70	1477.50	
	肩宽	355.50	312.00	
	上肢全长	718.60	655.70	
	上臂长	—	—	
	测量特征	手长	—	—
		腰围	—	—
		骨盆宽	262.5	257.60
		下肢全长	852.60	807.00
		大腿长	—	—
		小腿长	—	—
		足长	—	—
	足高	—	—	
	体重(kg)	52.79	47.24	
	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	22.23	21.12
		手长宽指数	—	—
足长宽指数		—	—	
上肢全长/身高		44.91	44.39	
上下肢长度指数	—	—		

保 安 族

身材中等偏矮。鼻型为狭鼻型。面型为狭面型和过狭面型。头型属于高头型、中头型和狭头型。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	186.91	179.41
		头宽	148.88	143.28
		头围	552.50	
		头矢状弧	—	—
		头冠状弧	—	—
		面高	128.63	120.65
		面宽	140.33	134.01
		额宽	101.96	100.43
		口宽	50.14	48.59
		鼻高	56.06	52.01
	鼻宽	35.41	32.50	
	鼻长	—	—	
	指数特征	头指数(头宽/头长×100)	79.68	79.92
		面指数(面高/面宽×100)	91.70	90.12
		鼻指数(鼻宽/鼻高×100)	63.52	62.80
观察特征	面部	狭面型和过狭面型		
	眼部	—		
	鼻部	狭鼻型		
	耳部	—		
	口部	—		
	毛发	—		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1633.95	1537.03	
	肩宽	370.00	339.33	
	上肢全长	711.88	657.99	
	上臂长	291.43	270.00	
	手长	182.32	170.77	
	测量特征	腰围	56.47	49.61
	骨盆宽	283.73	277.39	
	下肢全长	868.78	818.36	
	大腿长	435.09	414.94	
	小腿长	363.66	339.41	
	足长	245.29	226.44	
	足高	67.22	60.08	
	体重(kg)	56.47	49.61	
	指数特征	身高坐高指数(坐高/身高×100)	53.88	53.87
	身高肩宽指数(肩宽/身高×100)	22.66	22.08	
	手长宽指数	46.55	45.08	
	足长宽指数	41.11	39.57	
上肢全长/身高	43.58	42.81		
上下肢长度指数	82.01	80.45		

俄罗斯族

男性总体身材较高，宽胸型、宽肩型和宽骨盆型者相对占优。男性的身材较高且魁梧，躯干较长，腿部和臂部相对较长。女性总体身材中上等，女性的身材属中上等且较健壮，躯干长度、腿部和臂部的长度均为中等。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	183.48	174.15
		头宽	157.45	151.67
		头围	559.30	543.56
		头矢状弧	—	—
		头冠状弧	—	—
	指数特征	面高	125.08	116.64
		面宽	144.35	138.46
		额宽	109.73	106.43
	观察特征	口宽	—	—
		鼻高	51.98	47.48
		鼻宽	35.17	32.64
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	85.90	87.20
		面指数(面高/面宽×100)	86.81	84.32
		鼻指数(鼻宽/鼻高×100)	67.97	68.67
观察特征	面部	—	—	
	眼部	—	—	
	鼻部	—	—	
	耳部	—	—	
	口部	—	—	
	毛发	—	—	
体 部 特 征	测量特征	身高	1676.71	1558.53
		肩宽	387.28	354.53
		上肢全长	723.89	670.18
		上臂长	313.09	287.31
		手长	—	—
		腰围	—	—
		骨盆宽	294.21	287.53
		下肢全长	905.40	847.97
		大腿长	453.06	429.90
		小腿长	387.34	352.27
	观察特征	足长	—	—
		足高	—	—
		体重(kg)	69.66	62.61
	指数特征	身高坐高指数(坐高/身高×100)	52.71	53.59
		身高肩宽指数(肩宽/身高×100)	23.12	22.76
手长宽指数		—	—	
足长宽指数		—	—	
上肢全长/身高		43.23	43.01	
上下肢长度指数	80.11	79.13		

裕 固 族

身材中等偏高。男性以浓眉、横眉为主，女性以细眉、弯眉为主。男性胡须中等。上眼睑皱褶发育好，蒙古褶不明显。鼻梁较高，鼻翼沟发育明显，鼻尖水平，女性鼻尖上翘较多，属狭鼻型。面长而窄，头短且高，头型属于高头型和中头型。面型属于过狭面型。

体质特征参数				
参 数		男(mm)	女(mm)	
头 面 部 特 征	测量特征	头长	190.36	184.37
		头宽	150.98	148.00
		头围	568.44	559.83
		头矢状弧	339.58	327.84
		头冠状弧	336.63	329.50
		面高	135.05	126.78
		面宽	130.67	126.38
		额宽	131.15	128.94
		口宽	50.10	48.31
		鼻高	62.94	59.94
		鼻宽	37.40	34.86
		鼻长	—	—
指数特征		头指数(头宽/头长×100)	79.49	80.37
		面指数(面高/面宽×100)	103.63	100.70
		鼻指数(鼻宽/鼻高×100)	59.74	58.44

续表

		参 数	男(mm)	女(mm)
头 面 部 特 征	观察特征	面部	过狭面型	
		眼部	蒙古褶不明显	
		鼻部	狭鼻型	
		耳部	—	
		口部	—	
		毛发	男性胡须中等	
体 部 特 征	测量特征	身高	1672.72	1562.76
		肩宽	368.97	340.65
		上肢全长	736.77	688.91
		上臂长	—	—
		手长	184.98	172.92
		腰围	—	—
		骨盆宽	—	—
		下肢全长	863.42	814.85
		大腿长	—	—
		小腿长	—	—
		足长	247.52	227.66
		足高	—	—
		体重(kg)	60.19	54.05
		指数特征	身高坐高指数(坐高/身高×100)	53.38
身高肩宽指数(肩宽/身高×100)	22.10		21.80	
手长宽指数	47.10		46.00	
足长宽指数	39.80		39.80	
上肢全长/身高	44.10		44.10	
上下肢长度指数	85.20		84.50	

乌孜别克族

身材偏矮，均具有上眼睑皱褶，蒙古褶出现率则较低。鼻根几乎均为高型，鼻翼高度多为中等。耳垂多为圆形。上唇皮肤高度以中等为主，男性肩部较窄，而躯干下部较宽。

体质特征参数				
		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	183.30	170.70
		头宽	160.00	152.30
		头围	566.60	553.10
		头矢状弧	—	—
		头冠状弧	—	—
	指数特征	面高	121.70	112.00
		面宽	145.70	137.60
		额宽	115.3	110.8
		口宽	—	—
		鼻高	52.70	48.10
	观察特征	鼻宽	34.90	32.00
		鼻长	—	—
		头指数(头宽/头长×100)	87.56	89.29
		面指数(面高/面宽×100)	83.70	81.48
鼻指数(鼻宽/鼻高×100)		66.75	66.63	
观察特征	面部	—		
	眼部	均具有上眼睑皱褶		
	鼻部	鼻根几乎均为高型		
	耳部	耳垂多为圆形		
	口部	上唇皮肤高度以中等为主		
	毛发	—		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1685.40	1555.50	
	肩宽	364.60	333.70	
	上肢全长	745.60	683.30	
	上臂长	326.80	295.10	
	手长	—	—	
	腰围	—	—	
	测量特征	骨盆宽	297.00	281.60
		下肢全长	885.00	840.30
		大腿长	434.60	409.80
		小腿长	389.20	363.00
		足长	—	—
		足高	—	—
		体重(kg)	86.70	59.20
	指数特征	身高坐高指数(坐高/身高×100)	58.78	54.36
		身高肩宽指数(肩宽/身高×100)	21.72	21.32
		手长宽指数	—	—
	足长宽指数	—	—	
	上肢全长/身高	44.30	43.86	
	上下肢长度指数	84.39	81.28	

鄂 伦 春 族

身材偏矮，肤色较深，浅褐色头发。头型多圆头型和过圆头型。面型多阔面型。眼裂开度中等，眼裂斜度外高内低，缺失蒙古褶，无上眼睑皱褶。多属直型鼻。耳垂多为附着的圆形。

第一章 中国少数民族简况及其体质特征

体质特征参数

		参 数	男 (mm)	女 (mm)
头 面 部 特 征	测量特征	头长	183.51	177.10
		头宽	157.99	152.65
		头围	—	—
		头矢状弧	—	—
		头冠状弧	—	—
	指数特征	面高	199.75	188.96
		面宽	147.38	139.11
		额宽	109.44	106.65
		口宽	51.35	48.06
		鼻高	—	—
		鼻宽	37.58	34.09
		鼻长	49.31	46.06
	观察特征	面部	阔面型	
		眼部	无上眼睑皱褶,蒙古褶缺失	
		鼻部	直型鼻	
		耳部	耳垂多圆形	
		口部	—	
体 部 特 征	测量特征	毛发	浅褐色,直形	
		身高	1597.90	1482.80
		肩宽	335.70	304.40
		上肢全长	751.10	690.50
		上臂长	—	—
		手长	—	—
		腰围	—	—
		骨盆宽	—	—
		下肢全长	—	—
		大腿长	—	—
	指数特征	小腿长	—	—
		足长	—	—
		足高	—	—
		体重 (kg)	—	—
		身高坐高指数 (坐高/身高×100)	53.89	54.35
身高肩宽指数 (肩宽/身高×100)	—	—		
手长宽指数	—	—		
足长宽指数	—	—		
上肢全长/身高	47.07	46.50		
上下肢长度指数	—	—		

独龙族

身材矮小，眼色多为褐色和深褐色。肤色多为黄色。男女均以高头型、狭头型、中胸型、中骨盆型率最高。男性还以中头型、狭面型、中鼻型、长躯干型、中腿型、宽肩型率最高；女性还以圆头型、阔面型、狭鼻型、中躯干型、亚短腿型、窄肩型率最高。鼻根高度男性中型率最高，女性低型率最高。耳垂类型男性的方形率与圆形率接近，女性三角形率最高。上唇皮肤部高度男女均以中等型率最高。

体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	182.60	173.80
		头宽	146.30	141.50
		头围	551.50	534.20
		头矢状弧	—	—
		头冠状弧	—	—
		面高	119.40	109.60
		面宽	138.00	129.70
		额宽	103.00	102.60
		口宽	—	—
		鼻高	52.60	48.90
	鼻宽	37.40	33.60	
	鼻长	—	—	
	指数特征	头指数(头宽/头长×100)	80.20	81.50
面指数(面高/面宽×100)		86.60	84.60	
鼻指数(鼻宽/鼻高×100)		71.30	69.00	
观察特征	面部	狭面型	阔面型	
	眼部	大多无蒙古褶		
	鼻部	中鼻型	狭鼻型	
	耳部	耳垂方形和圆形	耳垂三角形	
	口部	上下唇厚度中等		
	毛发	—		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1574.30	1464.90	
	肩宽	363.10	325.00	
	上肢全长	690.60	637.60	
	上臂长	293.30	266.30	
	手长	—	—	
	腰围	—	—	
	测量特征	骨盆宽	—	—
	下肢全长	823.60	774.80	
	大腿长	392.30	368.10	
	小腿长	366.90	343.20	
	足长	—	—	
	足高	—	—	
	体重(kg)	52.10	44.90	
	指数特征	身高坐高指数(坐高/身高×100)	52.90	53.50
		身高肩宽指数(肩宽/身高×100)	22.90	22.20
		手长宽指数	—	—
足长宽指数		—	—	
上肢全长/身高		—	—	
上下肢长度指数		83.90	82.30	

塔 塔 尔 族

直形发,发色为黑色,眉毛中等发育,眉形以横眉为多。面型男女性均以卵圆形为主。眼裂方向多为水平型,两性眼裂宽度中等。蒙古褶出现率极少,上睑褶皱达睫毛处较多。眼色以褐色和天蓝色为多。鼻宽等于两眼内角宽占多数,鼻翼突度与鼻侧壁平行占多数,鼻翼高度中等,鼻孔形状男性卵圆形,女性椭圆,鼻孔最大横径矢状位多见,鼻尖方向男向前,鼻基部方向,鼻梁形状直,鼻根高度中等。上唇皮肤高度中等为多,上红唇厚度薄型为多;耳垂圆形为多。

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	187.69	181.23	
	头宽	164.27	157.95	
	头围	580.82	567.12	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	126.81	120.04
		面宽	—	—
		额宽	150.28	140.90
		口宽	—	—
	鼻高	57.16	51.39	
	鼻宽	38.75	33.99	
	鼻长	51.20	46.04	
	指数特征	头指数(头宽/头长×100)	87.52	87.15
		面指数(面高/面宽×100)	84.38	85.20
		鼻指数(鼻宽/鼻高×100)	67.79	66.14
观察特征	面部	卵圆形为主		
	眼部	上眼睑褶皱达睫毛处较多， 眼色以褐色和天蓝色为多		
	鼻部	鼻宽等于两眼内角宽占多数		
	耳部	耳垂圆形为多		
	口部	上红唇厚度薄型为多		
毛发	发形为直形，发色为黑色， 眉毛中等发育，眉形以横眉为多			
体 部 特 征	测量特征	身高	1707.14	158.69
		肩宽	433.02	396.65
		上肢全长	569.04	511.26
		上臂长	320.40	292.90
		手长	188.81	170.96
		腰围	843.33	788.59

续表

参 数		男(mm)	女(mm)
体 部 特 征	骨盆宽	—	—
	下肢全长	932.82	816.04
	大腿长	456.33	423.21
	测量特征		
	小腿长	386.98	335.71
	足长	254.14	235.23
	足高	—	—
	体重(kg)	69.46	58.93
	身高坐高指数(坐高/身高×100)	53.04	53.01
	身高肩宽指数(肩宽/身高×100)	22.39	22.35
	指数特征		
	手长宽指数	47.58	46.32
	足长宽指数	41.21	39.29
	上肢全长/身高	33.34	32.30
上下肢长度指数	—	—	
赫 哲 族			

身材中等偏高。发形直，发旋多为单旋顺时针方向。眼裂开度较窄，眼裂斜度多为外高内低，大多有蒙古褶。鼻梁较直，鼻尖和鼻基底部方向多为水平向前，属中鼻型和狭鼻型。上唇皮肤部高度中等，红唇较薄。耳垂形状多为三角形和圆形。面部较低而宽，面型男性多为中面型，女性多为阔面型。头圆而高，头型属于过圆头型、高头型和中头型。

体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	181.00	174.10	
	头宽	159.10	150.10	
	头围	556.70	557.60	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	121.60	113.10
		面宽	143.10	137.30
		额宽	105.20	102.80
		口宽	48.80	45.90
	鼻部	鼻高	52.70	50.10
		鼻宽	38.10	35.50
		鼻长	—	—
	指数特征	头指数(头宽/头长×100)	87.91	86.64
		面指数(面高/面宽×100)	84.95	82.37
		鼻指数(鼻宽/鼻高×100)	70.93	70.87
面部	中面型	阔面型		
观察特征	眼部	眼裂开度较窄, 大多有蒙古褶		
	鼻部	中鼻型和狭鼻型		
	耳部	耳垂形状多为三角形和圆形		
	口部	上唇皮肤部高度中等, 红唇较薄		
	毛发	发形直		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1667.10	1553.20	
	肩宽	388.70	346.50	
	上肢全长	731.10	669.20	
	上臂长	322.70	292.60	
	手长	176.20	168.30	
	测量特征	腰围	795.80	718.70
		骨盆宽	282.70	277.80
		下肢全长	900.40	857.00
		大腿长	481.10	458.00
		小腿长	357.10	341.80
		足长	244.30	222.30
	足高	70.10	58.40	
	体重(kg)	64.20	55.60	
	指数特征	身高坐高指数(坐高/身高×100)	—	—
		身高肩宽指数(肩宽/身高×100)	—	—
手长宽指数		—	—	
足长宽指数		—	—	
上肢全长/身高		—	—	
上下肢长度指数		—	—	
高 山 族				

台湾高山族中阿美人的体质特征如下：

身材中等，发黑直，肤色较淡。眼间距较宽，眼裂斜度大多水平，多数无上眼睑褶皱，80%以上的个体无蒙古褶。唇中等厚。耳廓大多无达尔文结节，耳垂多为游离型。鼻梁多为凹型，鼻梁属中鼻型。面型属狭面型，头型属中头型、高头型和狭头型。胸围指数较小。

阿美族体质特征参数

参 数		男(mm)	女(mm)	
头 面 部 特 征	头长	187.60	178.60	
	头宽	145.10	138.30	
	头围	—	—	
	头矢状弧	—	—	
	头冠状弧	—	—	
	测量特征	面高	—	—
	面宽	141.20	133.70	
	额宽	104.50	101.00	
	口宽	52.40	49.10	
	鼻高	—	—	
	鼻宽	41.90	38.90	
	鼻长	52.40	49.40	
	指数特征	头指数(头宽/头长×100)	77.34	77.43
	面指数(面高/面宽×100)	90.00	88.00	
	鼻指数(鼻宽/鼻高×100)	80.00	79.00	
面部	过狭脸型			
眼部	多数无上眼睑褶皱			
观察特征	鼻部	中鼻型		
	耳部	耳廓大多无达尔文结节， 耳垂多为游离型		
	口部	唇中等厚		
	毛发	发黑直		

续表

参 数		男(mm)	女(mm)	
体 部 特 征	身高	1646.00	1559.00	
	肩宽	367.00	341.00	
	上肢全长	757.00	713.00	
	上臂长	—	—	
	手长	—	—	
	腰围	—	—	
	测量特征	骨盆宽	271.00	270.00
	下肢全长	—	—	
	大腿长	—	—	
	小腿长	—	—	
	足长	—	—	
	足高	—	—	
	体重(kg)	—	—	
	指 数 特 征	身高坐高指数(坐高/身高×100)	52.00	52.00
		身高肩宽指数(肩宽/身高×100)	—	—
		手长宽指数	—	—
足长宽指数		—	—	
上肢全长/身高		—	—	
	上下肢长度指数	—	—	

台湾高山族中泰雅人的体质特征如下：

身材亚中等偏矮。发黑直且粗，肤色较淡。鼻长，脸长。眼裂斜度大多为水平，半数以上个体无蒙古褶。唇中等偏薄。半数以上无达尔文结节，耳垂多为游离或附着。鼻梁大多为凹型，鼻梁属中鼻型。面型属过狭面型，头型属圆头型、高头型和中头型。

泰雅族体质特征参数

		参 数	男(mm)	女(mm)
头 面 部 特 征	测量特征	头长	179.80	174.20
		头宽	145.70	141.40
		头围	—	—
		头矢状弧	—	—
		头冠状弧	—	—
		面高	125.90	119.40
		面宽	134.30	128.80
		额宽	102.50	99.70
		口宽	51.00	47.50
		鼻高	—	—
	鼻宽	40.90	37.80	
	鼻长	52.10	49.30	
	指数特征	头指数(头宽/头长×100)	81.00	81.00
		面指数(面高/面宽×100)	94.00	93.00
鼻指数(鼻宽/鼻高×100)		79.00	77.00	
观察特征	面部	过狭面型		
	眼部	半数以上个体无蒙古褶		
	鼻部	中鼻型		
	耳部	半数以上无达尔文结节, 耳垂多为游离或附着		
	口部	唇中等偏薄		
	毛发	发黑直且粗		
体 部 特 征	测量特征	身高	1601.00	1498.00
		肩宽	367.00	337.00
		上肢全长	729.00	680.00
		上臂长	—	—
		手长	—	—
		腰围	—	—
		骨盆宽	262.00	261.00
		下肢全长	—	—
		大腿长	—	—

第二章 中国少数民族遗传学的分子基础

第一节 遗传物质

遗传物质是核酸这一想法，是在 1928 年发现转化的基础上得出的。肺炎双球菌能够通过引起肺炎杀死老鼠。这种细菌的有毒物质被认为是它的外壳多糖，是一种细菌表面的物质，能够避免细菌被宿主破坏掉。不同类型的肺炎双球菌有不同的外壳多糖。但它们都有平滑的表面。每一种光滑肺炎双球菌都能产生变异，不再产生外壳多糖，它们有粗糙的表面（由外壳多糖之下的物质组成），无毒性，不能杀死老鼠。因其表面缺少了多糖，使宿主能够杀死这种细菌。

当光滑细菌被加热致死，它们失去了杀死动物的能力。但是，没有活性的热致死光滑细菌与无毒性的粗糙细菌混在一起时，与其单独时相比产生非常不同的效果。当它们被一起注射到动物体内，老鼠由于感染肺炎双球菌而死亡（图 2-1）。在老鼠体内能够分离到有毒性的光滑细菌。

在这个试验中，死去的光滑细菌是三型的。活的粗糙细菌是从二型中得到的。从致死后的动物体内分离到的细菌有三型的光滑外表面。因此死去的三型光滑细菌的一些性质能够转化活的粗糙细菌，从而产生三型外壳多糖，变得有毒性。

引起转化的死细胞成分可以通过形成一个“自由细胞体系”来纯化。在这个体系中，死光滑细菌的提取物能够加到活的 R 细菌中。Avery 的经典试验在 1944 年证明这种分离的可转移分子为 DNA。此结果令人感到十分惊奇，因为当时人们还不知道 DNA 是肺炎球菌的成分，虽然人们早就知道 DNA 是真核生物染色体的重要组成成分。这个结果说明了原核生物的遗传物质也是

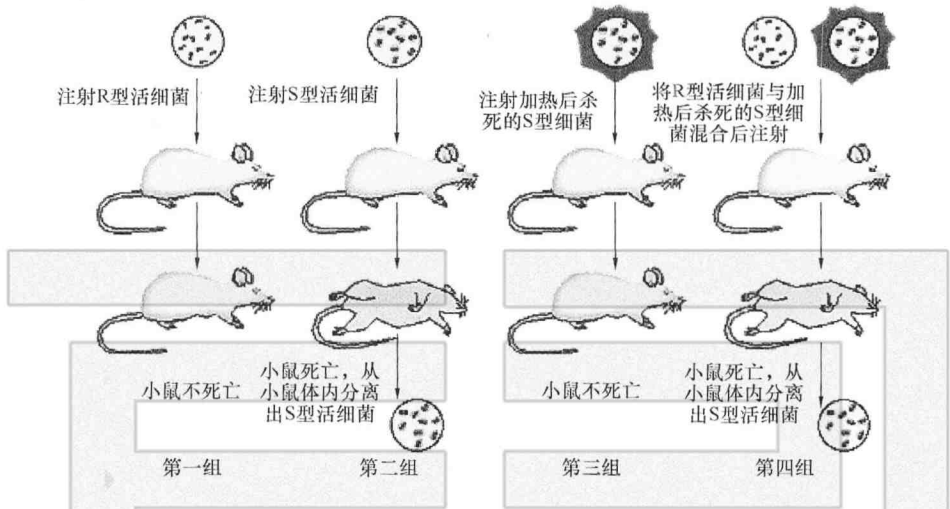


图 2-1 转化过程的核心是 DNA

DNA，这为细菌和高等生物的遗传提供了一种独特的观点。

当可转移分子被发现由 DNA 组成之后，下一步就是要阐述 DNA 在一个相当不同的体系中提供遗传物质。噬菌体 T2 是侵染大肠杆菌的一种病毒。当噬菌体加到细菌中去时，它们吸附在外表面，一些物质进入到细菌中，20 分钟后细菌会溶解并释放出大量的子代噬菌体。

在 1952 年，Hershey 和 Chase 用放射性标记 T2 噬菌体，感染细菌 (^{32}P 标记 DNA, ^{35}S 标记蛋白质)。图 2-2 说明了这个试验的结果。感染后的细菌在搅拌器中破碎，通过离心分离到两种成分。一种是从细菌表面释放的空噬菌体外壳，它们由蛋白质组成，因此有 ^{35}S 的放射性标记。另一种成分是被感染细胞本身。

大多数 ^{32}P 标记出现在被感染细胞内。感染后产生的子代噬菌体中 30% 带有 ^{32}P 标记。子代仅含有很少 (不足 1%) 的原噬菌体蛋白质。该试验直接表明父代噬菌体的 DNA 进入到细菌中，变成了子代噬菌体的一部分，恰好符合遗传物质的遗传特性。

噬菌体的复制依赖于被感染宿主细胞的复制系统去生产很多它自己的拷贝。噬菌体拥有和细胞基因组类似的遗传物质，复制是严格的，并且它们有同样的控制遗传的方法。T2 的例子进一步说明了不管是在细胞基因组中还是病毒中，遗传物质就是 DNA 这一结论。

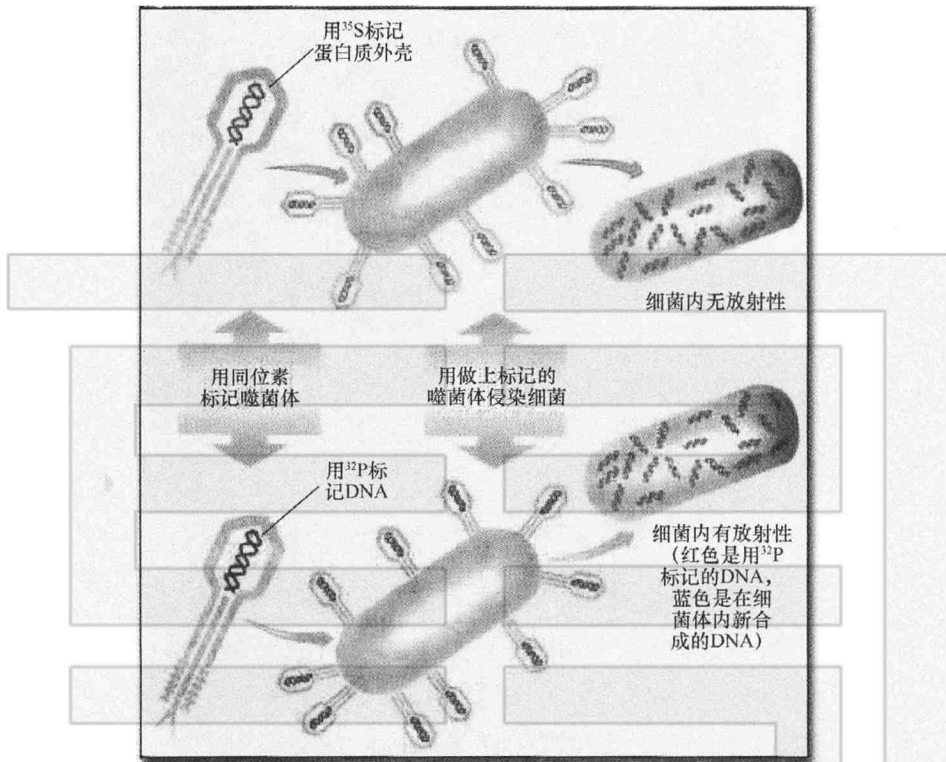


图 2-2 T2 噬菌体的遗传物质是 DNA

当 DNA 加入到几种培养生长的原核细胞时，核酸进入细胞，在一些条件下导致新蛋白质的产生。这些试验现在已是实验室的常规试验，引入纯化的 DNA 可以产生特殊蛋白质。

由于历史的原因，在真核生物中操作这样的试验被称为转染 (Transfection)，与细菌的转化相对应。引进宿主细胞的 DNA 成为其遗传物质的一部分，同其他部分一样遗传下去。它的表达赋予细胞新的特性。当初，只有当个别细胞能在培养基上存活下来时试验才是成功的。现在，DNA 可以通过显微注射的方法注入老鼠卵细胞，并且成为老鼠遗传物质中稳定的一部分。

上述试验证明，DNA 不仅是真核生物中的遗传物质，还可以在不同种属之间转移，并能保持功能。

已知生物及大多数病毒中的遗传物质是 DNA。但是，有些病毒使用代替的核酸—RNA 作为遗传物质。尽管它的化学结构与 DNA 有一些细微的不同，

但 RNA 能起到同样的作用。

因此，遗传物质的本质是核酸。事实上，除了 RNA 病毒以外，其遗传物质是 DNA。

第二节 DNA 的双螺旋结构

核酸是由多聚核苷酸链组成。图 2-3 是链的骨架结构，它由交替的五元糖及磷酸组成，一个嘌呤或嘧啶碱基与核糖相连接。DNA 和 RNA 的名字来源于其组成是脱氧核糖核苷酸还是核糖核苷酸。区别在于 RNA 的核糖在其五元环 2' 位点上是 OH 基。

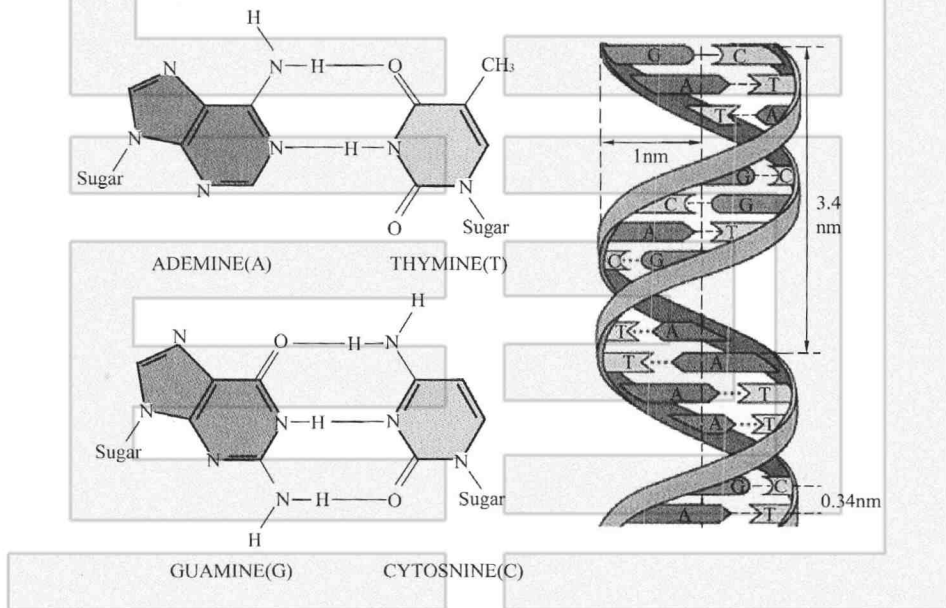


图 2-3 在一系列由 5'-3'核糖-磷酸连接而成的骨架上，
碱基从糖环突出，这样便形成了一条多核苷酸链

每一种核酸都有四种不同的碱基。两个嘌呤——鸟嘌呤和腺嘌呤出现在 DNA 和 RNA 中。两种嘧啶——胞嘧啶和胸腺嘧啶仅出现在 DNA 中，而在 RNA 中尿嘧啶代替了胸腺嘧啶。胸腺嘧啶和尿嘧啶的唯一区别是 C5 上是否存在甲基。这些碱基通常用其第一个英文字母表示。DNA 包括 A、T、G、

C, RNA 包括 A、G、U、C。

糖五元环的 1' 位置由糖苷键与嘌呤的 N1 或嘧啶的 N9 连接。为避免杂环与核糖之间计数的混乱, 糖五元环上的位点用序号加 “'” 表示。多聚核苷酸链是通过五元环上 5' 位点和另一个五元环 3' 位点间的磷酸基团连接起来。所以糖-磷酸骨架结构是 5' 到 3' 走向。碱基在骨架结构上平插出来。核苷酸链的一个末端是一个自由的 5' 基团, 另一个末端是自由的 3' 基团。习惯上通常沿着 5'-3' 的方向来写核苷酸链, 即从左边 5' 末端到右边的 3' 末端。

Watson 和 Crick 于 1953 年提出了 DNA 的双螺旋结构 (double helix)。X 光衍射数据显示 DNA 具有常规的螺旋结构, 每个螺旋 $3.4 \cdot (3.4\text{nm})$, 直径 $20 \cdot (2\text{nm})$ 。由于每个螺旋间的距离是 3.4nm, 所以每个螺旋有 10 个碱基。

DNA 的密度说明螺旋包括两条多聚核苷酸链。如果每条链的碱基是面对面配对, 并且嘌呤与嘧啶配对可避免嘌呤之间 (太厚) 以及嘧啶之间 (太薄) 的配对, 那么螺旋不变的直径就可以得到解释。

尽管 DNA 中每一个碱基的实际数目不同, 但 G 与 C 含量相等, A 与 T 的含量也相等。因此任何 DNA 的组成都可用其碱基含量来描述, 即 G+C。G+C 数目不同, 种间变化范围为 26% 到 74%。

Watson 和 Crick 提出双螺旋中两条多聚核苷酸链通过碱基之间的氢键联系起来。G 只能与 C 形成氢键, 而 A 只能与 T 形成氢键。这种反应被称为碱基配对 (Base pairing), 并且碱基对 (G 与 C, A 与 T) 是互补的。这个模型要求两条链是反向平行 (Antiparallel)。在双螺旋中, 一条是 5'-3' 走向, 另一条是 3'-5' 走向 (图 2-4)。

糖-磷酸骨架在外侧, 磷酸集团上带有负电荷。当 DNA 在体内溶液中, 由于阳离子 (通常是 Na 离子) 的结合使这些电荷被中和。在体内自然状态下, 带正电的蛋白质提供了一些中和能力, 这些蛋白质对 DNA 在细胞中的组织有重要作用。碱基在中间。碱基是平面的结构, 与螺旋的中轴垂直。若想象双螺旋是旋转楼梯, 那么碱基对就类似台阶 (图 2-5) 沿着螺旋碱基一个堆着一个, 如同平板一样堆积。每一个碱基对比另一对沿轴旋转 36 度, 所以 10 个碱基便旋转完整的 360 度。两条链相互缠绕形成一个带有小沟 (12·) 和大沟 (22·) 的双螺旋 (图 2-6), 双螺旋是右手螺旋, 即沿轴顺时针旋转。这种特征与已公认的 B 型 DNA 是一致的。

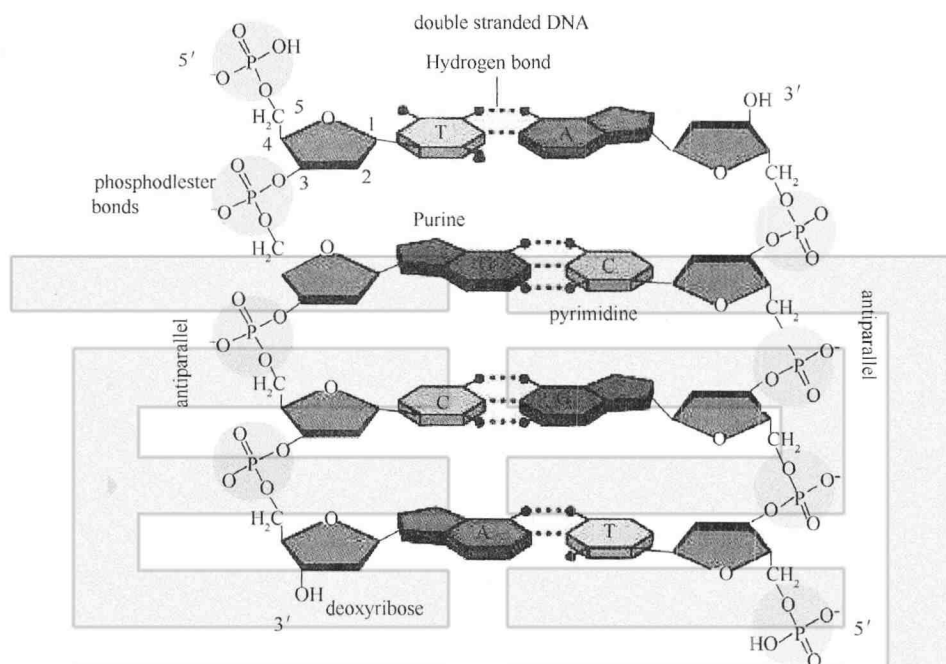


图 2-4 由于嘌呤碱基总是与嘧啶碱基配对，因此双螺旋的宽度是恒定的，图中序列为 T-A、C-G、A-T、G-C

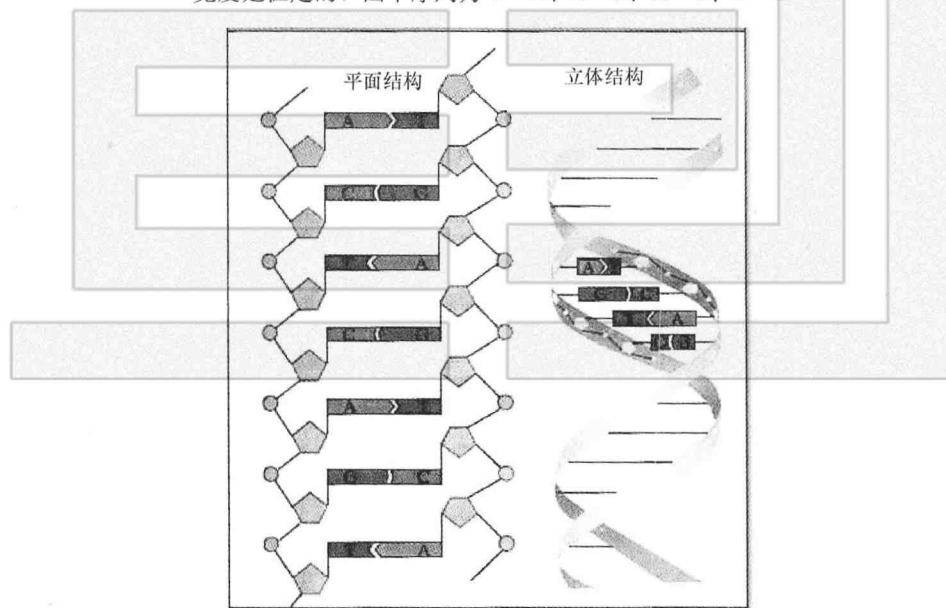


图 2-5 碱基之间的配对平面与核糖磷酸骨架垂直

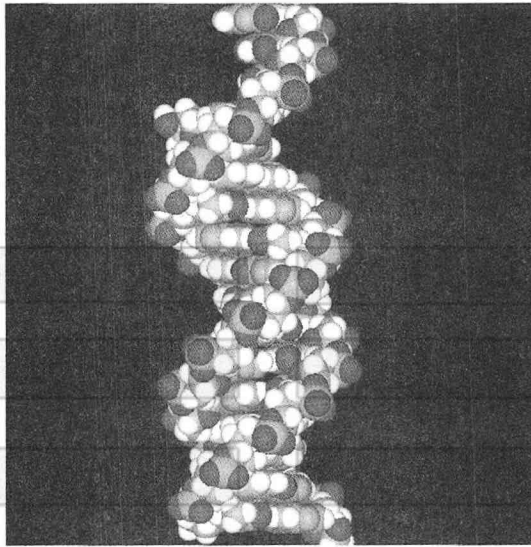


图 2-6 DNA 的两条链形成双螺旋结构

第三节 遗传密码

DNA 普遍结构的一个重要特点，是其独立于组成它的核苷酸的结构。DNA 中核苷酸序列是很重要的，并不是因为它的结构，而是它编码了组成相应多肽的氨基酸序列。DNA 序列与相应蛋白质序列的关系被称为遗传密码 (Genetic code)。

每一种蛋白质的结构或酶活性是与核苷酸的原始结构相关的。通过确定蛋白质中的氨基酸序列，证明基因携带了活性多肽链所需要的所有信息。因此，基因这种简单的结构完全能够表示数不清的多肽链。

细胞内多种蛋白质产物一起行使催化和结构活性，最终产生细胞表型。当然，除了编码蛋白质的基因序列以外，DNA 也包括特定的序列，它的功能由调节因子（通常是蛋白质）来识别。DNA 的功能是由它的序列直接决定的，而不是中间密码。通常情况下，表达蛋白质的基因，或者被识别的序列，组成了遗传信息。

遗传密码是由一个翻译核酸序列的复杂解码器解码的。如果 DNA 上的信

息有意义，解码器是必须的。在任何给定的区域内，只有一条 DNA 链编码蛋白质，因此遗传密码是一个碱基序列（而不是碱基对）。

遗传密码以 3 个核苷酸为单位阅读，每一组代表一种氨基酸。每三个核苷酸序列成为一个密码子（Codon）。一个基因包括一系列密码子，从一端的起始密码（Initiation codon）往另一端的终止密码（Termination codon）阅读。按照通常的 5'到 3'顺序，编码蛋白质的 DNA 核苷酸序列与从 N 端到 C 端的氨基酸序列一致。

遗传密码按照不重叠规则，从固定起始部位开始阅读。不重叠意味着每个密码子包括三个核苷酸，连续的密码子由连续的三个核苷酸表达。使用固定的起始位点说明编码蛋白质必须从一端开始，向另一端连续延伸，因此密码序列的不同部位不能独立地阅读。

密码子的本质说明两种突变会有不同的结果。如果某一个序列如下：

UUU AAA GGG CCC（密码子）

Aa1 aa2 aa3 aa4（氨基酸）

那么一个点突变只影响一个氨基酸。比如，A 被碱基 X 替换后使 aa2 被 aa5 代替：

UUU AAX GGG CCC

Aa1 aa5 aa3 aa4

因为只有第二个密码子发生了改变。但如果一个突变产生缺失或者增加了一个碱基，将会改变整个序列的读码框（Reading frame），这种改变称为移码（Frame shift）。插入可能会是这种形式：

UUU AAX AGG GCC C

Aa1 aa5 aa6 aa7

因新的三联体序列和旧的很不同，蛋白质突变位点后的整个氨基酸序列都会发生改变，因此蛋白质将失去其功能。移码突变可由吡啶（Acridine）引起，吡啶能够与 DNA 结合，破坏双链结构，使其在复制中引入一个新的碱基或缺失一个碱基。每一个由吡啶诱导的突变导致缺失或者增加一个碱基。

如果突变是由于增加一个碱基而发生的，那么缺失这个碱基就会使其恢复野生型。但是缺失与之相邻的碱基也可能使突变回复。对这些突变的研究结果，为遗传密码的本质提供了证据。一个碱基的插入或者缺失能够使突变

位点后的蛋白质序列都发生改变（图 2-7）。

但是一个插入和缺失只会引起它们之间的读码框发生改变，第二个突变位点后恢复正确的读码。移码突变最初来自分析 T6 噬菌体 rII 区域内的突变。所有的吡啶突变可以分为两类，用 - 和 + 表示。每一种突变类型引起移码突变，+ 型实质上是碱基的增加，而 - 型是碱基的缺失。双突变的 ++ 或者 -- 类型依旧表现出突变型，但是 -+ 或者 +- 类型互相抑制，一个突变被称为另一个突变的抑制子（Suppressor，此处抑制子有非同一般的含义，因为两个突变发生在同一个基因中）。上述结果都说明遗传密码具有一个固定的阅读顺序，从起始位点开始。因此单一增加和缺失碱基可互相补偿，而双缺失或增加仍引起突变，但这些并未表明一个密码子包含多少个碱基。

当获得三突变体时，只有 (+++) 和 (---) 表现野生表型，其他组合都是突变。如果我们将三个缺失或三增加仅产生一个氨基酸的缺失或增加时，则表明密码子是三联体。在两边的突变中间是一段不正确的氨基酸序列，而两边的氨基酸顺序都是正确（图 2-7）。

如果遗传密码是不重叠的三联体，那么会有三种可能的方式将核苷酸翻

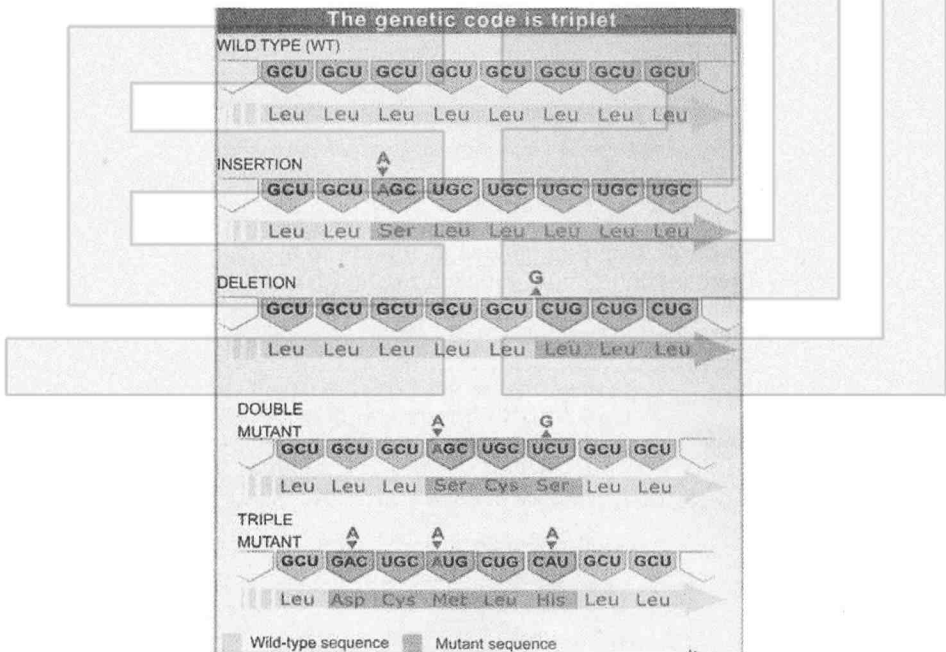


图 2-7 移码突变的读码是从三联体密码的固定位点开始的

译成蛋白质，这三种可能的读码（Reading frame）方式称为读码框架。比如序列：ACGACGACGACGACGACG，可能的读码框架就有以下三种：

```

ACG ACG ACG ACG ACG ACG ACG ACG
CGA CGA CGA CGA CGA CGA CGA CGA
GAC GAC GAC GAC GAC GAC GAC GAC
    
```

一个由能翻译成氨基酸序列的三联体构成的阅读框称为开放读框（Open reading frame, ORF）。一段翻译成蛋白质的序列有一个阅读框架，它有一个特殊的起始密码子（AUG），从此延伸出一系列代表氨基酸的三联体，一直到在三种类型的终止密码子上结束。

如果终止密码子频繁出现，就会阻止阅读框被翻译成蛋白质。一个序列的三个阅读框全部被阻断，那么它就会失去翻译成蛋白质的功能。

当获得一个未知的 DNA 序列后，就可分析其三个读码框是被阻断的还是开放的。在任何一段 DNA 中，通常不会超过一个读码框是开放的（图 2-8），因为替换的读码框被频繁出现的终止密码子阻断。一般情况，开放读框不可能太长。如果它不翻译成蛋白质，将不存在阻止终止密码子聚集的选择压力。证明序列是开放框是确定该框架能翻译为蛋白质的首要证据。一个不能表达蛋白质的开放读框被称为不确定读框（URF）。

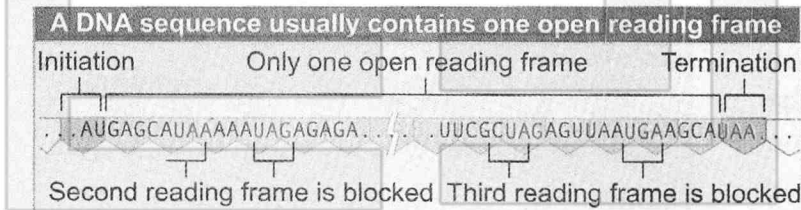


图 2-8 一个开放阅读框起始于 AUG，以三联体形式延续至终止密码，而受阻阅读框是因终止密码的频繁插入而形成的

第四节 基 因

一、基因的概念

基因（gene）是遗传学家约翰逊（W. Johanssen）在 1909 年提出来的。

他用基因这一名词来表示遗传的独立单位，相当于孟德尔在豌豆试验中提出的遗传因子。

在遗传学发展的早期阶段，基因仅仅是一个逻辑推理的概念，而不是一种已经证实了的物质和结构。由于科学研究水平的不断提高，从浅入深，由宏观到微观，基因的概念也在不断的修正和发展。从遗传学史的角度看，基因概念大致分以下几个阶段：孟德尔的遗传因子阶段、摩尔根基因阶段、顺反子阶段和现代基因阶段。

1. 孟德尔的遗传因子阶段

19世纪60年代初，孟德尔对具有不同形态的豌豆作杂交实验，在解释实验中每种性状的遗传行为时，用A代表红花，a代表白花，表明生物的某种性状是由遗传因子负责传递的，遗传下来的不是具体的性状，而是遗传因子。遗传因子是颗粒性的，在体细胞内成双存在，在生殖细胞内成单存在。孟德尔所说的“遗传因子”是代表决定某个性状遗传的抽象符号。

孟德尔在阐明遗传因子在世代中传递规律时，就已经认识到了基因的两个基本属性：基因是世代相传的，基因是决定遗传性表达的。现在所说的“基因是生物体传递遗传信息和表达遗传信息的基本物质单位”，实际上就是孟德尔所阐明的基因观。

2. 摩尔根基因阶段

1909年，丹麦遗传学家约翰逊创造了“基因”这一术语，用来表达孟德尔的遗传因子，只是提出了遗传因子的符号，没有提出基因的物质概念。摩尔根对果蝇的研究结果表明，1条染色体上有很多基因，一些性状的遗传行为之所以不符合孟德尔的独立分配定律，就是因为代表这些性状的基因位于同一条染色体上，彼此连锁而不易分离。这样，代表特定性状的特定基因与某一条特定染色体上的特定位置联系起来。基因不再是抽象的符号，而是在染色体上占有一定空间的实体，从而赋予基因以物质的内涵。

3. 顺反子阶段

早期的基因概念是把基因作为决定性状的最小单位、突变的最小单位和重组的最小单位，后来，这种“三位一体”的概念不断受到新发现的挑战。

20世纪50年代以后，随着分子遗传学的发展，1953年在沃森和克里克提出DNA的双螺旋结构以后，人们普遍认为基因是DNA的片段，确定了基

因的化学本质。

1957年,本泽尔(Seymour Benzer)以T4噬菌体为材料,在DNA分子水平上研究基因内部的精细结构,提出了顺反子(cistron)概念。

顺反子是1个遗传功能单位,1个顺反子决定1条多肽链。能产生1条多肽链的是1个顺反子,顺反子也就是基因的同义词。1个顺反子可以包含一系列突变单位——突变子。由于基因内的各个突变子之间有一定距离,所以彼此之间能发生重组,重组频率与突变子之间的距离成正比。重组子代表1个空间单位,有起点和终点,可以是若干个密码子的重组,也可以是单个核苷酸的互换。如果是后者,重组子也就是突变子。

4. 现代基因阶段

(1) 操纵子 从分子水平来看,基因就是DNA分子上的一个个片段,经过转录和翻译能合成1条完整的多肽链。可是,通过近年来的研究,认为这个结论并不全面,因为有些基因,如rRNA和tRNA基因只有转录功能而没有翻译功能。另外,还有一类基因,其本身并不进行转录,但可以对邻近的结构基因的表达起控制作用,如启动基因和操纵基因。从功能上讲,能编码多肽链的基因称为结构基因;启动基因、操纵基因和编码阻遏蛋白、激活蛋白的调节基因属于调控基因。操纵基因与其控制下的一系列结构基因组成1个功能单位,称为操纵子。

(2) 移动基因 移动基因指DNA能在有机体的染色体组内从一个地方跳到另一个地方,它们能从一个位点切除,然后插入同一或不同染色体上的另一个位置。移动基因机构简单,由几个促进移位的基因组成。基因的跳动能产生突变和染色体重排,进而影响其他基因的表达。

业已证明,相当一部分已知的自发突变是移动基因所致,而且,移动基因不仅能在个体的染色体组内移动,并能在个体间甚至种间移动,因此是一个重要的进化因素。移动基因的发现动摇了基因在染色体上有一固定位置的传统观念。

(3) 断裂基因 过去人们一直认为,基因的遗传密码子是连续不断地并列在一起,形成1条没有间隔的完整基因实体。但后来通过对真核蛋白质编码基因结构的分析发现,在它们的核苷酸序列中间插入有与编码无关的DNA间隔区,使1个基因分隔成不连续的若干区段。这种编码序列不连续的间断

基因被称为断裂基因。

不连续的断裂基因的表达程序是：先转录为初级转录物，即核内不均一 RNA，又称前体 RNA；然后经过删除和连接，除去无关的 DNA 间隔序列的转录物，便形成了成熟的 mRNA 分子，它从细胞核中输送到细胞质，再转译为相应的多肽链。

(4) 假基因 1977 年，G. Jacp 根据对非洲爪蟾 5S rRNA 基因簇的研究，提出了假基因的概念，现已在大多数真核生物中发现了假基因。这是一种核苷酸序列同其相应的正常功能基因基本相同，但却不能合成出功能蛋白质的失活基因。

(5) 重叠基因 长期以来，在人们的观念中一直认为同一段 DNA 序列内，是不可能存在重叠的读码结构的。但是，随着 DNA 序列测定技术的发展，人们已经在一些噬菌体和动物病毒中发现，不同基因的核苷酸序列有时是可以共用的。也就是说，它们的核苷酸序列是彼此重叠的，这样的两个基因被称为重叠基因。它修正了关于各个基因的多核苷酸链是彼此分立、互不重叠的传统观念。

由此可见，随着生物科学的不断发展，人们对基因概念的理解也不断深入。在世界科学技术日新月异的今天，生物科学将会有更多新的突破性进展，基因的概念不可避免地将会被赋予新的内容。

二、基因的分类

20 世纪 60 年代初 F·雅各布和 J·莫诺发现了调节基因。根据操纵子学说，并不是所有的基因都能编码多肽链。于是便把能编码多肽链的基因称为结构基因，包括编码结构蛋白和酶蛋白的基因，也包括编码阻遏蛋白或激活蛋白的调节基因。有些基因只能转录而不能翻译，如 tRNA 基因和 rRNA 基因。还有些 DNA 区段，其本身并不进行转录，但对其邻近的结构基因的转录起控制作用，被称为启动基因和操纵基因。启动基因、操纵基因与其控制下的一系列结构基因组成一个功能单位叫做操纵子 (operon)。就其功能而言，调节基因、操纵基因和启动基因都属于调控基因。这些基因的发现，大大拓宽了人们对基因功能及相互关系的认识。

但是从基因的原初功能这一角度来看，它们都是编码蛋白质。根据原初

功能（即基因的产物）基因可分为：①编码蛋白质的基因，包括编码酶和结构蛋白的结构基因以及编码作用于结构基因的阻遏蛋白或激活蛋白的调节基因。②没有翻译产物的基因，转录成为 RNA 以后不再翻译成为蛋白质的转移核糖核酸（tRNA）基因和核糖体核酸（rRNA）基因。③不转录的 DNA 区段，如启动区、操纵基因等等。前者是转录时 RNA 多聚酶开始和 DNA 结合的部位；后者是阻遏蛋白或激活蛋白和 DNA 结合的部位。

三、基因的结构

无论是结构基因还是调节基因，都能够储存、传递和表达遗传信息，也都可能发生突变，从而决定生物体的性状。基因之所以能够行使这些重要功能，是与它的结构有密切关系的。那么，基因的结构究竟是怎样的呢？

早在 20 世纪 50 年代，科学家们就开始了对于基因内部结构的研究，但是直到 70 年代中期，随着 DNA 分析技术的发展，人们才对基因的结构有所认识。

原核细胞的基因结构

原核细胞的基因是由成百上千个核苷酸对组成的。组成基因的核苷酸序列可以分为不同的区段。在基因表达的过程中，不同区段所起的作用并不相同。有的区段能够直接编码蛋白质，这样的区段叫做编码区（如图 2-9）。有的区段不能直接编码蛋白质，这样的区段叫做非编码区。

非编码区是由编码区上游（如图 2-9 中编码区左侧）和编码区下游（图中编码区右侧）的 DNA 序列组成的。非编码区虽然不能编码蛋白质，但是对于遗传信息的表达是不可缺少的。这是因为在非编码区上，有调控遗传信息表达的核苷酸序列，主要包括启动子、终止子等。

启动子位于编码区上游紧靠着转录起点的位置。在原核生物中一般只有

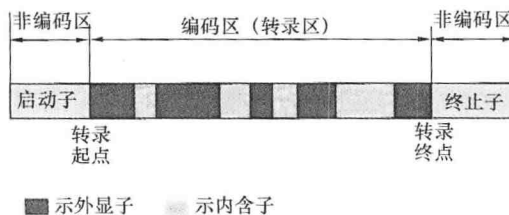


图 2-9 原核基因的基本结构

数十个碱基对，它的主要功能是引导 RNA 聚合酶与基因的正确部位结合。也就是说，只有在启动子的作用下，RNA 聚合酶才能准确地从转录起始点开始，沿着编码区正常地进行转录。终止子位于编码区下游紧靠着转录终点的位置。它的特殊的碱基排列顺序能够阻碍 RNA 聚合酶的移动，并使其从 DNA 模板链上脱离下来。

真核细胞的基因结构

真核细胞结构基因含有 4 个区域：①编码区，包括外显子与内含子；②前导区，位于编码区上游，相当于 RNA5'末端非编码区（非翻译区）；③尾部区，位于 RNA3'编码区下游，相当于末端非编码区（非翻译区）；④调控区，包括启动子和增强子等。基因编码区的两侧也称为侧翼顺序（图 2-10）。

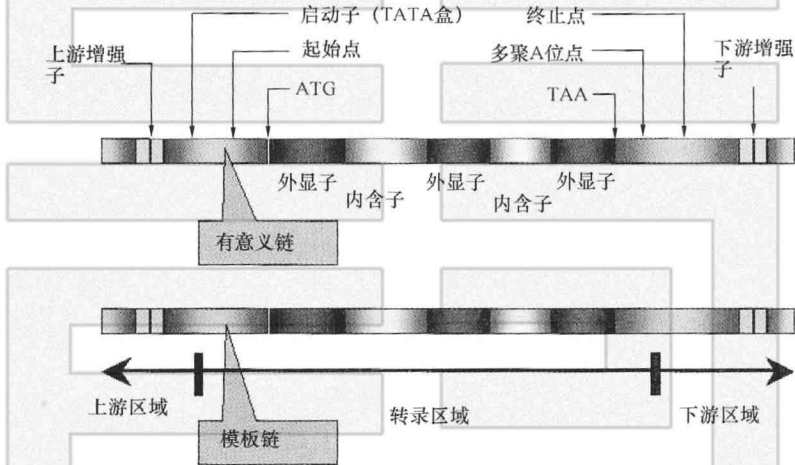


图 2-10 一个开放阅读框起始于 AUG，以三联体形式延续至终止密码，而受阻阅读框是因终止密码的频繁插入而形成的

1. 外显子和内含子：大多数真核生物的基因为不连续基因（interrupted or discontinuous gene）。所谓不连续基因就是基因的编码顺序在 DNA 分子上是不连续的，被非编码顺序所隔开。编码的顺序称为外显子（exon），是一个基因表达为多肽链的部分；非编码顺序所称为内含子（intron），又称插入顺序（intervening sequence, IVS）。内含子只转录，在前 mRNA（pre-mRNA）时被剪切掉。如果一个基因有 n 个内含子，一般总是把基因的外显子分

隔成 $n+1$ 部分。内含子的核苷酸数量可比外显子多许多倍。

内含子是相对的，一个基因的内含子可能是另一个基因的外显子。

如 SV40 病毒早期转录区编码了 T 和 t 两个功能不同的蛋白，这两个基因是重叠的，仅内含子不同。初始转录物的长度为 2530nt。大 T 蛋白成熟的 mRNA 中切除了 330nt 的内含子，余下的编码序列为 2200nt，翻译成大 T 蛋白的分子量为 100KDa。但小 t 蛋白内含子仅 30nt，位于大 T 内含子区域内，大 T 内含子其余部分（300nt）对于小 t 来说却是外显子。所以小 t 蛋白的 mRNA 长为 2500nt，由于在其内含子前有一个终止密码子，所以能翻译的部分并不长，小 t 蛋白的分子量也只有 18KDa。

2. 外显子—内含子接头：每个外显子和内含子接头区都有一段高度保守的一致顺序（consensus sequence），即内含子 5' 末端大多数是 GT 开始，3' 末端大多是 AG 结束，称为 GT—AG 法则，是普遍存在于真核基因中 RNA 剪接的识别信号。

3. 侧翼序列：在第一个外显子和最末一个外显子的外侧是一段不被翻译的非编码区，称为侧翼序列（flanking sequence）。侧翼序列含有基因调控顺序，对该基因的活性有重要影响。

(1) 启动子：启动子（promoter）包括下列几种不同顺序，能促进转录过程：

① TATA 框（TATA box）：其一致顺序为 TATAATAAT，它约在基因转录起始点上游约—30—50bp 处，基本上由 A—T 碱基对组成，是决定基因转录起始的，为 RNA 聚合酶的结合处之一，RNA 聚合酶与 TATA 框牢固结合之后才能开始转录。

② CAAT 框（CAAT box）：其一致顺序为 GGGTCAATCT，是真核生物基因常有的调节区，位于转录起始点上游约—80—100bp 处，可能也是 RNA 聚合酶的一个结合处，控制着转录起始的频率。

③ GC 框（GC box）：有两个拷贝，位于 CAAT 框的两侧，由 GGCGGG 组成，是一个转录调节区，有激活转录的功能。

此外，RNA 聚合酶 III 负责转录 tRNA 的 DNA 和 5SrDNA，其启动子位于转录的 DNA 顺序中，称为下游启动子。

④ 增强子：在真核基因转录起始点的上游或下游，一般都有增强子（en-

hander)，它不能启动一个基因的转录，但有增强转录的作用。此外，增强子顺序可与特异性细胞因子结合而促进转录的进行。研究表明，增强子通常有组织特异性，这是因为不同细胞核由不同的特异因子与增强子结合，从而对基因表达有组织、器官、时间不同的调节作用。

四、基因表达

通过比较基因的核苷酸序列和蛋白质的氨基酸序列，可以直接判定基因和蛋白质是不是共线性 (Colinear)，即核苷酸序列与蛋白质中氨基酸序列恰好吻合。在细菌及其病毒中，二者正好吻合，每一个基因包括一条 DNA，其长度与它表达的蛋白质中的氨基酸个数正相关。一个 $3N$ bp 长的基因可以编码含有 N 个氨基酸的蛋白质，符合遗传密码。

但是，基因并不是直接翻译成蛋白质的，而是通过一种直接用合成蛋白质的核酸序列——mRNA (Messenger RNA) 进行表达。信使 mRNA 与 DNA 双链中的一条互补，与另一条链相同 (只是将 T 用 U 替换)。通常在表示 DNA 时，是按 5'—3' 顺序书写与 RNA 相同的那条链。信使 RNA 包括一列与蛋白质中的氨基酸相应的核苷酸序列。这部分核苷酸称为编码区 (Coding region)。但信使 mRNA 还包括两边的额外序列，这些序列不编码蛋白质，称为 5' 非编码区和 3' 非编码区。

基因曾被认为是包括在 mRNA 中的所有序列。但有时会在额外的非编码区发现影响基因功能的突变，表明这些区域亦是遗传单位不可分割的组成部分。基因可认为是蛋白质翻译必须的连续长串的 DNA 序列，包括编码蛋白质的编码区与其两端的序列 (图 2-11)。

基因产生蛋白质的过程称为基因表达 (Expression)。基因表达的第一步称为转录 (Transcription)，合成一条链的复制版 RNA。最简单的基因 (通常在细菌中) 中，其 RNA 实际上就是 mRNA。真核基因的转录的产物是前体 mRNA，它随后通过一定程序产生成熟 mRNA (图 2-12)。

RNA 修饰中最重要的一个步骤是 RNA 剪接 (RNA splicing)。真核生物中许多基因 (高等真核基因的大部分) 包含不编码蛋白质的中间部分。剪接过程将前体 mRNA 的这些区域剪切掉，产生包含连续开放阅读框架的 RNA。此外，RNA 修饰过程还包括前体 mRNA 5' 和 3' 端的修饰。转录和修饰发生

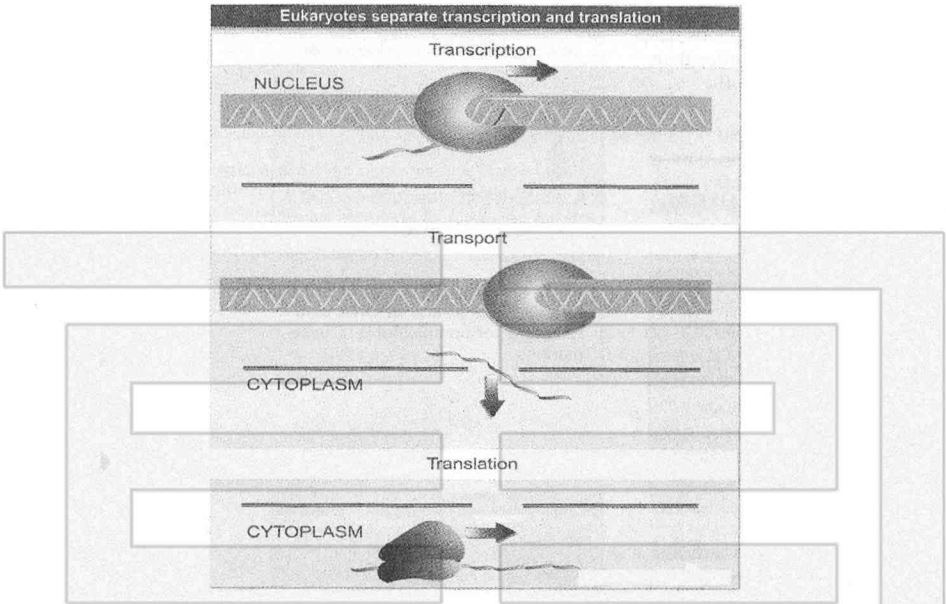


图 2-13 真核细胞的转录发生在细胞核中，而翻译发生在细胞质中

第五节 基因突变

一、基因突变的概念

突变 (mutation) 是遗传物质中任何可检测的能遗传的改变，但不包括遗传重组。突变是可以传递给子细胞，甚至延续给后代，从而导致产生了突变细胞或个体。对于一个多细胞生物来说，如果突变仅发生在体细胞中，那么这种突变是不会传递给后代的。这种类型的突变称为体细胞突变 (somatic mutation)。但若突变发生在生殖细胞中，那么这种突变就能通过配子传递给下一代，在后代个体的体细胞和生殖细胞中产生同样的突变，这种突变就叫做种系突变 (germ-line mutations)。由于遗传物质一般是 DNA，突变会影响到 DNA 的化学或物理的构成，它的复制，表型功能或者一个或多个碱基对序列会发生改变，例如增加、减少或置换碱基，颠倒顺序或转移到新的位

置上。

突变可以发生在染色水平或者基因水平。染色体结构和数目的改变叫做染色体畸变 (chromosomal aberration), 也称为染色体突变 (chromosome mutation)。当染色体畸变涉及基因组中染色体套数的改变称为基因组突变 (genome mutation), 即整倍数改变。发生在基因水平的突变称为基因突变 (gene mutation), 它涉及基因的一个或多个序列的改变, 包括一对或多对碱基对的替换, 增加或缺失。由于 DNA 碱基对的改变引起的基因突变称为点突变 (point mutation)

突变可以是自发的, 但也可被某些诱变剂 (mutagen) 诱发。一些物理因素或化学试剂都能增加这种自发突变的频率。而诱变剂处理所诱发的突变称为诱发突变 (induced mutation), 自然发生的突变是自发突变 (spontaneous mutation)。这两种突变之间并没有本质的区别。突变表型显示出 DNA 改变所产生的后果, 这是由于蛋白功能改变的结果。

二、基因突变的特性

整个生物界, 由于它们的遗传物质是相同的, 所以显示在遗传变异特性上都遵循着共同的规律, 这在基因突变的水平上尤为明显。基因突变一般有 7 个共同特点:

1. 不对应性

即突变的性状与引起突变的原因间无直接的对应关系。例如, 细菌在有青霉素的环境下, 出现了抗青霉素的突变体; 在紫外线的作用下, 出现了抗紫外线的突变体; 在较高的培养温度下, 出现了耐高温的突变体等。从表面上看, 会认为正是由于青霉素、紫外线或高温的“诱变”, 才产生了相对应的突变性状。事实恰恰相反, 这类性状都可通过自发的或其他任何诱变因子诱发得到。这里的青霉素、紫外线或高温仅是起着淘汰原有非突变型 (敏感型) 个体的作用。

2. 自发性

由于自然界环境因素的影响和微生物内在的生理生化特点, 在没有人诱发因素的情况下, 各种遗传性状的改变可以自发地产生。

3. 稀有性

指自发突变的频率较低，而且稳定，一般在 $10^{-6} \sim 10^{-9}$ 间。

4. 独立性

突变的发生一般是独立的，即在某一群体中，既可发生抗青霉素的突变型，也可发生抗链霉素或任何其他药物的抗药性。某一基因的突变，既不提高也不降低其他任何基因的突变率。突变不仅对某一细胞是随机的，且对某一基因也是随机的。

5. 可诱变性

通过各种物理、化学诱变剂的作用，可提高突变率，一般可提高 $10 \sim 10^5$ 倍。

6. 稳定性

由于突变的根源是遗传物质结构上发生了稳定的变化，所以产生的新性状也是稳定的和可遗传的。

7. 可逆性

由原始的野生型基因变异为突变型基因的过程称为正向突变，相反的过程则称为回复突变。实验证明，任何性状既有可能正向突变，也有可能发生回复突变，两者发生的频率基本相同。

三、基因突变的机制和种类

从 DNA 碱基顺序改变来分，突变一般可分为碱基置换突变、移码突变、整码突变、染色体错误配对和不等交换四种。

1. 碱基置换突变

一个碱基被另一碱基取代而造成的突变称为碱基置换突变（图 2-14）。凡是一个嘌呤被另一个嘌呤所取代，或者一个嘧啶被另一个嘧啶所取代的置换称为转换（transition）；一个嘌呤被另一个嘧啶所取代或一个嘧啶被另一个嘌呤所替代的置换称为颠换（transversion）。由此可产生 4 种不同的转换和 8 种不同的颠换。但自然界的突变，转换多于颠换。碱基置换会导致蛋白一级结构氨基酸组成的改变而影响蛋白质酶生物的功能。

由于碱基置换导致核苷酸顺序的改变，对多肽链中氨基酸顺序的影响，有下列几种类型；

(1) 同义突变 由于密码子具有兼并性，因此，单个碱基置换后使 mR-

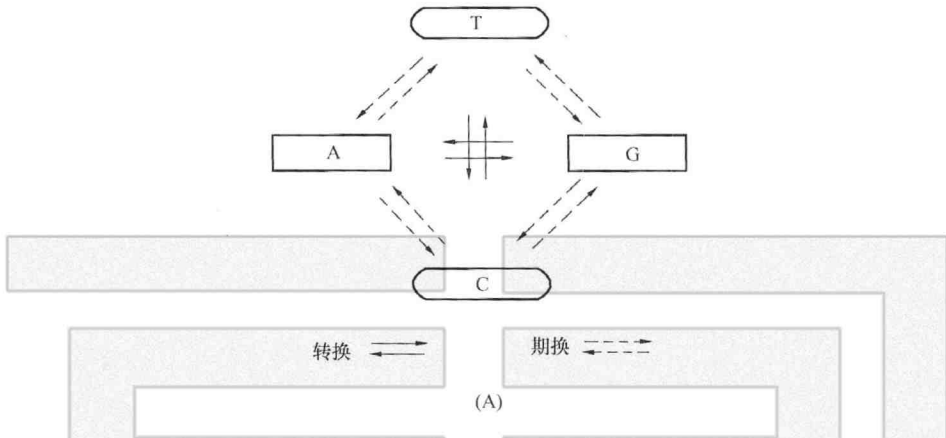


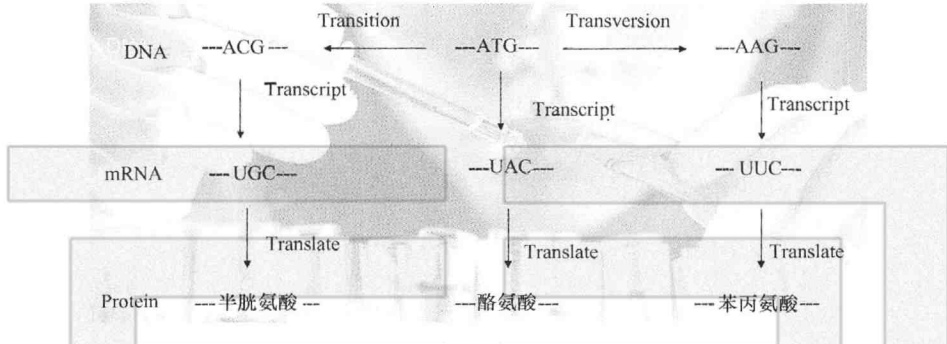
图 2-14 碱基置换示意图

NA 上改变后的密码子与改变前所编码的氨基酸一样，肽链中出现同一氨基酸。例如 DNA 分子模板链中 GCG 的第三位 G 被 A 取代而成 GCA，则 mRNA 中相应的密码子 CGC 就被转录为 CGU，CGC 和 CGU 都是精氨酸的密码子，翻译成的多肽链没有变化，这种突变称为同义突变 (same-sense or synonymous mutation)。同义突变不易检出。据估计，自然界中这样的突变频度占相当高比例。

DNA	--- GCA ---	Transition	--- GCG ---	Transversion	--- GCC ---
mRNA	--- CGU ---		--- CGC ---		--- CGG ---
多肽链	--- 精氨酸 ---		--- 精氨酸 ---		--- 精氨酸 ---

(2) 错义突变 是指 DNA 分子中的核苷酸置换后改变了 mRNA 上遗传密码，从而导致合成的多肽链中一个氨基酸被另一氨基酸所取代，这种情况称为错义突变 (missense mutation)。此时，在该氨基酸前后的氨基酸不改变。例如 mRNA 分子正常编码顺序为：UAU (酪) GCC (丙) AAA (赖) UUG (亮) AAA (赖) CCA (脯)，当第三密码子 A 颠换为 C 时，则 AAA (赖) → ACA (苏)，即上述顺序改变为 UAU (酪) GCC (丙) ACA (苏) UUG (亮) AAA (赖) CCA (脯)。错义突变结果产生异常蛋白质和酶。但也有不少基因由于错义突变而产生部分降低活性和异质组分的酶，从而不完全抑制了催化反应，这种基因称为漏出基因 (leaky gene)。如果由于基因错义突变置换了酶活性中心的氨基酸，因此合成了没有活性的酶蛋白，虽不具有酶活性但有时还具有蛋白质抗原性，其所产生的抗体可与正常蛋白质发生

交叉反应。有些错义突变不影响蛋白质或酶的生物活性，因而不表现出明显的表型效应，这种突变可称为中性突变（neutral mutation）。



(3) 无义突变 当单个碱基置换导致出现终止密码子（UAG、UAA、UGA）时，多肽链将提前终止合成，所产生的蛋白质（或酶）大都失去活性或丧失正常功能，此种突变称为无义突变（non-sense mutation）。例如，DNA 分子模板链中 ATG 的 G 被 T 代替时，相应的 mRNA 上的密码子便从 UAC 变成终止信号 UAA，因此翻译便到此为止，使多肽链缩短。无义突变如果发生在靠近 3' 末端处，它所产生的多肽链常有一定的活性，表现为渗漏型，这类多肽多半具有野生型多肽链的抗原特异性。



(4) 当基因内部不同位置上的不同碱基发生了两次突变，其中一次抑制了另一次突变的遗传效应，这种突变称为抑制基因突变（suppressor gene mutation）。例如 Hb Harlem 是 β 链第 6 位谷氨酸变成缬氨酸，第 73 位天冬氨酸变成天冬酰胺；如果单纯 $\beta 6$ 谷氨酸 \rightarrow 缬氨酸，则可产生 HbS 病，往往造成死亡。但 Hb Harlem 临床表现却较轻，即 $\beta 73$ 的突变抑制了 $\beta 6$ 突变的有害效应。

2. 移码突变

移码突变（frame-shift mutation）是指 DNA 链上插入或丢失 1 个、2 个甚至多个碱基（但不是三联体密码子及其倍数），在读码时，由于原来的密码子移位，导致在插入或丢失碱基部位以后的编码都发生了相应改变。移码突变造成的肽链延长或缩短，取决于移码终止密码子推后或提前出现。

(1) 插入突变：在基因中插入一个或几个碱基而造成的突变。



(2) 缺失突变：在基因中丢失一个或几个碱基而造成的突变。



3. 染色体错误配对不等交换

染色体错误配对不等交换 (mispaired synapsis and unequal crossing-over)：减数分裂期间，同源染色体间的同源部分发生联会和交换，如果联会时配对不精确，会发生不等交换，造成一部分基因缺失和部分基因重复。这种突变常用解释大段多核苷酸的丢失和重复 (图 2-15)。

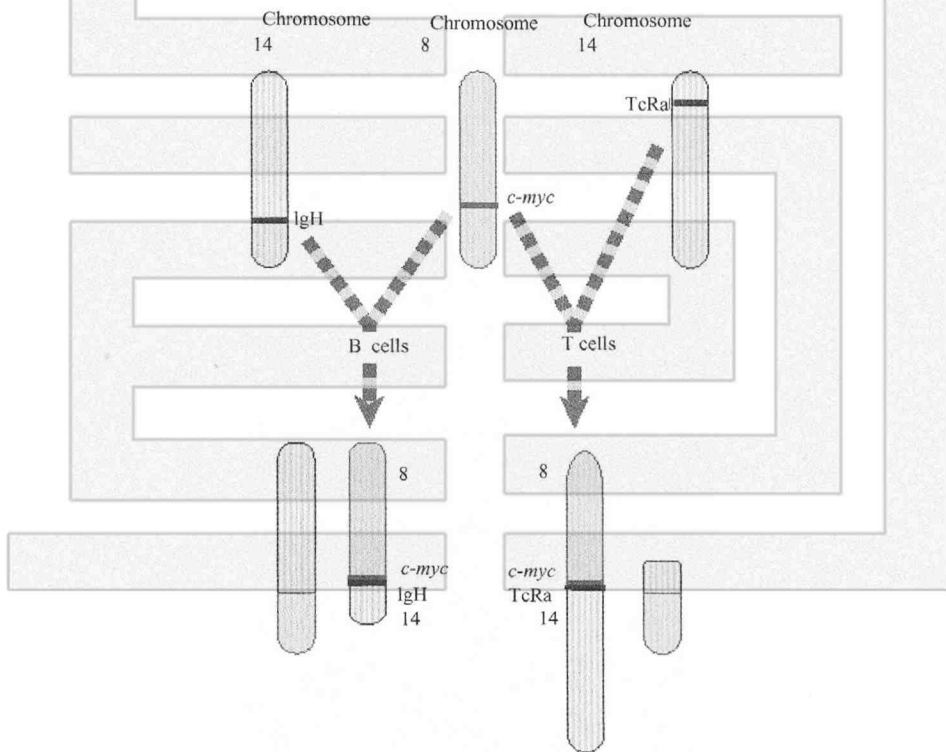


图 2-15 染色体的大片段交换

四、调控基因突变对结构基因表达的影响

所有细胞都是全能核（携带全部遗传信息），但不是全部基因都有活性，所以必定有一种抑制某些基因活性和启动另一些基因活性的机制。对于基因调控机制，1961年 Jacob 与 Monod 对大肠杆菌的研究提出了乳糖操纵子假说（Lac operon hypothesis），认为基因的作用单位是操纵子（operon），它由一个操纵基因和相邻的结构基因构成，它们按一定的线性顺序排列，并产生一系列相关的酶。操纵基因可以启动全组结构基因的活性，但它又被调节基因激活或抑制。调节基因能合成一种物质（阻遏物），抑制操纵基因；当调节基因起作用时，有关的结构基因不合成蛋白质，只有在阻遏物被一种特殊代谢物（诱导物）灭活后，调节基因在关闭的情况下，结构基因才起作用。真核细胞的基因调控还未完全阐明。如果这一模式能应用在人类，即假设有不缺乏一种相关酶的那些遗传病，有可能是由于调控系统基因突变的结果。又如有些酶活性缺乏或增加，或蛋白质合成量有改变，但从结构基因水平并未发现有任何碱基改变，这样推测突变可能发生在调控基因部分，例如腺苷脱氨酶（adenosine deaminase, ADA）遗传性酶活性过高（相当于正常水平的45—70倍）可引起溶血，但该突变酶结构没有改变，而转录的 mRNA 大大增多，故认为是调控基因突变的结果。又如 Crigler—Najjar 综合征Ⅱ型，表现为先天性黄疸，为肝葡萄糖醛酰转移酶缺乏，血中非结合胆红素增高。如用苯巴比妥可诱导此酶活性升高，黄疸消失，故认为此病可能是调节失控所致。

如果调节基因突变失去活性，基因不再被控制，结果蛋白质合成就会增加。在杂合子中，由于同源染色体上的正常调节基因所产生的阻遏物足够抑制两条染色体上的蛋白质合成，所以只有在纯合子中才表现出来。另一种情况是，由于调节基因发生突变，合成了异常的阻遏物，它不能被诱导物所灭活，导致蛋白质合成减少。由于阻遏物是可以扩散的，它将有可能在同源染色体上起作用，因此杂合子即可表现出来。因此，尽管一般认为，遗传性代谢缺陷是由于结构基因突变造成的，但应考虑到调控基因突变也可引起表型相同的遗传病。近年来分子遗传学的发展，已从 DNA 顺序的改变证明了这一点，例如地中海贫血有些突变就发生在调控基因部分。

五、基因突变的后果

根据基因突变对机体影响的程度，可分为下列几种情况：

1. 变异后果轻微，对机体不产生可察觉的效应。从进化观点看，这种突变称为中性突变。

2. 造成正常人体生物化学组成的遗传学差异，这种差异一般对人体并无影响。例如血清蛋白类型、ABO 血型、HLA 类型以及各种同工酶型。但在某种情况下也会发生严重后果。例如不同血型间输血，不同 HLA 型间的同种移植产生排斥反应等。

3. 可能给个体的生育能力和生存带来一定的好处。例如，HbS 突变基因杂合子比正常的 HbA 纯合子更能抗恶性疟疾，有利于个体生存。

4. 产生遗传易感性 (genetic susceptibility)。

5. 引起遗传性疾病，导致个体生育能力降低和寿命缩短，这包括基因突变导致蛋白质异常的分子病及遗传酶病。据估计，人类有 50000 个结构基因，正常人的基因座位处于杂合状态的可占 18%，一个健康人至少带有 5—6 个处于杂合状态的有害突变，这些突变如在纯合状态时就会产生有害后果。

6. 致死突变，造成死胎、自然流产或出生后夭折等。

第三章 中国少数民族 遗传学的细胞基础

细胞是生命活动的基本结构和功能单位，细胞的特殊性决定了个体的特殊性，因此，对细胞的深入研究是揭开生命奥秘、改造生命和征服疾病的关键。一般来说，细菌等绝大部分微生物以及原生动物的一个细胞组成，即单细胞生物；高等植物与高等动物则是多细胞生物。细胞可分为两类：原核细胞和真核细胞。真核细胞由细胞膜、细胞质和细胞核三部分组成。细胞核通常位于细胞的中央。细胞核中有一种物质，易被洋红、苏木精、甲基绿等碱性染料染成深色，叫做染色质(chromatin)。在细胞周期变化中染色质和染色体交替改变，在间期以染色质的形式存在，而在分裂期以染色体的形式存在。染色质和染色体是遗传物质的载体。

第一节 染色质(chromatin)和染色体(chromosome)

染色质和染色体是由 DNA、组蛋白、非组蛋白及少量 RNA 组成的核蛋白复合物，是核基因的载体，是同一物质在细胞周期的不同阶段所表现出的两种不同的存在形式。在间期细胞核中伸展呈细丝状、易被碱性染料着色的，称为染色质；在细胞分裂期，染色质高度螺旋化，折叠而缩短变粗，成为染色体。

一、染色质

染色质由 DNA、组蛋白、非组蛋白及少量 RNA 组成，比例为 1 : 1 : (1 - 1.5) : 0.05。可见 DNA 与组蛋白的含量比较恒定，非组蛋白的含量变化较大，RNA 含量最少。

1. 组蛋白

组蛋白带正电荷，含精氨酸、赖氨酸，属碱性蛋白，其含量恒定，在真核细胞中组蛋白共有 5 种，分为两类：

一类是高度保守的核心组蛋白（core histone）包括 H₂A、H₂B、H₃、H₄ 四种；另一类是可变的连接组蛋白（linker histone）即 H₁。

核心组蛋白的结构非常保守，特别是 H₄，牛和豌豆 H₄ 的 102 个氨基酸中仅有 2 个不同，而进化上两者分歧的年代约 3 亿年历史。核心组蛋白高度保守的原因可能有两个：其一是核心组蛋白中绝大多数氨基酸都与 DNA 或其他组蛋白相互作用，可置换而不引起致命变异的氨基酸残基很少；其二是在所有的生物中与组蛋白相互作用的 DNA 磷酸二酯骨架都是一样的。

四种核心组蛋白均由球形部和尾部构成，球形部借精氨酸残基与磷酸二酯骨架间的静电作用使 DNA 分子缠绕在组蛋白核心周围，形成核小体。尾部则含有大量赖氨酸和精氨酸残基，为组蛋白翻译后进行修饰的部位，如乙酰化、甲基化、磷酸化等。

H₁ 不仅具有属特异性，而且还有组织特异性，所以 H₁ 是多样性的。

2. 非组蛋白

与组蛋白不同，非组蛋白是染色体上与特异 DNA 序列结合的蛋白质，所以又称序列特异性 DNA 结合蛋白。非组蛋白的特性是：①含有较多的天门冬氨酸、谷氨酸，带负电荷，属酸性蛋白质；②整个细胞周期都进行合成，不像组蛋白只在 S 期合成，并与 DNA 复制同步进行；③能识别特异的 DNA 序列，识别信息存在于 DNA 本身，位点在大沟部分，识别与结合借助于氢键和离子键。

非组蛋白的功能是：①帮助 DNA 分子折叠，以形成不同的结构域，从而有利于 DNA 的复制和基因的转录；②协助启动 DNA 复制；③控制基因转录，调节基因表达。

3. 核小体

人体的一个细胞核中有 23 对染色体，每条染色体的 DNA 双螺旋若伸展开，平均长为 5cm，核内全部 DNA 联结起来约 1.7~2.0m，而细胞核直径不足 10μm。因此，不难想象 DNA 是以螺旋和折叠的方式压缩起来的，压缩比例高达上万倍，这种压缩的最初级结构就是核小体。

用非特异性核酸酶（如微球菌核酸酶）处理染色质，大多数情况下可得到大约 200bp 的片段，但处理裸露的 DNA 分子会得到随机降解的片段。以这个实验为基础，R. Kornberg 建立了核小体模型。

核小体是一种串珠状结构，由核心颗粒和联结线 DNA 两部分组成，可描述如下（图 3-1）：①每个核小体单位包括约 200bp 的 DNA、1 个组蛋白核心和 1 个 H_1 ；②由 H_2A 、 H_2B 、 H_3 、 H_4 各两分子形成八聚体，构成核心颗粒；③DNA 分子以左手螺旋缠绕在核心颗粒表面，每圈 80bp，共 1.75 圈，约 146bp，两端被 H_1 锁合；④相邻核心颗粒之间为一段 60bp 的连接线 DNA。

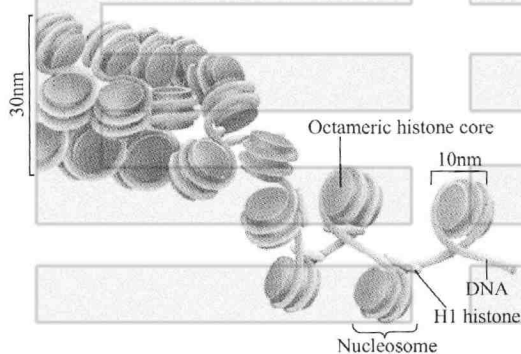


图 3-1 核小体和螺线管的结构

通过核小体，DNA 长度压缩了 7 倍，形成直径为 11nm 的纤维（图 3-2）。但是染色质不以这种状态存在，在电镜下观察用温和方法分离的染色质是直径 30nm 的纤维，这种纤维的形成有两种解释：①由核小体螺旋化形成，每 6 个核小体绕一圈，

构成外径 30nm 的中空管，长度压缩 6 倍；②由核小体纤维 Z 字形折叠而成，长度压缩 40 倍。30nm 纤维的形成与核小体之间蛋白质的相互作用有关，连接组蛋白 H_1 和核心组蛋白尾部均参与这种相互作用，去除组蛋白 H_1 的染色质中，30nm 纤维解体为更细的纤维。

对于更高级染色体包装方式，至今尚不明确。目前多认为 30nm 的纤维折叠为一系列的环（loop）结合在核骨架上（或称染色体骨架），结合点是富含 AT 的区域，这种环状的结构散布于细胞核中。用盐溶液去除组蛋白，在电镜下可以看到，在有丝分裂的染色体、灯刷染色体、间期的唾腺染色体上，都有大量的结合在骨架上的放射环，说明这种环并不是有丝分裂染色体所特有的。

4. 常染色质和异染色质

间期核中染色质可分为异染色质（heterochromatin）和常染色质（euchromatin）。常染色质是进行活跃转录的部位，呈疏松的环状，电镜下表现为

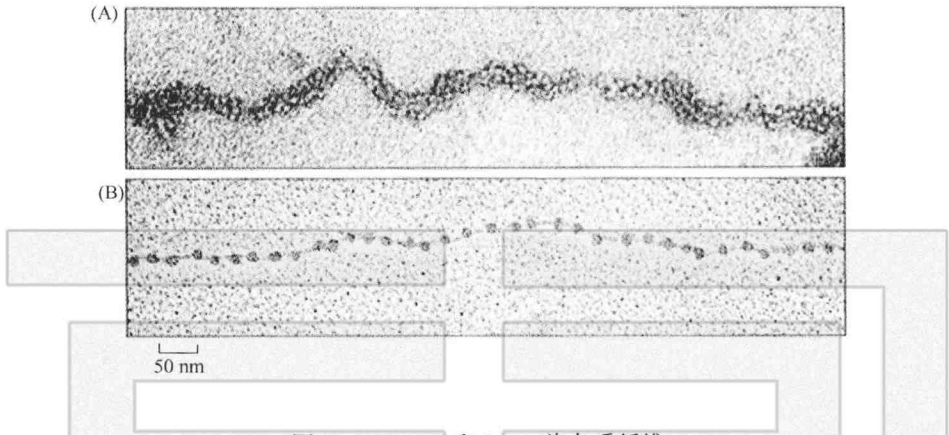


图 3-2 30nm 和 11nm 染色质纤维

浅染 (图 3-3); 易被核酸酶在一些敏感的位点 (hypersensitive sites) 降解。异染色质的特点是: 在间期核中处于凝缩状态, 无转录活性, 也叫非活动染色质 (inactive chromatin); 是遗传惰性区; 在细胞周期中表现为晚复制、早凝缩, 即异固缩现象 (heteropycnosis)。

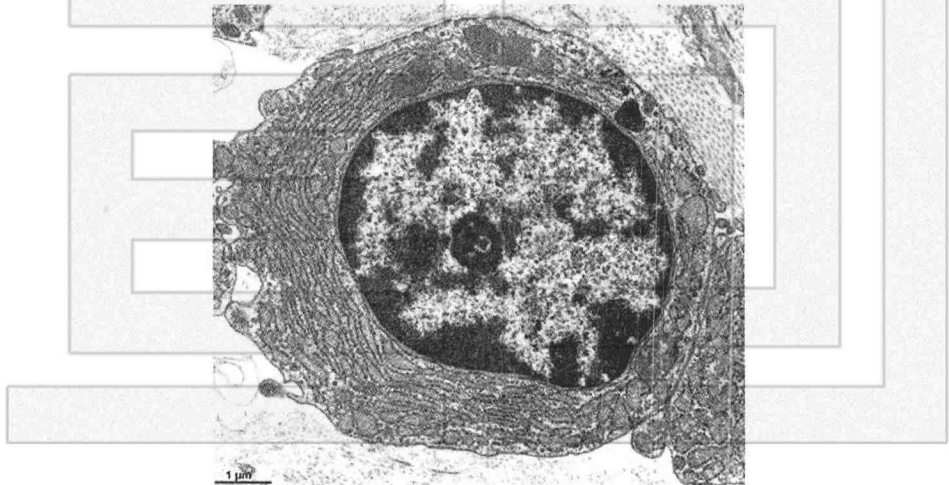


图 3-3 异染色质 (核内深染部分) 和常染色质
(核内浅染部分)

结构 (恒定) 异染色质 (constitutive heterochromatin): 在所有细胞内都呈异固缩的染色质, 多定位于着丝粒区 (图 3-4), 端粒、次缢痕及染色体臂的某些节段, 在间期聚集成多个染色中心 (chromocenter), 由相对简单的高

度重复序列构成。

兼性（功能）异染色质（facultative heterochromatin）：是指不同细胞类型或不同发育时期出现的异染色质区。雌性哺乳类动物的 X 染色体就是一类特殊的兼性异染色质。在哺乳动物细胞内如有两个 X 染色体（通常为雌性），而其中的一个染色体常表现为异染色质（图 3-5），称巴氏小体（barr body）。

人的胚胎发育到 16 天以后，一条 X 染色体转变为巴氏小体，呈块状紧靠核膜，染色反应表现为深染。因此通过检查羊水中胚胎细胞的巴氏小体可预测胎儿的性别。

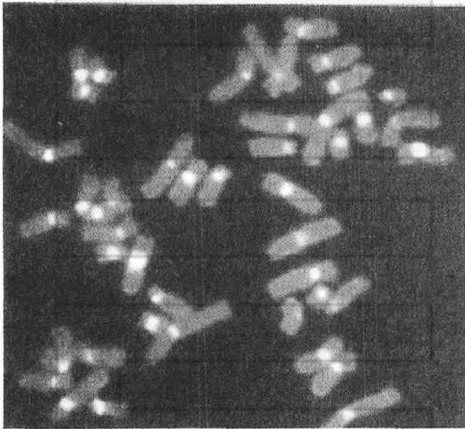


图 3-4 荧光原位杂交显示的
着丝粒卫星 DNA

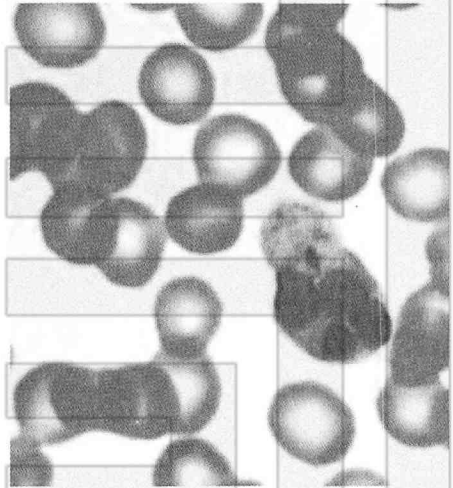


图 3-5 白细胞中的巴氏小体
(鼓槌状结构)

二、染色体

间期染色质分散于细胞核，但在分裂期，染色质通过盘旋折叠压缩近万倍，包装成大小不等、形态各异的短棒状染色体（图 3-6）。人类体细胞具有 46 条染色体，其中 44 条（22 对）为常染色体，另两条与性别分化有关，为性染色体。性染色体在女性为 XX，在男性为 XY。生殖细胞中卵细胞和精子各有 23 条染色体，分别为 $22+X$ 和 $22+Y$ 。中期染色体由于形态比较稳定，是观察染色体形态和计数的最佳时期。每一中期染色体都由两条染色单体构成，它们各含一条 DNA 双螺旋链。两条单体仅在着丝粒外互相连接，该处为

染色体的缩窄处，故又称为主缢痕。着丝粒是纺锤丝附着点，在细胞分裂中与染色体的运动密切相关，失去着丝粒的染色体片段通常不能在分裂后期向两极移动而丢失，着丝粒又将染色体横向地分为两个臂。

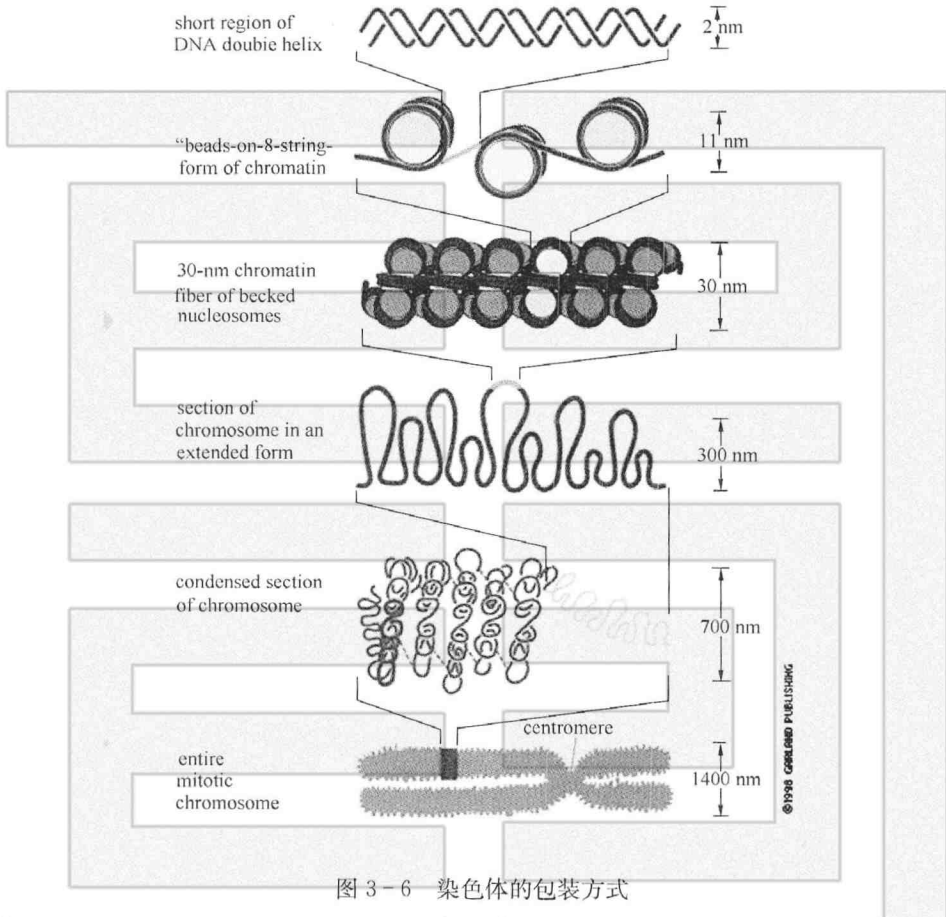


图 3-6 染色体的包装方式

同一物种的染色体数目是相对稳定的，性细胞染色体为单倍体 (haploid)，用 n 表示，体细胞为 2 倍体 (diploid) 以 $2n$ 表示，还有一些物种的染色体成倍增加成为 $4n$ 、 $6n$ 、 $8n$ 等，称为多倍体。同一个体的体细胞并非都是 2 倍体，如大鼠肝细胞有 $4n$ 、 $8n$ 、 $16n$ 等多倍体细胞，果蝇卵巢滋养细胞表现为 $2n$ 、 $4n$ 、 $8n$ 、 $16n$ 、 $32n$ 、 $64n$ 、 $128n$ 等不同倍性，人类子宫内膜细胞的染色体数目变化在 $2n=17-103$ ，为非整倍性。

1. 染色体形态和结构相关的术语 (图 3-7)

(1) 染色单体 (Chromatid)：中期染色体由两条染色单体组成，两者在

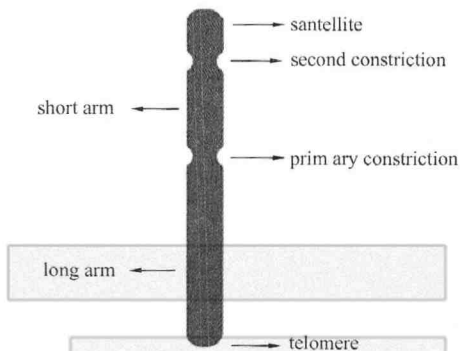


图 3-7 染色体各部分的名称

着丝粒的部位相互结合，每一条染色单体是由一条 DNA 双链经过螺旋和折叠而形成的，到后期，着丝粒分裂，两条染色单体分离。

(2) 染色线 (Chromonema): 前期或间期核内的染色质细线，代表一条染色单体。

(3) 染色粒 (Chromomere): 前期染色体上呈线性排列的念珠状颗粒，

是 DNA 局部收缩形成的，异染色质的染色粒一般较大，而常染色质的染色粒较小，在染色体上位于着丝粒两边的染色粒一般较大，而向染色体端部的染色粒较小，呈梯度排列。

(4) 主缢痕 (primary constriction): 中期染色体上一个染色较浅而缢缩的部位，主缢痕处有着丝粒，所以亦称着丝粒区。由于这一区域染色线的螺旋化程度低，DNA 含量少，所以染色很浅或不着色。一般动植物的染色体具有一个位置固定的着丝粒 (localized centromere)，有些生物整个染色体都具有着丝粒活性，称为全着丝粒 (holocentromere) 如：蛔虫、线虫、蝶、蛾、蚜虫等。可根据着丝粒的位置将染色体分为 4 类：即①中着丝粒染色体 (metacentric chromosome)，②亚中着丝粒染色体 (submetacentric chromosome)，③亚端着丝粒染色体 (subtelocentric chromosome)，④端着丝粒染色体 (telocentric chromosome)。划分的标准有：①臂比值 r (长臂长/短臂长)，②着丝粒指数 i [(短臂长/染色体长) $\times 100\%$]，③短臂长臂比等 3 种。

(5) 次缢痕 (secondary constriction): 除主缢痕外，染色体上第二个呈浅缢缩的部分称次缢痕，次缢痕的位置相对稳定，是鉴定染色体个别性的一个显著特征。

(6) 核仁组织区 (nucleolar organizing regions, NORs): 是核糖体 RNA 基因所在的区域 (图 3-8)，其精细结构呈灯刷状，能够合成核糖体的 28S、18S 和 5.8S rRNA。核仁组织区位于染色体的次缢痕区，但并非所有的次缢痕都是 NORs。

(7) 随体 (satellite): 指位于染色体末端的球形染色体节段，通过次缢

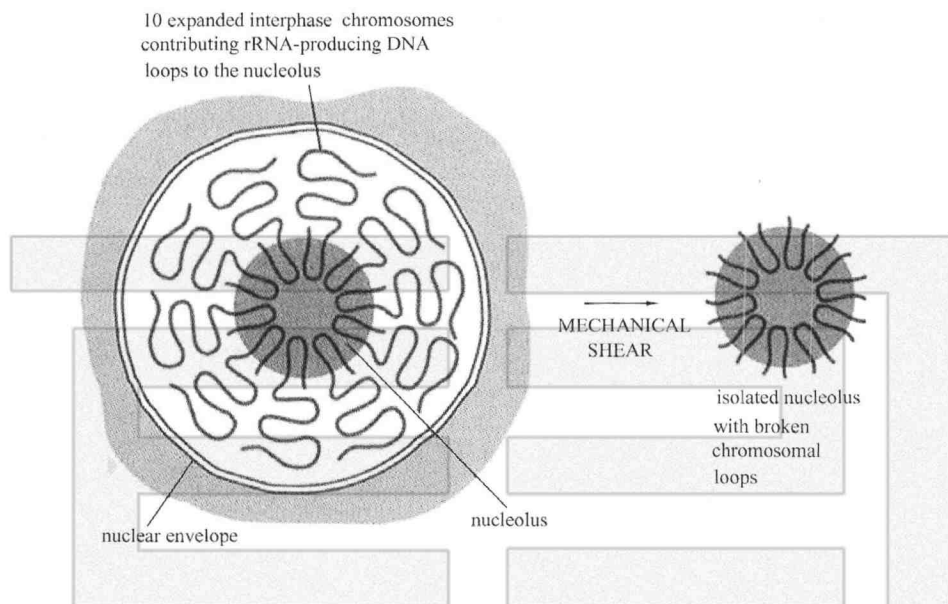
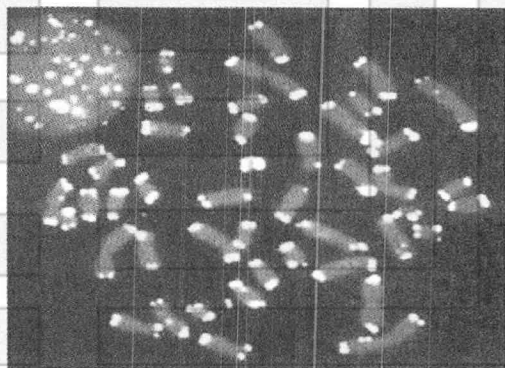


图 3-8 核仁组织者中心形成核仁

痕区与染色体主体部分相连。位于染色体末端的随体称为端随体，位于两个次缢痕中间的称中间随体。

(8) 端粒 (telomere): 是染色体端部的特化部分 (图 3-9), 其生物学作用在于维持染色体的稳定性。如果用 X 射线将染色体打断, 不具端粒的染色体末端有黏性, 会与其他片段相连或两端相连而成环状。端粒由高度重复的短序列串联而成, 在进化上高度保守, 不同生物的端粒序列都很相似, 人的序列为 TTAGGG。端粒起到细胞分裂计时器的作用, 端粒核苷酸复制和基因 DNA



Species	Telomere repeat sequence	Telomerase RNA template sequence
Human	TTAGGG	CUAACCCUAAAC
Tetrahymena	TTGGGG	CAACCCCAA
Arabidopsis	TTTAGGG	CUAAACCCUAAAC

图 3-9 荧光原位杂交显示的端粒(上)和端粒序列(下)

不同，每复制一次减少 50—100bp，其复制要靠具有反转录酶性质的端粒酶 (telomerase) 来完成，正常体细胞缺乏此酶，故随细胞分裂而变短，细胞随之衰老。

2. 着丝粒的结构

着丝粒 (centromere) 和着丝点 (kinetochore) 是两个不同的概念，前者指中期染色单体相互联系在一起的特殊部位，后者指主缢痕处两个染色单体外侧表层部位的特殊结构，它与纺锤丝微管相接触。

着丝粒含 3 个结构域 (图 3-10)：着丝点结构域 (kinetochore domain)、中心结构域 (central domain) 和配对结构域 (pairing domain)。

着丝点结构域 (图 3-11)：位于着丝粒的表面，由外板 (outer plate)、内板 (inner plate)、中间区 (interzone) 和围绕外层的纤维冠 (fibrous corona) 组成。内外板的电子密度高，中间区电子密度低。内板与中心结构域的着丝粒异染色质结合，外板与微管纤维结合，纤维冠上结合有马达蛋白，如胞质 Dynein 和属于 kinesin 家族的 CENP-E，为染色体的分离提供动力。

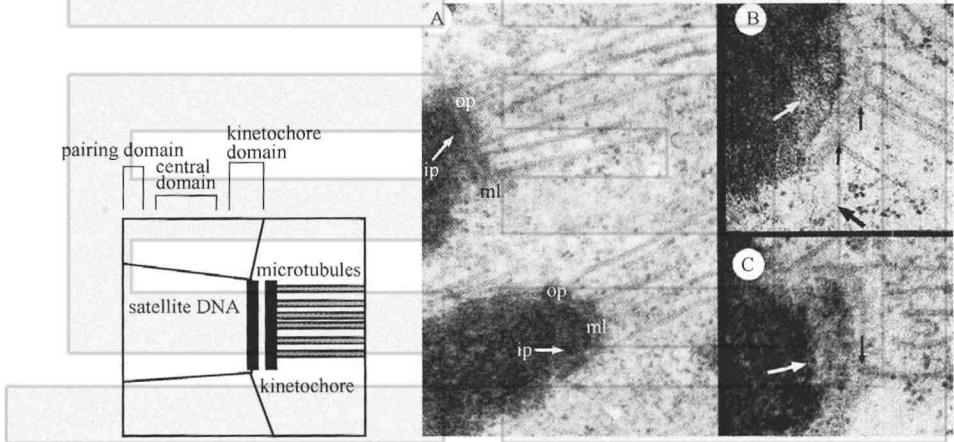


图 3-10 着丝粒的 3 个结构域

图 3-11 着丝点结构域

中心结构域：位于着丝粒结构域的下方，含有高度重复的 α 卫星 DNA 构成的异染色质。

配对结构域：位于着丝粒结构的内层，中期两条染色单体在此处相互联结，在此区域发现有两类蛋白，一类为内着丝粒蛋白 INCENP (inner centromere protein)，另一类为染色单体连接蛋白 CLIP (chromatid linking pro-

tein)。

对着丝粒蛋白主要是用 ACA 来研究的。ACA 是从 CREST 综合征病人血清中分离出来的抗着丝粒蛋白的抗体 (antacentromere antibodies)。用 ACAs 发现鉴定出来的 CENP 主要有 6 种, 即: CENP-A 至 F (表 3-1)。

表 3-1 几种着丝粒蛋白质的功能

类型	功能
CENP-A	着丝粒特异性组蛋白
CENP-B	与中心结构域的卫星 DNA 结合
CENP-C	与着丝点结合
CENP-D	与着丝点结合
CENP-E	驱动蛋白类分子马达
CENP-F	与着丝点结合
INCENP-A	连接姊妹染色单体
INCENP-B	连接姊妹染色单体

3. 核型与带型

早期对染色体的研究主要采用组织切片的方法, 由于染色体经常不处于同一平面, 因此很难对染色体进行观察和计数, 以至于对人类到底有多少条染色体这样的问题一直争论到 1956 年才得以确认, 这主要归功于 20 世纪 50 年代以后逐步发展起来的低渗处理、压片技术以及秋水仙来处理 and 细胞培养技术。20 世纪 60 年代末至 70 年代发展起来的各种染色体分带技术, 使染色体的研究进入了一个黄金时代, 为人类遗传病的鉴定、物种的血缘关系与进化研究、遗传育种等方面提供了重要的依据。

核型 (karyotype) 是细胞分裂中期染色体特征的总和, 包括染色体的数目、大小和形态特征等方面。如果将成对的染色体按形状、大小依顺序排列起来叫核型图 (karyogram, 图 3-12), 而染色体组型 idiogram 通常指核型的模式图, 代表一个物种的模式特征。

染色体分带技术是经物理、化学因素处理后, 再用染料对染色体进行分化染色, 使其呈现特定的深浅不同带纹 (band) 的方法。分带技术可分为两大类, 一类是产生的染色带分布在整过染色体的长度上如: G (图 3-13)、Q

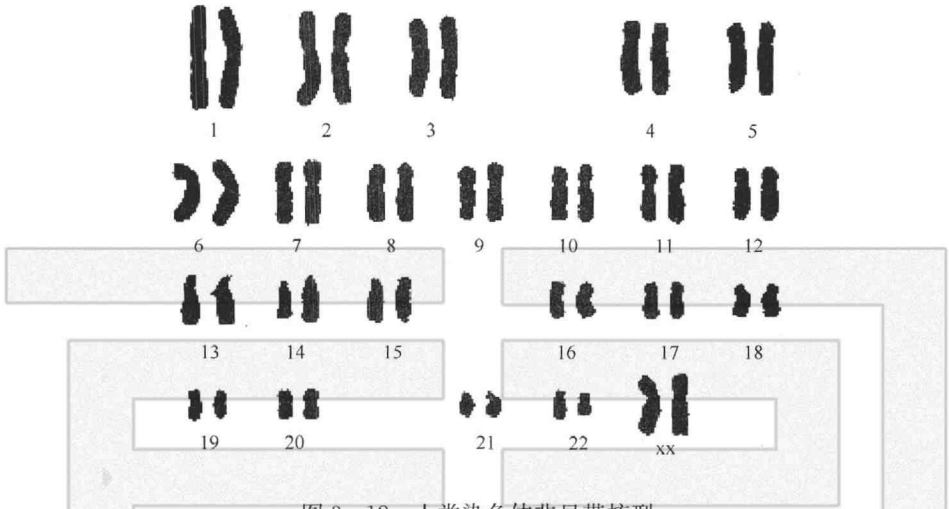


图 3-12 人类染色体非显带核型

(图 3-14) 和 R 带, 另一类是局部性的显带, 它只能使少数特定的区域显带, 如 C (图 3-15)、Cd、T 和 N 带。正常人体细胞的染色体带型图如 3-16 所示。

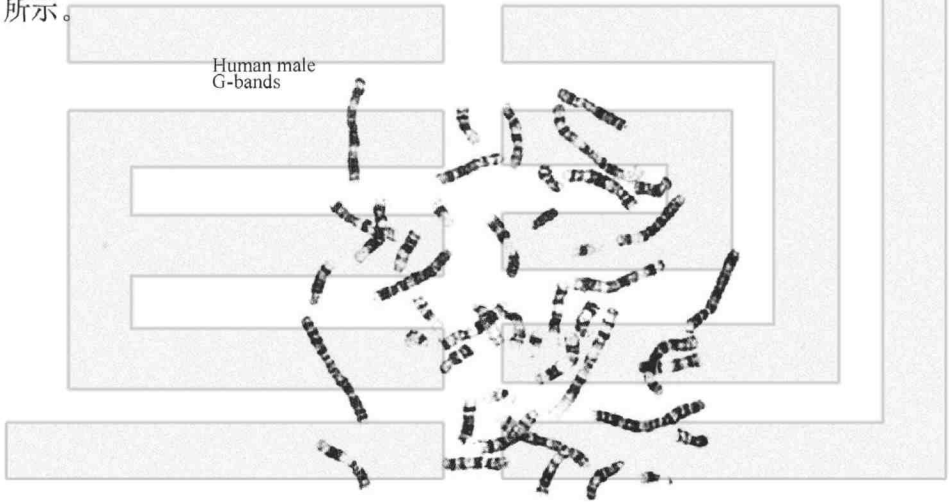


图 3-13 人类 G 带核型, G 带显示的是染色体上富含 AT 的区域

4. 几类的特殊的染色体

(1) 多线染色体

1881 年意大利的 Balbiani 发现于双翅目摇蚊幼虫的唾腺细胞。其特点是：
①体积巨大 (图 3-17), 比其他体细胞染色体长 100—200 倍, 体积大 1000—

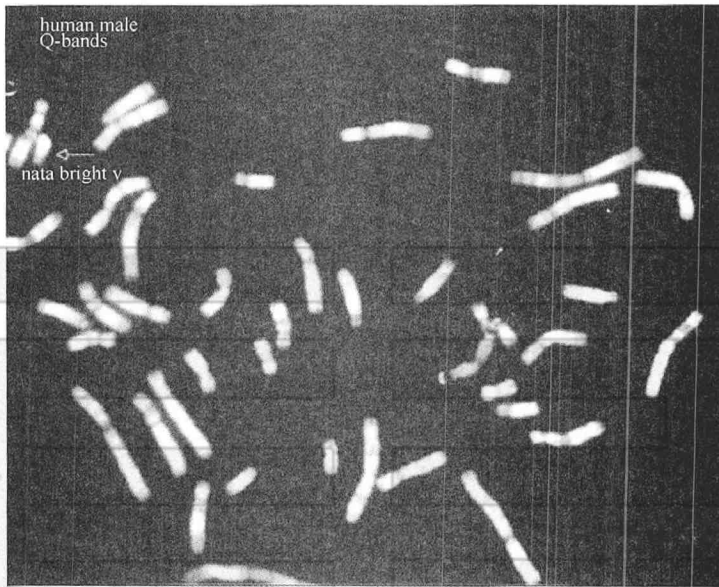


图 3-14 人类 Q 带核型, Q 带显示的带纹与 G 带相似

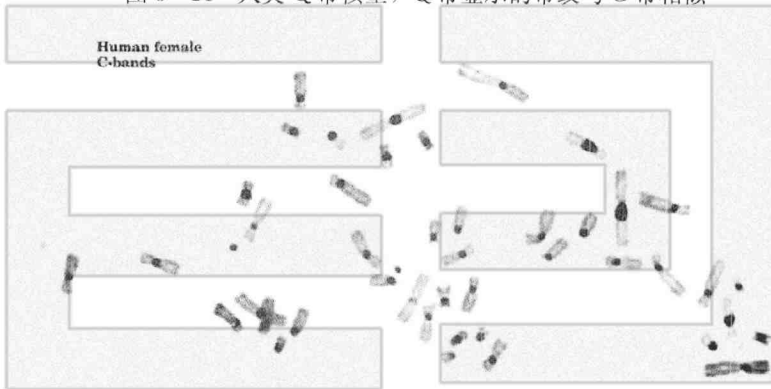
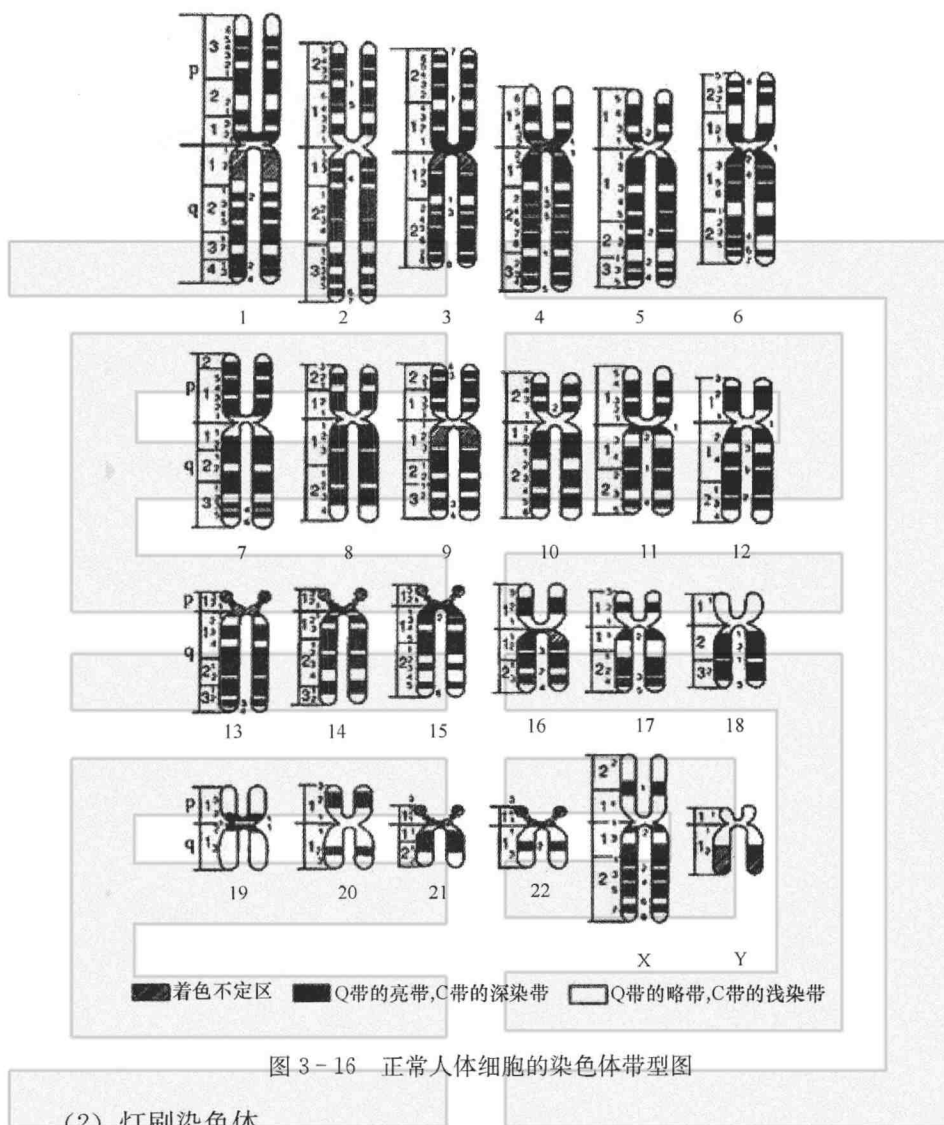


图 3-15 人类 C 带核型, C 带显示的是着丝粒异染色质

2000 倍, 这是由于核内有丝分裂的结果, 即染色体多次复制而不分离。②多线性, 每条多线染色体由 500—4000 条解旋的染色体合并在一起形成。③体细胞联会, 同源染色体紧密配对, 并合并成一个染色体。④横带纹, 染色后呈现出明暗相间的带纹。⑤膨突和环, 在幼虫发育的某个阶段, 多线染色体的某些带区疏松膨大, 形成膨突 (puff) 或巴氏环 (Balbiani ring)。用 H^3 —TdR 处理细胞, 发现膨突被标记, 说明膨突是基因活跃转录的形态学标志。



(2) 灯刷染色体

这种染色体最早发现于鱼类、两栖类和爬行类卵母细胞减数分裂的双线期。双线期是卵黄合成的旺盛期。由于染色体主轴两侧有侧环，状如灯刷，故名灯刷染色体（图 3-18）。它由两条同源染色体组成，在交叉处结合，每条同源染色体含 2 条染色单体。轴上有一些染色粒，代表染色质紧密螺旋化的部位。同时两条染色单体向两边伸出许多侧环，侧环是 RNA 活跃转录的区域。

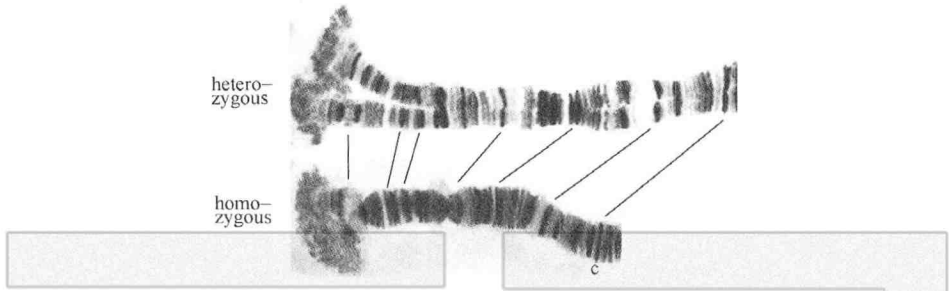


图 3-17 唾腺染色体

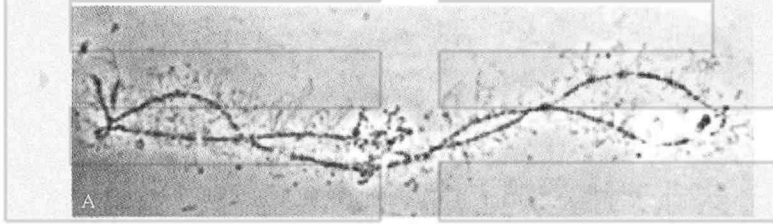


图 3-18 灯刷染色体

(3) B 染色体

1928 年伦道夫 (Randolph) 把生物的正常染色体叫做 A 染色体, 而把存在于许多动、植物中形态和行为不同于 A 染色体的超数染色体称为 B 染色体。

第二节 有丝分裂

真核细胞的细胞分裂 (cell division) 可分为有丝分裂 (mitosis) 和减数分裂 (meiosis) 两种类型。有丝分裂, 又称为间接分裂, 特点是有纺锤体染色体出现, 子染色体被平均分配到子细胞。减数分裂是指染色体复制一次而细胞连续分裂两次的分裂方式, 是高等动植物配子体形成的分裂方式。

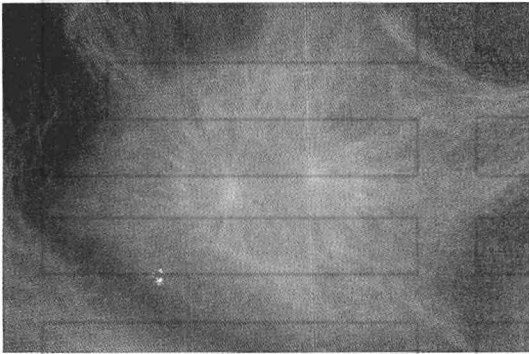
细胞周期是指连续分裂的细胞从一次有丝分裂结束到下一次有丝分裂完成所经历的整个过程, 是细胞物质累积与细胞分裂的循环过程。在细胞周期中, 染色体和染色质的形态交替变化, 遗传物质在细胞周期中维持了亲代到子代的传递和恒定。

有丝分裂过程是一个连续的过程，为了便于描述人为的划分为六个时期：间期（interphase）、前期（prophase）、前中期（premetaphase）、中期（metaphase）、后期（anaphase）和末期（telophase）。其中间期包括 G₁ 期、S 期和 G₂ 期，主要进行 DNA 复制等准备工作。

1. 前期

前期（图 3-19）的主要事件是：①染色质凝缩，②分裂极确立与纺锤体开始形成，③核仁解体，④核膜消失。

前期最显著的特征是染色质通过螺旋化和折叠，变短变粗，形成光学显微镜下可以分辨的染色体，每条染色体包含 2 个染色单体。



早在 S 期两个中心粒已完成复制，在前期移向两极，两对中心粒之间形成纺锤体微管，当核膜解体时，两对中心粒已到达两极，并在两者之间形成纺锤体，纺锤体微管包括：

图 3-19 前期两个中心体向两极移动

(1) 着丝点微管（kinetochore mt）：由中心体发出，

连接在着丝点上，负责将染色体牵引到纺锤体上，着丝点上具有马达蛋白。

(2) 星体微管（astral mt）：由中心体向外放射出，末端结合有分子马达，负责两极的分离，同时确定纺锤体纵轴的方向。

(3) 极体微管（polar mt 或 overlap mt）：由中心体发出，在纺锤体中部重叠，重叠部位结合有分子马达，负责将两极推开。

有两类马达蛋白参与染色体、分裂极的分离，一类是 dynein，另一类是 kinesin。

植物没有中心粒和星体，其纺锤体叫作无星纺锤体，分裂极的确定机理尚不明确。

2. 前中期

前中期指由核膜解体到染色体排列到赤道面（equatorial plane）这一阶段

(图 3-20 左)。纺锤体微管向细胞内部侵入，与染色体的着丝点结合。着丝点处的分子马达使染色体向微管的负端移动。在光镜下可以看到，此时染色体也就是既向一极移动也向另一极移动，是以振荡的方式移向纺锤体中部的。其原因是姊妹染色单体的着丝点都结合有微管和分子马达。

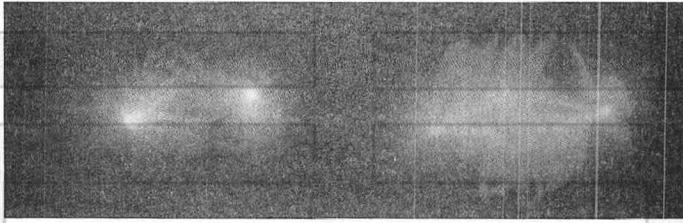


图 3-20 左侧，前中期；右侧，中期

3. 中期

中期指从染色体排列到赤道面上 (图 3-20 右、图 3-21)，到姊妹染色单体开始分向两极的一段时间，纵向观动物染色体呈辐射状排列。染色体两边的牵引力就像拔河一样达到平衡。

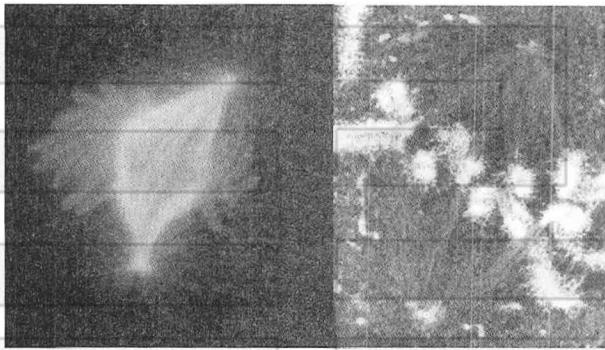


图 3-21 中期，右图显示与染色体连接的微管

4. 后期

后期指姊妹染色单体分开并移向两极的时期，当子染色体到达两极后，标志这一时期结束 (图 3-22)。

后期可以分为两个方面 (图 3-23)：①后期 A，指染色体向两极移动的过程。这是因为染色体着丝点微管在着丝点处去组装而缩短，在分子马达的作用下染色体向两极移动，体外实验证明即使在不存在 ATP 的情况下，染色

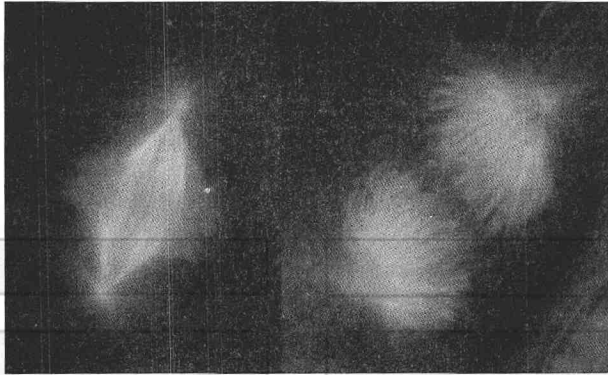


图 3-22 后期姊妹染色单体分离

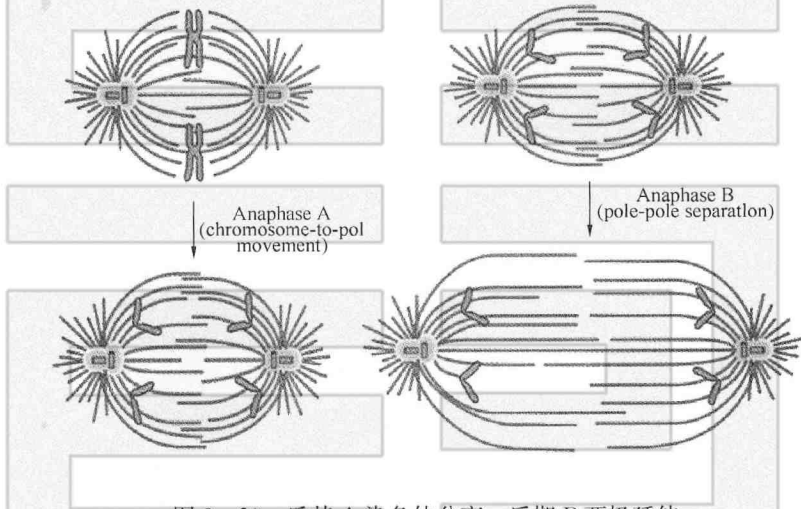


图 3-23 后期 A 染色体分离，后期 B 两极延伸

体着丝点也有连接到正在去组装的微管上的能力，使染色体发生移动。②后期 B，指两极间距离拉大的过程。这是因为一方面极体微管延长，结合在极体微管重叠部分的马达蛋白提供动力，推动两极分离；另一方面星体微管去组装而缩短，结合在星体微管正端的马达蛋白牵引两极距离加大。可见染色体的分离是在微管与分子马达的共同作用下实现的（图 3-24）。

后期 A，B 是用药物鉴定出来的，如紫杉酚（taxol）能结合在微管的（+）端，抑制微管（+）端去组装，从而抑制后期 A。动物中通常先发生后期 A，再后期 B，但也有些只发生后期 A，还有的后期 A、B 同时发生。植物细胞没有后期 B。

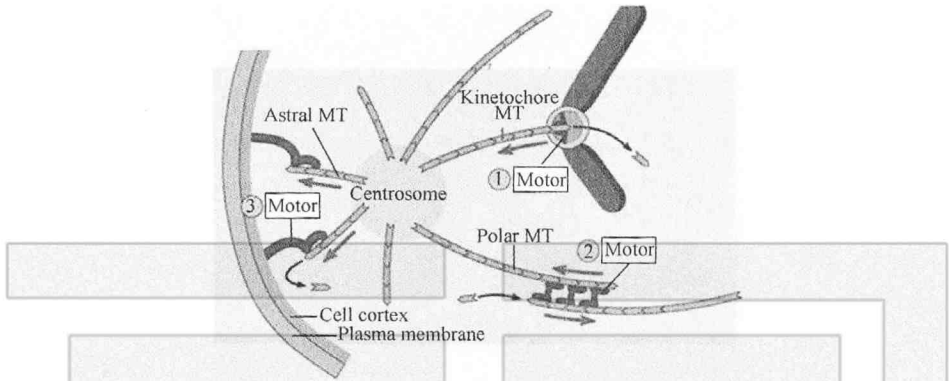


图 3-24 马达蛋白和微管系统共同协作，使染色体分离

5. 末期

末期（图 3-25）是从子染色体到达两极，至形成两个新细胞为止的时期。末期涉及子核的形成和胞质分裂两个方面。

(1) 子核的形成

末期子核的形成，大体经历了与前期相反的过程，即染色体解聚缩，核仁出现和核膜重新形成。核仁由染色体上的核仁组织中心形成（NORs），几个 NORs 共同组成一个大的核仁，因此核仁的数目通常比 NORs 的数目要少。

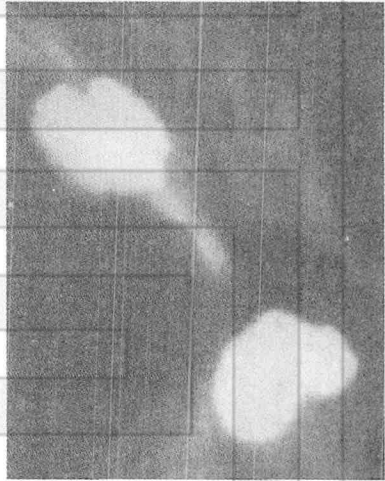


图 3-25 末期

前期核膜解体后，核纤层蛋白 B 与核膜残余小泡结合，末期核纤层蛋白 B 去磷酸化，介导核膜的重新装配。

(2) 胞质分裂

虽然核分裂与胞质分裂（cytokinesis）是相继发生的，但属于两个分离的过程，例如大多数昆虫的卵，核可进行多次分裂而无胞质分裂，某些藻类的多核细胞可长达数尺，以后胞质才分裂形成单核细胞。

动物细胞的胞质分裂是以形成收缩环的方式完成的（图 3-26），收缩环在后期形成，由大量平行排列的肌动蛋白和结合在上面的 myosin II 等成分组成，用细胞松弛素及肌动蛋白和肌球蛋白抗体处理均能抑制收缩环的形成。

不难想象胞质收缩环工作原理和肌肉收缩是一样的。

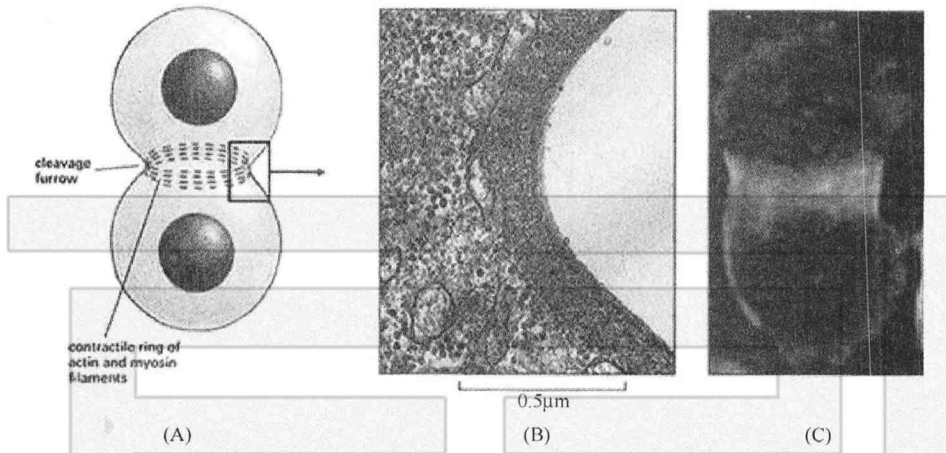


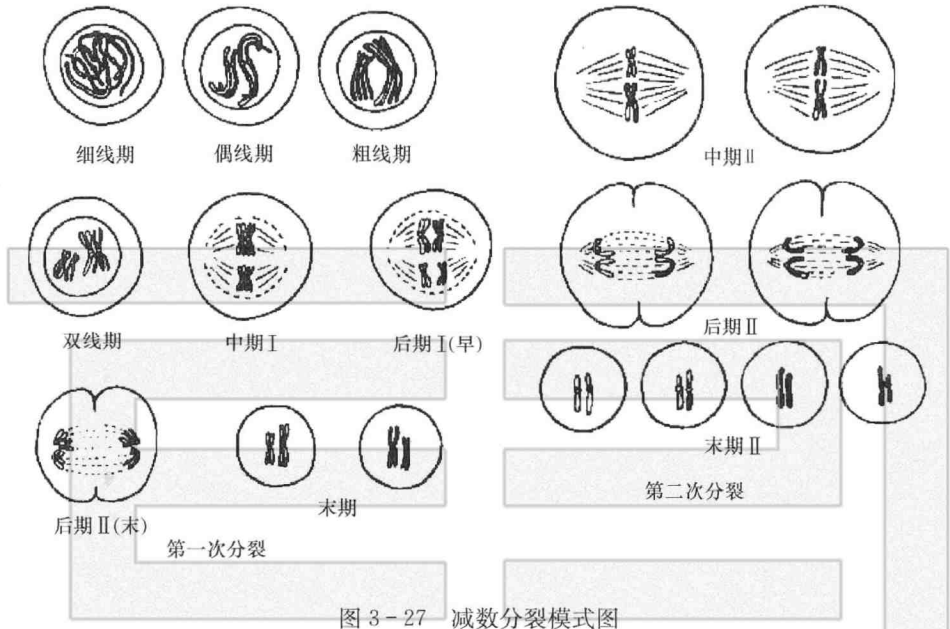
图 3-26 动物细胞的胞质收缩环

动物胞质分裂的另一特点是形成中体。末期纺锤体开始瓦解消失，但在纺锤体的中部微管数量增加，其中掺杂有高电子密度物质和囊状物，这一结构称为中体。在胞质分裂中的作用尚不清楚。

第三节 减数分裂

减数分裂 (Meiosis) 的特点是 DNA 复制一次，而细胞连续分裂两次，形成单倍体的精子和卵子，通过受精作用又恢复二倍体，减数分裂过程中同源染色体间发生交换，使配子的遗传多样化，增加了后代的适应性，因此减数分裂不仅是保证生物种染色体数目稳定的机制，并且也是物种适应环境变化不断进化的机制。

减数分裂由紧密连接的两次分裂构成。通常减数分裂 I 分离的是同源染色体，所以称为异型分裂 (heterotypic division) 或减数分裂 (reductional division)。减数分裂 II 分离的是姊妹染色体，类似于有丝分裂，所以称为同型分裂 (homotypic division) 或均等分裂 (equational division)。同有丝分裂一样，为了描述方便将减数分裂分为几个期和亚期 (图 3-27)。



一、间期

有丝分裂细胞在进入减数分裂之前要经过一个较长的间期，称前减数分裂间期 (premeiotic interphase) 或前减数分裂期 (premeiosis)。

前减数分裂期也可分为 G_1 期、S 期和 G_2 期，在 G_1 期和 S 期把麝香百合的花粉细胞在体外培养，则发现细胞进行有丝分裂，将 G_2 晚期的细胞在体外培养则向减数分裂进行，说明 G_2 期是有丝分裂向减数分裂转化的关键时期。

和有丝分裂不同的是，DNA 不仅在 S 期合成，而且也在前期合成一小部分。这些 DNA 的合成可能与联会复合体的形成有关。

二、分裂期

1. 减数分裂 I

(1) 前期 I

减数分裂的特殊过程主要发生在前期 I，通常人为划分为 5 个时期：① 细线期 (leptotene)、② 合线期 (zygotene)、③ 粗线期 (pachytene)、④ 双线期 (diplotene)、⑤ 终变期 (diakinesis)。必须注意的是这 5 个阶段本身是连

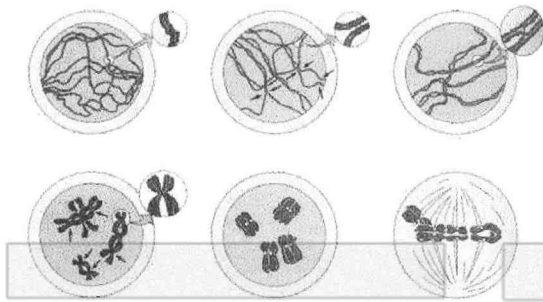


图 3-28 减数分裂前期 I

交织在一起，偏向核的一方，所以又称为凝线期 (synizesis)，在有些物种中表现为染色体细线一端在核膜的一侧集中，另一端放射状伸出，形如花束，称为花束期 (bouquet stage)。

②合线期：持续时间较长，占有丝分裂周期的 20%。亦称偶线期，是同源染色体配对的时期，这种配对称为联会 (synapsis) (图 3-29)。这一时期同源染色体间形成联会复合体 (synaptonemal complex, SC)。在光镜下可以看到两条结合在一起的染色体，称为二价体 (bivalent)。每一对同源染色体都经过复制，含四个染色单体，所以又称为四分体 (tetrad)。

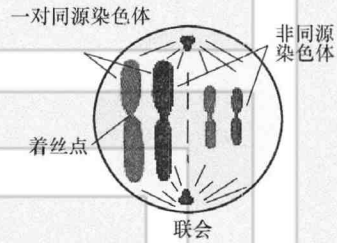


图 3-29 联会示意图

③粗线期：持续时间长达数天，此时染色体变短，结合紧密，在光镜下只在局部可以区分同源染色体，这一时期是同源染色体的非姊妹染色单体之间发生交换的时期。在果蝇粗线期 SC 上具有与 SC 宽度相近的电子致密球状小体，称为重组节，与 DNA 的重组有关。

④双线期：联会的同源染色体相互排斥，开始分离，但在交叉点 (chiasma) 上还保持着联系。双线期染色体进一步缩短，在电镜下已看不到联会复合体。

交叉的数目和位置在每个二价体上并非固定的，而随着时间推移，向端部移动，这种移动现象称为端化 (terminalization)，端化过程一直进行到中期。

植物细胞双线期一般较短，但在许多动物中双线期停留的时间非常长，

人的卵母细胞在五个月胎儿中已达双线期，而一直到排卵都停在双线期，排卵年龄大约在 12-50 岁之间。成熟的卵细胞直到受精后，才迅速完成两次分裂，形成单倍体的卵核。

在鱼类、两栖类、爬行类、鸟类以及无脊椎动物的昆虫中，双线期的二价体解螺旋而形成灯刷染色体，这一时期是卵黄积累的时期。

⑤终变期：二价体显著变短，并向核周边移动，在核内均匀散开。所以是观察染色体的良好时期。

由于交叉端化过程的进一步发展，故交叉数目减少，通常只有一至两个交叉。终变期二价体的形状表现出多样性，如 V 形、O 形等。

核仁此时开始消失，核被膜解体，但有的植物，如玉米，在终变期核仁仍然很显著。

(2) 中期 I

核仁消失，核被膜解体，标志进入中期 I，中期 I 的主要特点是染色体排列在赤道面上。每个二价体有 4 个着丝粒、姊妹染色单位的着丝粒定向于纺锤体的同一极，故称联合定向 (co-orientation)。

(3) 后期 I

二价体中的两条同源染色体分开，分别向两极移动。由于相互分离的是同源染色体，所以染色体数目减半。但每个子细胞的 DNA 含量仍为 2C。同源染色体随机分向两极，使母本和父本染色体重新组合，产生基因组的变异。如人类染色体是 23 对，染色体组合的方式有 2^{23} 个 (不包括交换)，因此除同卵孪生外，几乎不可能得到遗传上等同的后代。

(4) 末期 I

染色体到达两极后，解旋为细丝状、核膜重建、核仁形成，同时进行胞质分裂。

(5) 减数分裂间期

在减数分裂 I 和 II 之间的间期很短，不进行 DNA 的合成，有些生物没有间期，而由末期 I 直接转为前期 II (图 3-30)。

2. 减数分裂 II

可分为前、中、后、末四个四期，与有丝

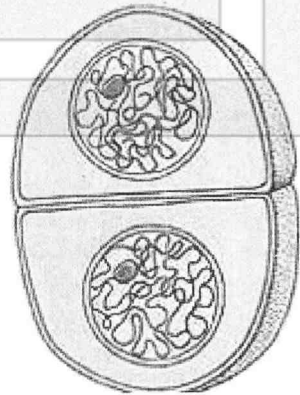


图 3-30 下次分裂前的间期

分裂相似。

通过减数分裂一个精母细胞形成 4 个精子。

而一个卵母细胞形成一个卵子及 2~3 个极体（图 3-31）。

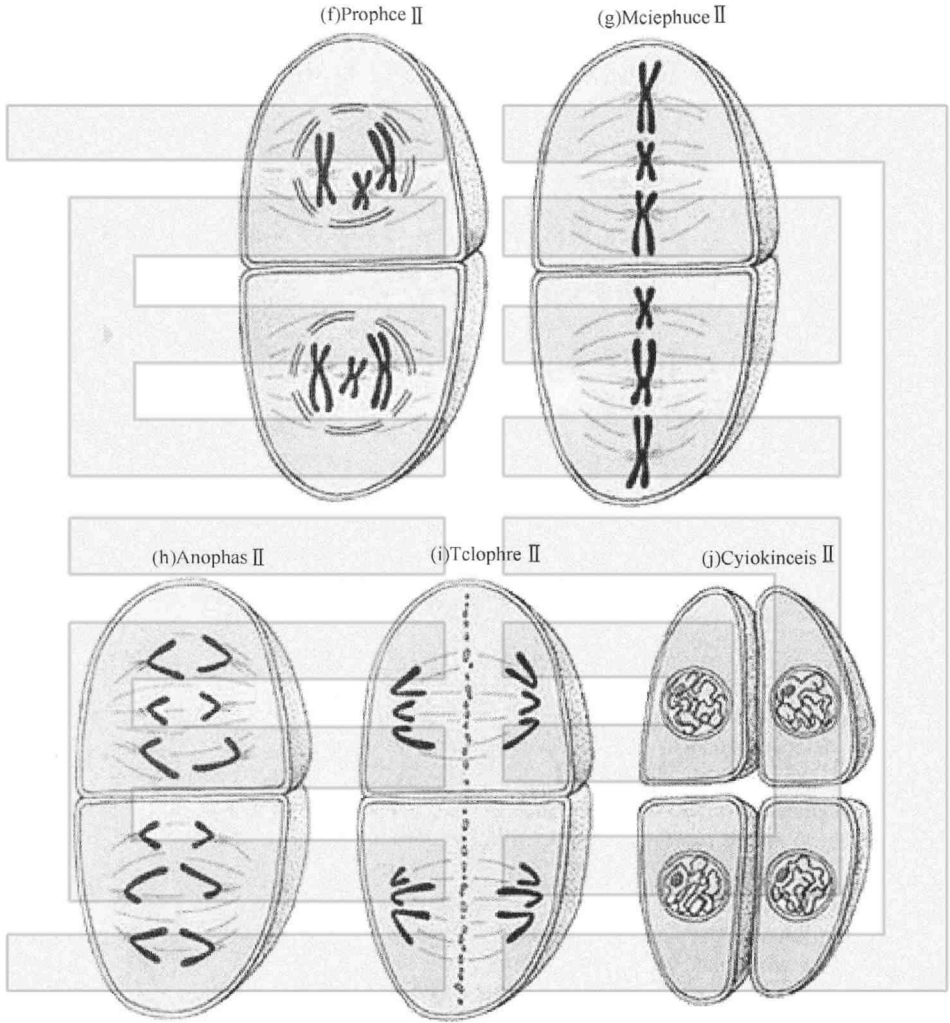


图 3-31 减数分裂 II 各期示意图

三、联会复合体

联会复合体（synaptonemal complex, SC）是减数分裂合线期两条同源染色体之间形成的一种结构，它与染色体的配对、交换和分离密切相关。

联会：

SC是同源染色体间形成的梯子样的结构。在电镜下观察，两侧是约40nm的侧生组分(lateral element)，电子密度很高，两侧之间为宽约100nm的中间区(intermediate space)，在电镜下是明亮区，在中间区的中央为中央组分(central element)，宽约30nm。侧生组分与中央组分之间有横向排列的粗约7~10nm的SC纤维，使SC外观呈梯子状(图3-32)。

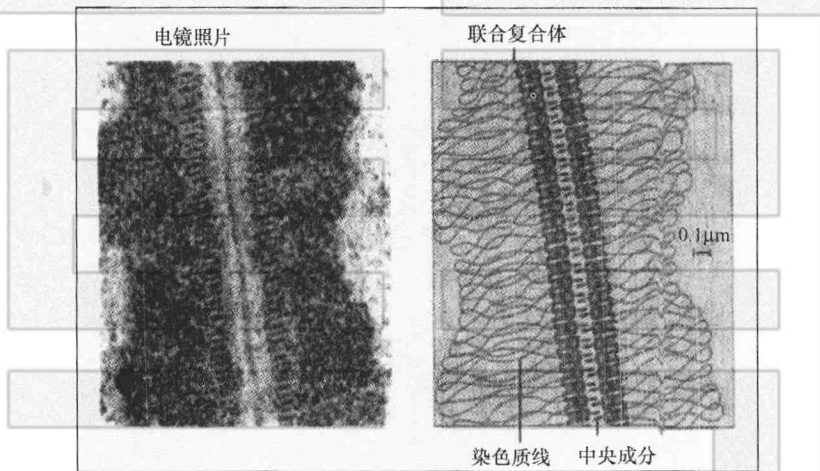


图3-32 联会复合体电镜图

长期以来人们认为SC将同源染色体组织在一起，使伸入SC的DNA之间产生重组，但实验证明不仅SC的形成晚于基因重组的启动，而且基因突变不能形成SC的酵母中，同源染色体间照样可以发生交换。现在一般认为它与同源染色体间交换的完成有关。

在磷钨酸染色的SC中央，还可以看到呈圆形或椭圆形的重组节(recombination nodules, RNs)，RNs是同源染色体发生交叉的部位，RNs上有基因交换所需要的酶。

从形态学来看，SC形成合线期，成熟于粗线期，并存在数天，消失于双线期。联会复合体的形成与合线期DNA(Zyg-DNA)有关，在细线期或合线期加入DNA合成抑制剂，则抑制SC的形成。

精子和卵子发生的示意图如图3-33和图3-34所示。

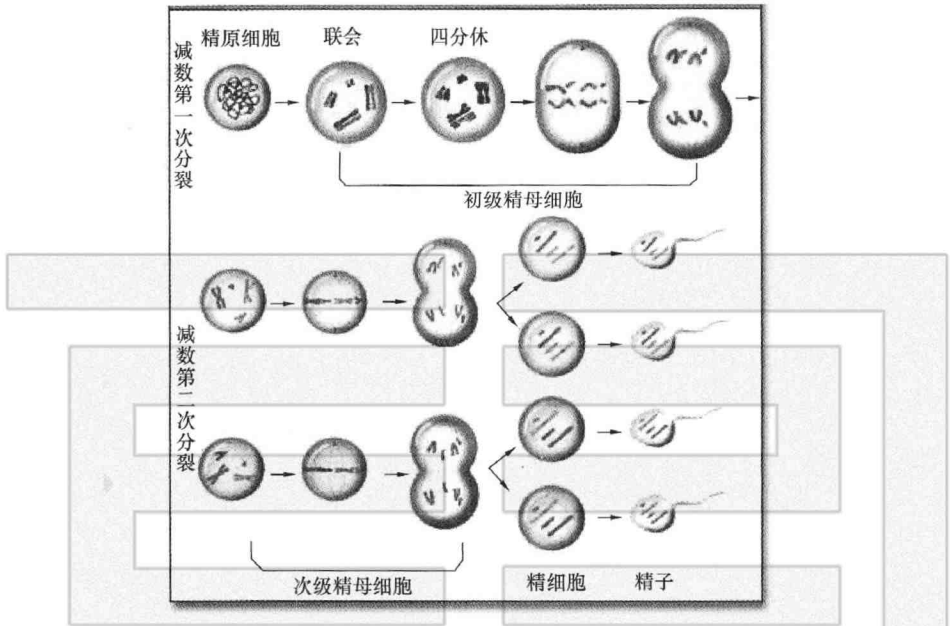


图 3-33 精子发生示意图

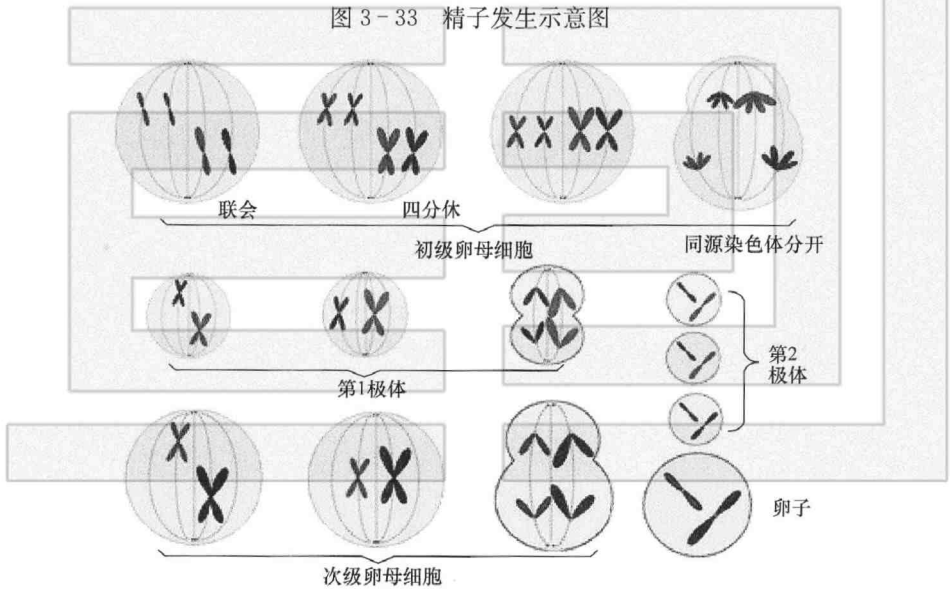


图 3-34 卵子发生示意图

第四章 中国少数民族

遗传学基本原理

第一节 孟德尔定律

孟德尔 (Mendel G. J 1822—1884) 在 1856—1864 年进行了 8 年的豌豆杂交试验, 在前人的基础上把植物杂交的工作向前推进了一大步, 确定了生物性状遗传的两个基本规律——遗传因子分离和多对因子分离后的自由组合。这两个规律后来称为孟德尔定律, 即分离规律 (the law of segregation) 和独立分配规律 (the law of independent assortment)。

一、分离规律

1. 性状分离现象

性状 (character) 是指生物体所表现的形态特征和生理特性的总称。豌豆 (*Pisum sativum*) 是严格的自花授粉植物。孟德尔在研究豌豆性状的遗传时, 把植株所表现的性状总体区分为各个单位来进行研究, 这些被区分开的每一个具体性状称为单位性状 (unit character)。例如, 豌豆的花色、种子形状、子叶颜色、豆荚形状、豆荚 (未成熟的) 颜色、花序着生部位和株高就是 7 个不同的、稳定的、易于区分的单位性状。不同个体在单位性状上常有着各种不同的表现, 如豌豆花色有紫花和白花、种子形状有圆形和皱缩、子叶颜色有黄色和绿色等 (图 4-1)。这种同一单位性状在不同个体间所表现出来的相对差异, 称为相对性状 (contrasting character)。

孟德尔在进行豌豆杂交试验时, 选用具有明显差别的 7 对相对性状的品种作为亲本, 用人工方法将它们分别成对地相互杂交, 并按照杂交后代的系

谱进行详细的记载，统计和计算了杂种后代表现相对性状的株数，并分析了它们的比例关系。

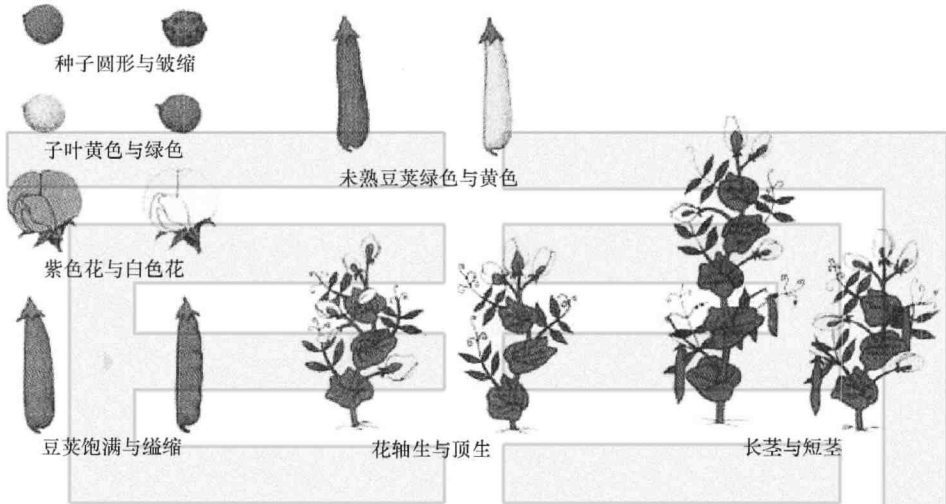


图 4-1 孟德尔研究的 7 对性状

(引自 Greffiths 等, 1996)

例如，为了研究豌豆种子形状的遗传，孟德尔用圆形种子的植株与皱缩种子的植株进行杂交。在杂交时，先将母本花蕾的雄蕊完全摘除，这称为去雄，然后将父本的花粉授到已去雄的母本柱头上，这称为人工授粉。去了雄和授了粉的母本花朵还必须套袋隔离，防止其他花粉授粉。结果发现圆形种子的植株不论做父本还是做母本，子一代杂种植株的种子全部为圆形，皱缩性状被圆形性状所遮盖。子一代自花授粉后，得到的子二代植株上结出的荚果内，同时出现了圆形和皱缩两种形状的种子。他统计了这些种子的数目，其中 5474 粒圆形，1850 粒皱缩，两者之比 2.96 : 1。这个比值非常接近 3 : 1 (图 4-2)。如果把圆形种子的植株做母本的杂交组合称为正交，则做父本的杂交组合即称为反交。正交和反交的结果

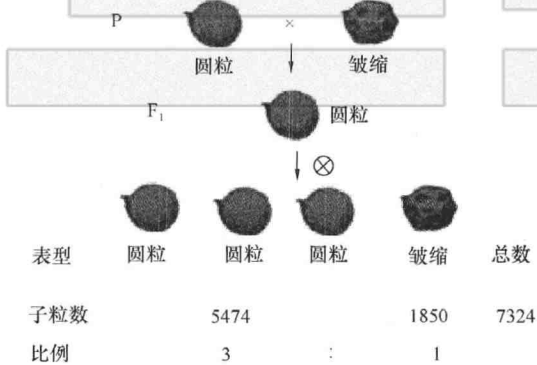


图 4-2 种子形状的遗传

果完全一样，说明了后代的性状表现不受亲本组合方式的影响。通常用 F (filial generation) 表示杂种后代，F₁ 即表示杂种第一代，是指杂交当代所结的种子及由它所长成的植株。F₂ 表示杂种第二代，是指由 F₁ 自交产生的种子及由它所长成的植株。依此类推，F₃、F₄ 分别表示杂种第三代、杂种第四代。

孟德尔把在子一代表现出来的性状称为显性性状 (dominant character)，如圆形。未表现出来的性状叫做隐性性状 (recessive character)，如皱缩。其他 6 对相对性状的杂交试验，都获得同样的试验结果 (表 4-1)。

孟德尔从以上 7 对相对性状的杂交结果，看到了两个共同特点：第一，不论正交、反交，F₁ 所有植株的性状表现是一致的，都只表现一个亲本的性状，另一亲本的性状隐藏未现。第二，F₂ 植株在性状表现上是不同的，一部分植株表现一个亲本的性状，其余植株则表现另一个亲本的相对性状，即显性性状和隐性性状在 F₂ 中都表现了出来，这种现象叫做性状分离现象 (character segregation)，F₂ 群体中表现显性性状的个体与表现隐性性状的个体比例大约总是 3 : 1。

表 4-1 孟德尔豌豆一对相对性状杂交试验的结果

性 状	杂交组合	F ₁ 表现的	F ₂ 的表现		比 例
		显性性状	显性性状	隐性性状	
花色	紫花×白花	紫花	705 紫花	224 红花	3.15 : 1
种子形状	圆粒×皱粒	圆粒	5474 圆粒	1850 皱粒	2.96 : 1
子叶颜色	黄色×绿色	黄色	6022 黄色	2001 绿色	3.01 : 1
豆荚形状	饱满×缢缩	饱满	882 饱满	299 缢缩	2.95 : 1
未熟豆荚色	绿色×黄色	绿色	428 绿色	152 黄色	2.82 : 1
花着生位置	轴生×顶生	轴生	651 轴生	207 顶生	3.14 : 1
植株长度	长茎×短茎	长茎	787 长茎	277 短茎	2.84 : 1

2. 分离现象的解释 (表现型和基因型)

这 7 对相对性状在 F₂ 为什么都出现 3 : 1 的分离比呢？孟德尔为了解释这些结果，提出了下面的假设：

(1) 一对相对性状由一对遗传因子 (hereditary determinant 或 hereditary factor) 控制

(2) 遗传因子在体细胞内是成对的，一个来自父方，一个来自母方。

(3) 杂种的“遗传因子”彼此不同，各自保持独立性，且存在显隐关系。即 F₁ 植株有一个控制显性性状的遗传因子和一个控制隐性性状的遗传因子。

(4) 在形成配子时，每对遗传因子相互分离，均等地分配到不同的配子中，结果每个配子（精核或卵细胞）中只含有成对遗传因子中的一个。

(5) 在形成合子（子代新体）时，配子的结合是随机的。

用符号 P 表示亲本（parent），♀ 表示母本，♂ 表示父本，× 表示杂交，⊗ 表示自交，字母大小写表示显性和隐性遗传因子。孟德尔的豌豆种子形状的杂交实验及其结果可以表示如下（图 4-3）：以 R（round）表示显性性状

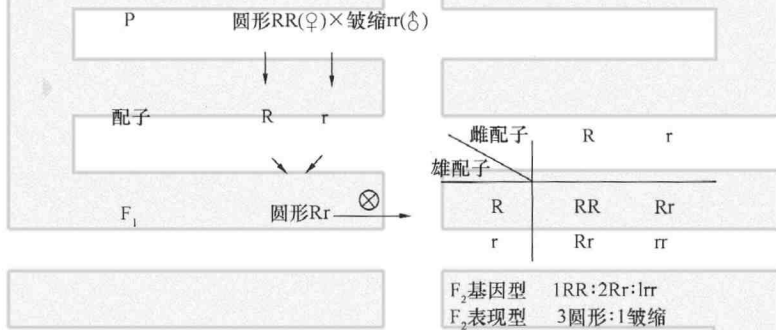


图 4-3 孟德尔对分离现象的解释

——圆形的遗传因子，r（wrinkled）表示隐性性状——皱缩的遗传因子。圆形种子亲本应具有一对圆形遗传因子 RR，皱缩亲本应具有一对皱缩遗传因子 rr。圆形亲本产生的配子中只有一个遗传因子 R，皱缩亲本产生的配子中只有一个遗传因子 r。受精时，雌雄配子结合形成的 F₁ 应该是 Rr。由于 R 对 r 有显性的作用，所以 F₁ 种子是圆形的。但是 F₁ 植株在产生配子时，由于 Rr 因子分配到不同的配子中去，所以产生的配子（不论雌配子还是雄配子）有两种：一种带有遗传因子 R，另一种带有遗传因子 r，两种配子数目相等，呈 1:1 的比例。F₁ 自交时各类雌雄配子随机结合。由此可见，F₂ 群体的各种组合按遗传因子的成分归纳，实际上是三种：1/4 个体 RR，2/4 个体 Rr，1/4 个体 rr。1/4 RR 和 2/4 Rr 种子都是圆形，只有 1/4rr 是皱缩，所以 F₂ 种子圆形与皱缩之比是 3:1。

孟德尔在解释上述遗传试验中所用的遗传因子，在 1909 年由丹麦遗传学家约翰逊创造的“基因”（gene）这个名词所取代。遗传学中将控制一对相对

性状位于同源染色体上相同位点的两个不同形式的基因称为等位基因 (alleles)。如种子圆形基因 R 和皱缩基因 r, 相互为等位基因。个体的基因组合, 称为基因型 (genotype)。例如, 决定种子圆形性状的基因型为 RR 和 Rr, 种子皱缩性状的基因型为 rr。基因型是生物性状表现的内在遗传基础, 是肉眼看不到的, 只能通过杂交试验根据表现型来推断。表现型 (phenotype) 是指生物体所表现的性状, 如种子圆形和皱缩等。表现型简称表型, 它是基因型和外界环境作用下的具体表现, 是可以直接观测的。像 RR 个体和 Rr 个体, 种子都为圆形, 其表现型相同, 但基因型不同。

从基因的组合来看, 像 RR 和 rr 两个基因型, 等位基因是一样的, 这在遗传学上称为纯合基因型 (homozygous genotype)。具有纯合基因型的个体称为纯合体 (homozygote), RR 个体为显性纯合体 (dominant homozygote), rr 个体为隐性纯合体 (recessive homozygote)。如 Rr 基因型, 等位基因不同, 称为杂合基因型 (heterozygous genotype)。具有杂合基因型的个体称为杂合体 (heterozygote)。

从细胞学的角度看, R 和 r 是一对等位基因, 位于一对同源染色体的相对位点上。F₁ 的基因型是 Rr, 当它的孢母细胞进行减数分裂形成配子时, 随着这对同源染色体在后期 I 的分离, R 与 r 也分别进入不同的二分体。最后形成 1/2 的 R 配子和 1/2 的 r 配子。雌雄配子都是这样。因此, 雌雄配子相互随机结合出现 4 种组合, 在基因型上出现 1RR : 2Rr : 1rr 比例, 在表现型上出现 3 : 1 的比例。

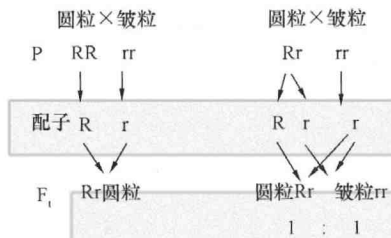
3. 分离规律的验证

任何一个理论或假说能否成立, 除了要对已有的事实作出解释外, 还必须经得起检验。分离规律是建立在一种假设的基础上的, 这个假设的实质就是成对的基因 (等位基因) 在配子形成过程中彼此分离, 互不干扰, 因而配子中只具有成对基因的一个。为了证明这一假设的真实性, 孟德尔采用了测交和自交的方法进行验证, 并得到了证实。

(1) 测交法

所谓测交 (test cross) 是指被测验的个体与隐性纯合个体间的杂交。测交所得的后代为测交子代, 用 F_t 表示。根据测交子代所出现的表现型种类和比例, 可以确定被测个体是纯合体还是杂合体。由于隐性纯合体只能产生一种含

隐性基因的配子，它们和含有任何基因的另一配子结合，其子代都只能表现出另一种配子所含基因的表现型。因此，测交子代表现型的种类和比例正好反映了被测个体所产生的配子种类和比例，从而可以确定被测个体的基因型。



例如，某被测个体为圆形种子的豌豆，不知其为纯合体还是杂合体。当它与皱缩种子的豌豆（隐性的 rr 纯合体）杂交，由于后者只产生一种含 r 基因的配子，所以如果测交子代全部是圆形种子，就说明该被测豌豆是 RR 纯合体，因为它只产生含 R 基因的一种配子。如果在测交子代中有 $1/2$ 的种子圆形， $1/2$ 的种子皱缩，就说明被测豌豆的基因型是 Rr （图 4-4）。

图 4-4 豌豆圆粒和皱粒一对基因的分

(2) 自交法

孟德尔为了验证遗传因子的分离，也曾继续使 F_2 植株自交产生 F_3 株系（来自一个 F_2 植株的子代群体）。然后根据 F_3 的性状表现，证实他所设想的 F_2 基因型。按照他的设想， F_2 的隐性性状植株只能产生隐性性状的 F_3 ；而在 F_2 的显性性状植株中， $2/3$ 是杂合体， $1/3$ 是纯合体。如果真如此，则前者自交产生的 F_3 群体就应该又分离为 $3/4$ 的显性植株和 $1/4$ 的隐性植株；而后者自交产生的 F_3 群体就应该一律表现显性。实际自交的结果表明（表 4-2），这两类显性性状的 F_2 植株的比例接近 $2:1$ 。实际观察的结果证实了他的推论。孟德尔对前述 7 对性状，连续自交了 4—6 代均未发现与他的推论不相符合的情况。

表 4-2 豌豆 F_2 表现显性性状的个体分别自交后的 F_3 表现型种类及其比例

性 状	在 F_3 表现显性：隐性		F_3 株系总数
	= 3 : 1 的株系数	性状的株系数	
花色	64(1.80)	36(1)	100
种子形状	372(1.93)	193(1)	565
子叶颜色	353(2.13)	166(1)	519
豆荚形状	71(2.45)	29(1)	100
未熟豆荚色	60(1.50)	40(1)	100
花着生位置	67(2.03)	33(1)	100
植株高度	72(2.57)	28(1)	100

(3) F_1 花粉鉴定法

关于等位基因发生分离的时机，细胞遗传学已经充分证明它是在杂种细胞进行减数分裂形成配子时发生的。随着染色体减半的发生，各对同源染色体分别分配到两个配子中去，位于同源染色体上的等位基因也就随之分开从而分配到不同的配子中去。这种现象在某些植物如玉米、水稻、高粱、谷子等中可用 F_1 植株产生的花粉粒进行观察鉴定。

例如，水稻的籽粒有糯性和非糯性两种，已知它们是受一对等位基因控制的，分别控制着籽粒及其花粉粒中的淀粉性质。非糯性的为直链淀粉，由显性基因 Wx 控制；糯性的为支链淀粉，由隐性基因 wx 控制。通常糯性的花粉或籽粒的胚乳与稀碘液反应后呈棕黄色；非糯性的花粉或籽粒与稀碘液反应后则呈蓝黑色。如以碘液处理水稻（糯性或非糯性） F_1 植株上的花粉，在显微镜下可以明显地看到棕黄色和蓝黑色的两种花粉粒，而且大致上各占一半。这清楚地表明了 F_1 ($Wxwx$) 产生了带有 Wx 基因和带有 wx 基因两种类型的配子，而且它们的数目是 1:1。

此外，用红色面包霉的不同品系杂交也可以验证等位基因的分离现象。当红色菌丝的正常种与白色菌丝变种有性交配时，二倍体接合子在一个长形的子囊内进行减数分裂和一次有丝分裂，形成 8 个子囊孢子。在子囊尚未破裂前，把各个子囊中的子囊孢子分别取出培养，就会看到每个子囊中 4 个孢子长出红色菌丝，另外 4 个孢子长出白色菌丝，形成 1:1 的比例。这也是由于等位基因在减数分裂时发生分离的结果。

孟德尔的基因分离规律在很多生物中都得到了证实，具有普遍意义。在人类中也发现不少呈现单基因分离模式的家系遗传现象。例如白化病、侏儒症、多指（趾）症等在一些家系中的表现。

根据分离规律，由具有一对相对性状的个体杂交产生的 F_1 ，其自交后代分离比为 3:1，测交后代分离比为 1:1。这些分离比的出现必须满足以下的条件：

- ① 研究的生物体是二倍体。
- ② F_1 个体形成的两种配子数目相等或接近相等，并且两种配子生活力相同；受精时各雌雄配子都能以均等的机会相互自由结合。
- ③ 不同基因型的合子及由合子发育的个体具有同样或大致同样的存活率。

④ 研究的相对性状差异明显，显性表现完全。

⑤ 杂种后代都处于相对一致的条件下，而且试验分析的群体比较大。

这些条件在一般情况下是具备的，所以大量试验结果都能符合这个基本遗传规律。

4. 分离规律的应用

分离规律是遗传学中最基本的一个规律。根据分离规律，必须重视表现型和基因型之间的联系和区别。对于一个表现显性的材料，如果不知道其显性基因位点是否纯合，就可以通过自交或测交的方法来确定。实践中正是利用纯种不分离、杂种要分离规律来鉴定品种或品系是否纯合。从而可以区别真伪杂种。例如，小麦的无芒对有芒显性，如果用有芒品种作为母本，无芒品种作父本，杂种一代应该全是无芒的。如果出现少量有芒植株，则这些植株可能是由于母本去雄不干净而产生的自交后代，这不是真正的杂种，应该淘汰。

分离规律表明，杂种通过自交将产生性状分离，同时也导致基因的纯合。在杂交育种工作中，在杂种后代连续进行自交和选择，目的就是促使个体间的分离和个体基因型的纯合。根据各性状的遗传研究，可以比较准确地预计后代分离的类型及其出现的频率，从而可以有计划地种植杂种后代，提高选择效果，加速育种进程。例如，水稻对稻瘟病的抗病性和感病性分别由显性基因和隐性基因决定，在它们的 F_2 群体内虽然很容易选到抗病植株，但根据分离规律，可以预料其中某些抗病植株的抗病性仍要分离。因此，还需要通过自交和进一步的选择，才能从中选出抗病性稳定的纯合体植株。

在生产上为了保持良种的增产作用，必须防止品种因天然杂交而发生分离退化。因此，需要加强良种繁育工作，注意品种保纯，做好去杂去劣或适当的隔离繁殖。

此外，根据分离规律的启示，杂种产生的配子在基因型上是纯粹的，可以利用花粉培养的方法培育出优良的纯合二倍体植株，为育种工作开辟了新的途径。

二、独立分配规律

孟德尔在说明了一对相对性状的遗传规律后，进一步研究了两对和两对以上相对性状之间的遗传关系，揭示了独立分配规律，又称自由组合规律。

1. 两对相对性状的遗传

为了研究两对相对性状的遗传，孟德尔仍以豌豆为材料，选取具有两对相对性状差异的纯合亲本进行杂交。例如，用种子圆形而子叶黄色的一个亲本与种子皱缩而子叶绿色的另一亲本杂交。其 F_1 都结圆形、黄色子叶的种子，表明种子圆形和子叶黄色都是显性。这与 7 对性状分别进行研究的結果是一致的。由 F_1 种子长成的植株（共 15 株）进行自交，得到 556 粒 F_2 种子，共有四种类型，其中两种类型和亲本相同，另两种类型为亲本性状的重新组合，而且存在着一定的比例关系（图 4-5）。

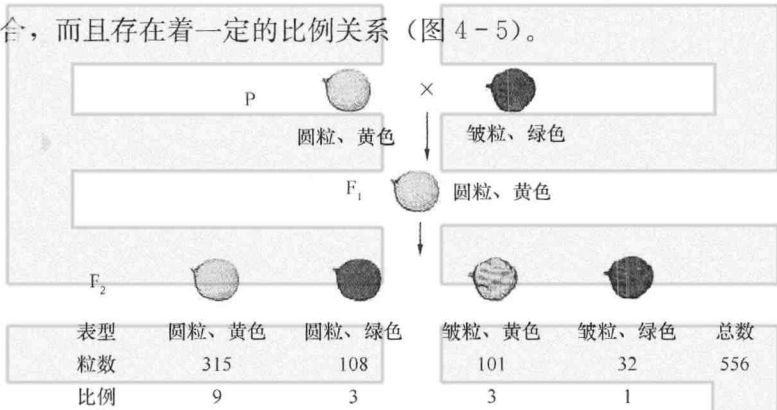


图 4-5 豌豆两对相对性状的杂交试验

如果把以上两对相对性状个体杂交试验的结果，分别按一对性状进行分析， F_2 代中：

$$\text{黄色} : \text{绿色} = (315 + 101) : (108 + 32) = 416 : 140 \approx 3 : 1$$

$$\text{圆粒} : \text{皱缩} = (315 + 108) : (101 + 32) = 423 : 133 \approx 3 : 1$$

根据上述分析，虽然两对相对性状是同时由亲代遗传给子代的，但由于每对性状的 F_2 分离仍然符合 3 : 1 的比例，说明它们是彼此独立地从亲代遗传给子代的，没有发生任何相互干扰的情况。同时在 F_2 群体内两种重组型个体的出现，说明两对性状的基因在从 F_1 遗传给 F_2 时，是自由组合的。按照概率定律，两个独立事件同时出现的概率，为分别出现的概率的乘积。因而圆种子、黄子叶同时出现的机会应为 $\frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$ ，皱种子、黄子叶同时出现的机会应为， $\frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$ ，圆种子、绿子叶同时出现的机会应为 $\frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$ ，

皱种子、绿子叶同时出现的机会应为 $\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$ ；用另一种方式表达是：

$$\begin{aligned} & \text{黄子叶} \frac{3}{4} : \text{绿子叶} \frac{1}{4} \\ & \times \text{圆种子} \frac{3}{4} : \text{皱种子} \frac{1}{4} \end{aligned}$$

$$\text{黄、圆} \frac{9}{16} : \text{黄、皱} \frac{3}{16} : \text{绿、圆} \frac{3}{16} : \text{绿、皱} \frac{1}{16}$$

将孟德尔试验的 556 粒 F₂ 种子，按上述的 9 : 3 : 3 : 1 的理论推算，即 556 分别乘以 $\frac{9}{16}$ 、 $\frac{3}{16}$ 、 $\frac{3}{16}$ 、 $\frac{1}{16}$ ，如下所列，所得的理论数值与实际结果比较，是非常近似的。

实得粒数	黄、圆	黄、皱	绿、圆	绿、皱
按理论比例 (9 : 3 : 3 : 1)	315	101	108	32
推算的粒数	312.75	104.25	104.25	34.75
差数	+2.25	-3.25	+3.75	-2.75

2. 独立分配现象的解释

独立分配规律的基本要点是：在配子形成时，控制不同相对性状的等位基因与另一对等位基因的分和组合是互不干扰的，各自独立分配到配子中去。

仍以上述杂交试验为例，用 Y 和 y 分别代表子叶黄色和绿色的一对基因，R 和 r 分别代表种子圆粒和皱粒的一对基因。从图 4-6 可以看出，黄色、圆粒亲本的基因型为 YYRR，产生 YR 配子；绿色、皱粒亲本的基因型为 yyrr，产生 yr 配子。两亲本的配子结合产生 F₁ 杂合体 YyRr，表现型为黄色、圆粒。F₁ 产生的雌配子和雄配子都是 4 种，即 YR、Yr、yR 和 yr，并且 4 种配子数目相等，为 1 : 1 : 1 : 1。雌雄配子结合，共有 16 种可能的组合。F₂ 群体中共有 9 种基因型。因为 Y 对 y 为完全显性，R 对 r 为完全显性，所以 F₂ 中只有 4 种表现型，其比例为 9 : 3 : 3 : 1，这与孟德尔的杂交试验结果是符合的。

现在我们从细胞学的角度看这 4 种配子是怎样形成的。Y 和 y 是一对等

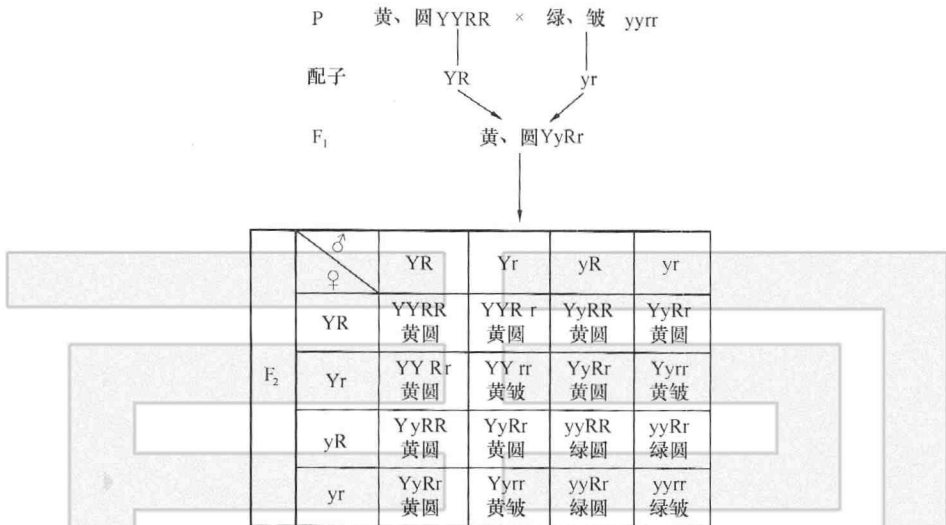


图 4-6 豌豆黄色、圆粒×绿色、皱粒的 F₂ 分离图解

位基因，位于同一对同源染色体的相对位点上。R 和 r 是另一对等位基因，位于另一对同源染色体的相对位点上。这两对等位基因互称为非等位基因 (non-allele)。F₁ 的基因型是 YyRr，当它的孢母细胞进行减数分裂形成配子时，随着这两对同源染色体在后期 I 的分离，Y 与 y 一定分别进入不同的二分体，R 与 r 也一定分别进入不同的二分体。此时，在一些孢母细胞内，可能是 Y 和 R 进入一个二分体，而 y 和 r 进入另一个二分体，最后形成 1/2 的 YR 配子和 1/2 的 yr 配子。在另一些孢母细胞内，可能是 Y 和 r 进入一个二分体，而 y 和 R 进入另一个二分体，最后形成 1/2 的 Yr 配子和 1/2 的 yR 配子。由于发生这两种分离的孢母细胞数目均等，所以这四种类型的配子数目相等，成为 1:1:1:1 的比例。雌雄配子都是这样。雌雄配子相互随机结合，因而有如图 4-7 所示，共出现 16 种组合，在表现型上出现 9:3:3:1 的比例。

由此可知，独立分配规律的实质在于：控制这两对性状的两对等位基因，分别位于不同的同源染色体上。在减数分裂形成配子时，每对同源染色体上的每一对等位基因发生分离，而位于非同源染色体上的基因之间可以自由组合。

3. 独立分配规律的验证

(1) 测交法

为了验证两对基因的独立分配规律，孟德尔同样采用了测交法。就是用

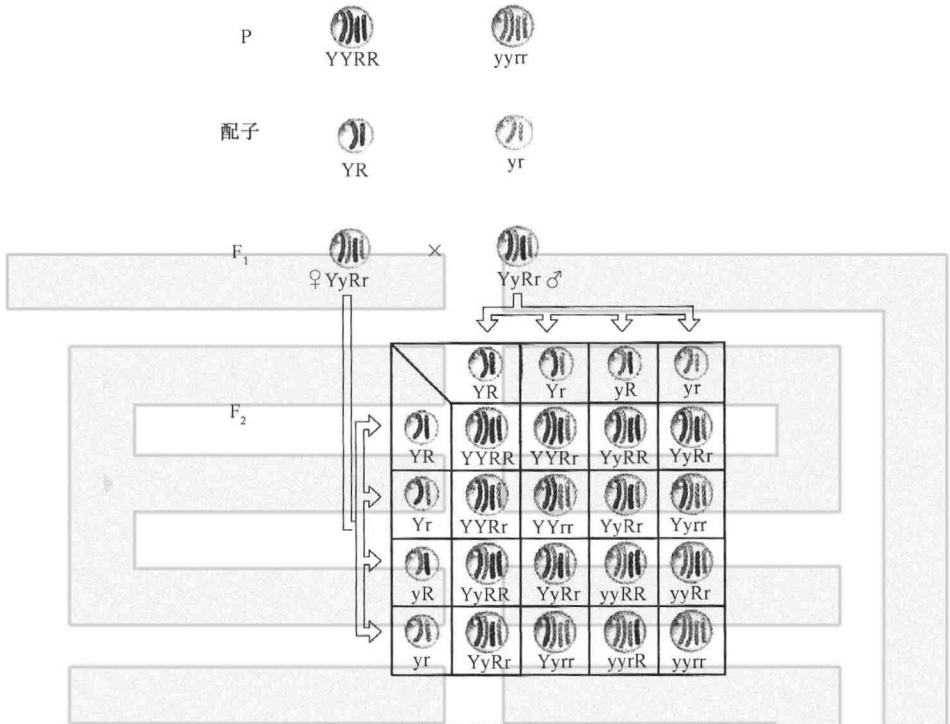


图 4-7 两对同源染色体及其载荷基因独立分配示意图

F₁ 与双隐性纯合体测交。当 F₁ 形成配子时，不论雌配子或雄配子，都有四种类型，即 YR、Yr、yR、yr，而比例相等。由于双隐性纯合体的配子只有 yr 一种，因此测交子代种子的表现型种类和比例应能反映 F₁ 所产生的配子种类和比例。表 4-3 说明孟德尔所得到的实际结果与测交的理论推断是完全一致的。

表 4-3 豌豆（黄色、圆粒×绿色、皱粒）F₁ 和双隐性亲本测交的结果

F₁黄、圆 YyRr × 绿、皱 yyrr

		配子	YR	Yr	yR	yr	yr
测交后代	基因型		YyRr	Yyrr	yyRr	yyrr	
	表现型		黄、圆	黄、皱	绿、圆	绿、皱	
期望值	表现型比例		1	1	1	1	
孟德尔的实际测交结果	F ₁ 为母本		31	27	26	26	
	F ₁ 为父本		24	22	25	26	

(2) 自交法

按照分离和独立分配规律的理论推断, 由纯合的 F_2 植株 (如 $YYRR$ 、 $yyRR$ 、 $YYrr$ 、 $yyrr$) 自交产生的 F_3 个体, 不会出现性状的分离, 这类植株在 F_2 群体中应各占 $1/16$ 。由一对基因杂合的植株 (如 $YyRR$ 、 $YYRr$ 、 $yyRr$ 、 $Yyrr$) 自交产生的 F_3 个体, 一对性状是稳定的, 另一对性状将分离为 $3:1$ 的比例。这类植株在 F_2 群体中应各占 $2/16$ 。由两对基因都是杂合的植株 ($YyRr$) 自交产生的 F_3 个体, 将分离为 $9:3:3:1$ 的比例。这类植株在 F_2 群体中应占 $4/16$ 。孟德尔所作的试验结果, 完全符合预定的推论, 例如:

F_2	F_3
38 株 $(\frac{1}{16})YYRR$	→全部为黄、圆, 没有分离
35 株 $(\frac{1}{16})yyRR$	→全部为绿、圆, 没有分离
28 株 $(\frac{1}{16})YYrr$	→全部为黄、皱, 没有分离
30 株 $(\frac{1}{16})yyrr$	→全部为绿、皱, 没有分离
65 株 $(\frac{2}{16})YyRR$	→全部为圆粒, 子叶颜色分离 3 黄 : 1 绿
68 株 $(\frac{2}{16})Yyrr$	→全部为皱粒, 子叶颜色分离 3 黄 : 1 绿
60 株 $(\frac{2}{16})YYRr$	→全部为黄色, 籽粒形状分离 3 圆 : 1 皱
67 株 $(\frac{2}{16})yyRr$	→全部为绿色, 籽粒形状分离 3 圆 : 1 皱
138 株 $(\frac{4}{16})YyRr$	→分离 9 黄、圆 : 3 黄、皱 : 3 绿、圆 : 1 绿、皱

从 F_2 群体基因型的鉴定, 也证明了独立分配规律的正确性。

4. 多对基因的遗传

当具有 3 对不同性状的植株杂交时, 只要决定 3 对性状遗传的基因分别在 3 对非同源染色体上, 它们的遗传都是符合独立分配规律的。例如以黄色、圆粒、红花植株和绿色、皱粒、白花植株杂交, F_1 全部为黄色、圆粒、红花。 F_1 的 3 对杂合基因分别位于 3 对染色体上, 减数分裂过程中, 这 3 对染色体

有 $2^3 = 8$ 种可能的分离方式，因而产生 8 种雌雄配子（YRC、YrC、yRC、YRc、yrC、Yrc、yRc、yrc），并且各种配子的数目相等。由于各种雌雄配子之间随机结合， F_2 将产生 64 种组合，8 种表现型，27 种基因型，归纳于表 4-4 中。

雌雄配子的自由组合，可以用棋盘方格图解法进行分析。但是，为了方便起见，复杂的基因组合也可以先将各对基因杂种的分离比例分解开，而后按同时发生事件的几率进行综合。例如，3 对独立基因杂种的杂交（YyRrCc × YyRrCc），可以看作是 3 个单基因杂种之间的杂交，即（Yy × Yy）（Rr × Rr）（Cc × Cc）。每一单基因杂种的 F_2 按 3 : 1 比例分离，因此 3 对独立基因杂种的 F_2 表现型的比例就是（3 : 1）³ 的展开。如有 n 对独立基因，则其 F_2 表现型的比例应为（3 : 1）ⁿ 的展开（表 4-5）。

由表 4-5 可见，只要各对基因都是属于独立遗传的，其杂种后代的分离就有一定的规律可循。就是说在 1 对等位基因的基础上，每增加 1 对等位基因， F_1 形成的不同配子种类就增加为 2 的倍数，即 2^n ； F_2 的基因型种类就增加为 3 的倍数，即 3^n ； F_1 配子可能的组合数就增加为 4 的倍数，即 4^n 。

表 4-4 豌豆黄色、圆粒、红花 × 绿色、皱粒、白花的 F_2 基因型、表现型及其 F_3 分离的比例

基因型	基因型比例	表现型	表现型比例	F_3 的分离比例
YYRRCC	1	黄色 圆粒 红花	27	不分离 黄色 : 绿色 = 3 : 1 圆粒 : 皱粒 = 3 : 1 红花 : 白花 = 3 : 1 黄圆 : 黄皱 : 绿圆 : 绿皱 = 9 : 3 : 3 : 1 圆红 : 圆白 : 皱红 : 皱白 = 9 : 3 : 3 : 1 黄红 : 黄白 : 绿红 : 绿白 = 9 : 3 : 3 : 1 黄圆红 : 黄圆白 : 黄皱红 : 绿圆红 : 黄皱白 : 绿圆白 : 绿皱红 : 绿皱白 = 27 : 9 : 9 : 9 : 3 : 3 : 3 : 1
YyRRCC	2			
YYRrCC	2			
YYRRCc	2			
YyRrCC	4			
YYRrCc	4			
YyRRCc	4			
YyRrCc	8			
yyRRCC	1	绿色 圆粒 红花	9	不分离 圆粒 : 皱粒 = 3 : 1 红花 : 白花 = 3 : 1 圆红 : 圆白 : 皱红 : 皱白 = 9 : 3 : 3 : 1
yyRrCC	2			
yyRRCc	2			
yyRrCc	4			

续表

基因型	基因型比例	表现型	表现型比例	F ₃ 的分离比例
YYrrCC YyrrCC YYrrCc YyrrCc	1 2 2 4	黄色 皱粒 红花	9	不分离 黄色:绿色 = 3:1 红花:白花 = 3:1 黄红:黄白:绿红:绿白 = 9:3:3:1
YYRrcc YyRRcc YYRrcc YyRrcc	1 2 2 4	黄色 圆粒 白花	9	不分离 黄色:绿色 = 3:1 圆粒:皱粒 = 3:1 黄圆:黄皱:绿圆:绿皱 = 9:3:3:1
yyrrCC yyrrCc	1 2	绿色 皱粒 红色	3	不分离 红花:白花 = 3:1
YYrrcc Yyrrcc	1 2	黄色 皱粒 白花	3	不分离 黄色:绿色 = 3:1
yyRRcc yyRrcc	1 2	绿色 圆粒 白花	3	不分离 圆粒:皱粒 = 3:1
yyrrcc	1	绿色 皱粒 白花	1	不分离

表 4-5 杂种杂合基因对数与 F₂ 表现型和基因型种类的关系

杂种杂合 基因对数	显性完全 时 F ₂ 表现 型种类	F ₁ 形成的 不同配子 种类	F ₂ 基因 型种类	F ₁ 产生的 雌雄配子的 可能组合数	F ₂ 纯合基 因型种类	F ₂ 杂合基 因型种类	F ₂ 表现型 分离比例
1	2	2	3	4	2	1	(3:1) ¹
2	4	4	9	16	4	5	(3:1) ²
3	8	8	27	64	8	19	(3:1) ³
4	16	16	81	256	16	65	(3:1) ⁴
5	32	32	243	1024	32	211	(3:1) ⁵
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	2 ⁿ	2 ⁿ	3 ⁿ	4 ⁿ	2 ⁿ	3 ⁿ ~2 ⁿ	(:1) ⁿ

5. 独立分配规律的应用

独立分配规律是在分离规律的基础上,进一步揭示了多对基因之间自由组合的关系。它解释了不同基因的独立分配是自然界生物发生变异的重要来源之一。

按照独立分配规律,在显性作用完全的条件下,亲本间有 2 对基因差异时, F_2 有 $2^2=4$ 种表现型; 4 对基因差异时, F_2 有 $2^4=16$ 种表现型。设两个亲本有 20 对基因的差别,而这些基因都是独立遗传的,那么 F_2 将有 $2^{20}=1,048,576$ 种不同的表现型。至于 F_2 的基因型数目就更为复杂了。这说明通过杂交造成基因的重新组合,是生物界多样性的重要原因之一。生物有了丰富的变异类型,可以广泛适应于各种不同自然条件,有利于生物的进化。

根据独立分配规律,在杂交育种工作中,可以有目的地组合两个亲本的优良性状,并可预测在杂交后代中出现的优良性状组合及其大致的比例,以便确定育种工作的规模。例如,某水稻品种无芒而感病,另一水稻品种有芒而抗病。已知有芒(A)对无芒(a)为显性,抗病(R)对感病(r)为显性。在有芒、抗病(AARR)×无芒、感病(aarr)的组合中, F_2 分离出无芒抗病(aaR_)植株的机会占 3/16,其中纯合的(aaRR)植株占 1/3,杂合的(aaRr)植株占 2/3。在 F_3 ,纯合体不再分离,而杂合体将继续分离。因此,如在 F_3 希望获得 10 个稳定遗传的无芒、抗病(aaRR)株系,那么可以预计,在 F_2 至少要选择 30 株以上无芒、抗病的植株,供 F_3 株系鉴定。

三、孟德尔规律的扩展

1. 显隐性关系的相对性

(1) 显性现象的表现

① 完全显性 (complete dominance)

孟德尔在豌豆杂交试验中的 7 对相对性状中, F_1 所表现的性状都和亲本之一完全一样,这样的显性表现称为完全显性。通常所说的显性就是指完全显性。

② 不完全显性 (incomplete dominance)

有些性状,其杂种 F_1 的性状表现是双亲性状的中间型,这称为不完全显性。例如,金鱼草 (Antirrhinum majus) 花色的遗传 (图 4-8)。红花亲本

(RR) 和白花亲本 (rr) 杂交, F_1 (Rr) 的花色不是红色, 而是粉红色。 F_2 群体的基因型分离为 $1RR : 2Rr : 1rr$, 即其中 $1/4$ 的植株开红花, $2/4$ 的植株开粉红花, $1/4$ 的植株开白花。因此, 在不完全显性时, 表现型和其基因型是一致的。

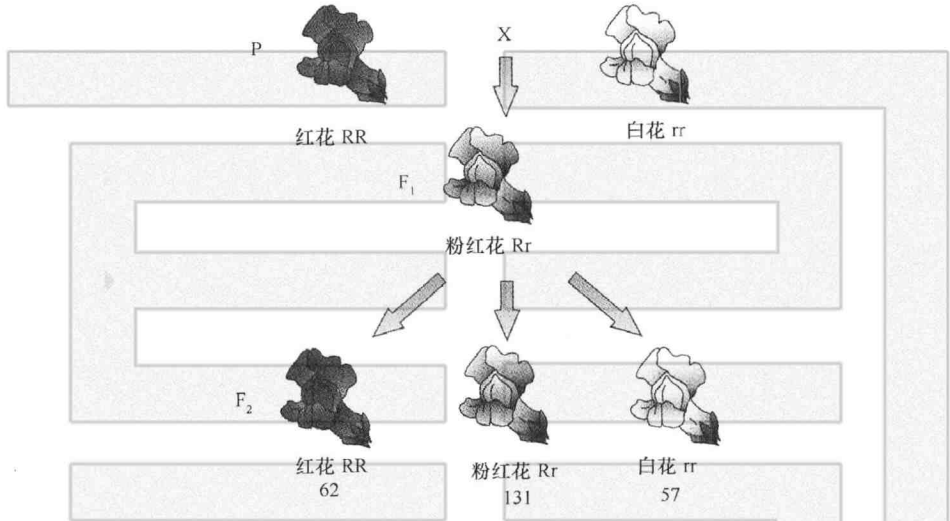


图 4-8 金鱼草花色的遗传

人的天然卷发也是由一对不完全显性基因决定的, 其中卷发基因 W 对直发基因 w 是不完全显性的。WW 的头发十分卷曲, Ww 的头发卷曲程度中等, ww 则是直发。

③ 共显性 (codominance)

如果双亲的性状同时在 F_1 个体上表现出来, 即一对等位基因的两个成员在杂合体中都表达的遗传现象称为共显性或并显性。例如, 正常人红血球细胞呈碟形, 镰形贫血症患者的红血球细胞呈镰刀形 (图 4-9)。这种贫血症患者和正常人结婚所生的子女, 其红血球细胞既有碟形, 又有镰刀形, 这就是共显性的表现。

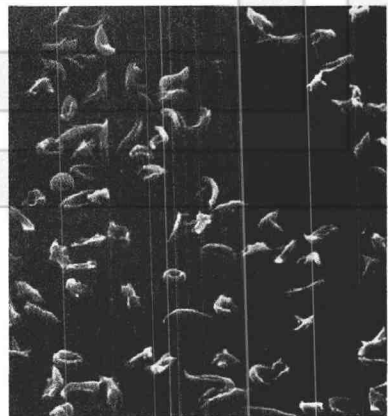


图 4-9 镰刀形贫血症的红血球细胞

(引自 Tamarin, 2002)

人类的血型中存在有一种 MN 血型系统。M 血型是由 L^M 基因控制, 其红细胞

存在 M 抗原；N 血型是由 L^N 基因控制，其细胞存在 N 抗原。当 M 血型人 ($L^M L^M$) 与 N 血型人 ($L^N L^N$) 结婚，他们所生的小孩血型为 MN 型，其红细胞既有 M 抗原又有 N 抗原，而不是两种抗原的中间类型。任何一个人，要么是 M 血型，要么是 N 血型，要么是 MN 血型。

④ 镶嵌显性 (mosaic dominance)

双亲的性状在后代的同一个体不同部位表现出来，形成镶嵌图式，这种显性现象称为镶嵌显性。例如，我国学者谈家桢对异色瓢虫 (*Harmonia axyridis*) 色斑遗传的研究，他用黑缘型鞘翅 ($S^{Au} S^{Au}$) 瓢虫 (鞘翅前缘呈黑色) 与均色型鞘翅 ($S^F S^F$) 瓢虫 (鞘翅后缘呈黑色) 杂交，子一代杂种 ($S^{Au} S^F$) 既不表现黑缘型，也不表现均色型，而出现一种新的色斑，即上下缘均呈黑色。在植物中，如玉米花青素的遗传也表现出这种现象。

鉴别性状的显性表现也取决于所依据的标准而改变。例如，孟德尔根据豌豆种子的外形，发现圆粒对皱粒是完全显性。但是，如果用显微镜检查豌豆种子淀粉粒的形状和结构，可以发现纯合圆粒种子的淀粉粒持水力强，发育完善，结构饱满；纯合皱粒种子的淀粉粒持水力较弱，发育不完善，表现皱缩；而 F_1 杂合种子的淀粉粒，其发育和结构是前面两者的中间型，但种子外形是圆粒。故从种子外表观察，圆粒对皱粒是完全显性，但深入研究淀粉的形态结构，则可发现它是不完全显性。

(2) 显性与环境的影响

当一对相对基因处于杂合状态时，为什么显性基因能决定性状的表现，而隐性基因不能，是否由于显性基因直接抑制了隐性基因的作用？试验证明，相对基因之间的关系，并不是彼此直接抑制或促进的关系，而是分别控制各自所决定的代谢过程，从而控制性状的发育。例如，兔子的皮下脂肪有白色和黄色的不同。白色由显性基因 Y 决定；黄色由隐性基因 y 决定。白脂肪的纯种兔子 (YY) 和黄脂肪的纯种兔子 (yy) 杂交， F_1 (Yy) 的脂肪白色。用 F_1 的雌兔 (Yy) 和雄兔 (Yy) 进行近亲交配， F_2 群体中， $3/4$ 个体是白脂肪， $1/4$ 个体是黄脂肪。兔子的主要食料是绿色植物，绿色植物中除含有叶绿素以外，还有大量的黄色素。显性基因 Y 控制合成一种黄色素分解酶能分解黄色素，隐性基因 y 则没有这种作用。所以，基因型为 YY 或 Yy 的兔子，由于细胞内有 Y 基因，能合成黄色素分解酶，因而能破坏吃进的黄色素使脂肪

内没有黄色素的积存，故而脂肪是白色。基因型为 yy 的兔子，由于细胞内不能合成黄色素分解酶，所以脂肪是黄色的。由此可知，显性基因 Y 与白脂肪表现型的关系、隐性基因 y 与黄色脂肪的关系都是间接的。从兔子脂肪颜色的遗传来看，显性基因与相对隐性基因之间的关系决不是显性基因抑制了隐性基因的作用，而是它们各自参加一定的代谢过程，分别起着各自的作用。一个基因是显性还是隐性取决于它们各自的作用性质，取决于它们能不能控制某个酶的合成。

显隐性关系有时受到环境的影响，或者为其他生理因素如年龄、性别、营养、健康状况等所左右。

例如，金鱼草的红花品种与淡黄色花品种杂交， F_1 在不同条件下的表型不同。在低温、光充足的条件下，花为红色，那么红色为显性；在温暖、遮光条件下，花为淡黄色，那么红色为隐性。如果培育在温暖、光充足的条件下，花为粉红色，表现不完全显性。可见环境条件改变时，显隐性关系也可相应的发生改变。

石竹 (*Dianthus barbatus*) 花的白色和暗红色是一对相对性状。用开白花的植株与开暗红色花的植株杂交，杂种 F_1 的花最初是纯白的，以后慢慢变为暗红色。这样个体发育中显隐性关系也可相互转化。

有角羊与无角羊杂交， F_1 雄性有角，雌性无角。因此，杂种有无角与性别有关。

2. 复等位基因

尽管在孟德尔的豌豆杂交试验中，单位性状均受一对等位基因控制，随后的遗传研究发现，同一位点的基因可能有两种以上的形式。遗传学把同源染色体相同位点上存在的三个或三个以上的等位基因称为复等位基因 (multiple alleles)。复等位基因在生物界广泛存在，如人类的 ABO 血型遗传。由于二倍体生物中等位基因总是成对存在，因此每一个个体最多只能具有复等位基因的两种成员。各复等位基因存在于同种生物群体的不同个体中，决定同一单位性状内多种差异的遗传，增加了生物多样性，为生物适应性和育种提供了更丰富的资源。

人类的 ABO 血型由 3 个复等位基因 I^A 、 I^B 和 I^O 决定。 I^A 与 I^B 之间表示共显性 (无显隐性关系)，而 I^A 和 I^B 对 I^O 都是显性，所以这 3 个复等位基因

组成 6 种基因型，但表现型只有 4 种（表 4-6）。

表 4-6 人类 ABO 血型 and 基因型及其凝集反应

表型 (血型)	基因型	抗原	抗体	血清	血细胞
AB	$I^A I^B$	A、B	—	不能使任一血型的红细胞凝集	可被 O、A、B 型的血清凝集
A	$I^A I^A$ $I^A I^O$	A	β	可使 B 及 AB 型的红细胞凝集	可被 O、B 型的血清凝集
B	$I^B I^B$ $I^B I^O$	B	α	可使 A 及 AB 型的红细胞凝集	可被 O、A 型的血清凝集 OA
O	$I^O I^O$	—	$\alpha、\beta$	可使 A、B 及 AB 型的红细胞凝集	不能被任一血型的血清凝集

3. 致死基因

致死基因 (lethal alleles)，是指当其发挥作用时导致生物体不能存活的基因。致死作用可以发生在个体发育的各个时期。第一次发现致死基因是 1904 年，法国遗传学家 Cuenot 在小鼠中发现黄色皮毛的品种不能稳定遗传 (图 4-10)。黄色与黑色小鼠交配，其后代总会出现黑色鼠，而且黄色、黑色的比率往往是 2 : 1，而不是通常应出现的 3 : 1 的分离比。黑色鼠的后代都是黑色，证明它是黑色的隐性纯合体。黄色与黑色小鼠杂交的子代则是 1 黄色 : 1 黑色，表明黄色皮毛老鼠是杂合体。根据孟德尔原理，既然黄色是杂合体，其自交结果却不出现 3 : 1 比率，唯一的可能性就是其中纯合的黄色个体在胚胎发育过程中死亡了。后来的研究证明了这一推断。

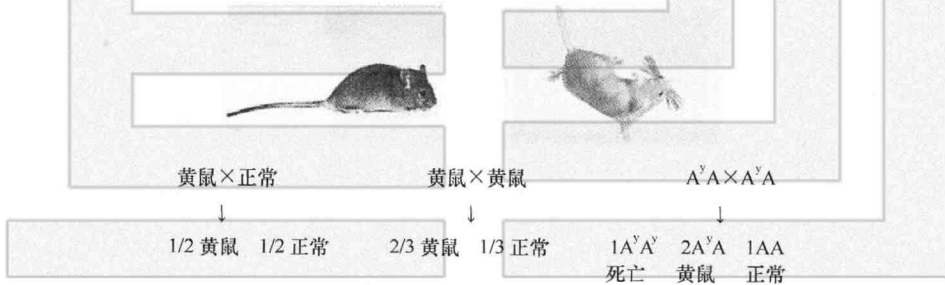


图 4-10 致死基因

致死基因包括显性致死基因 (dominant lethal alleles) 和隐性致死基因 (recessive lethal alleles)。黄鼠这个例子， A^y 基因在控制毛皮的颜色上是显性的，但在“致死”这个表型上属于隐性 (纯合致死)。隐性致死基因只有在隐性纯合时才能使个体死亡。植物中常见的白化基因也是隐性致死基因。因为不能形成叶绿素，它使植物成为白化苗，最后植株死亡。显性致死基因在杂

合体状态时就可导致个体死亡。如人的神经胶症 (epiloia) 基因只要一份就可引起皮肤的畸形生长、严重的智力缺陷以及多发性肿瘤，所以对该基因是杂合的个体在很年轻时丧失生命。

4. 基因互作

基因分离和独立分配规律不断得到实验的证明。根据独立分配规律，两对基因杂交 F_2 出现 $9:3:3:1$ 的性状分离比例。但是，两对等位基因的自由组合却不一定会出现 $9:3:3:1$ 的性状分离比例，或者即使出现 $9:3:3:1$ 的性状分离比例，也并非常见的双显性个体占 $9/16$ ，单显、单隐性个体占 $3/16$ ，双隐性个体占 $1/16$ 。研究表明这是由于不同对基因间相互作用共同决定同一单位性状表现的结果。这种现象，称为基因互作 (interaction of genes)。

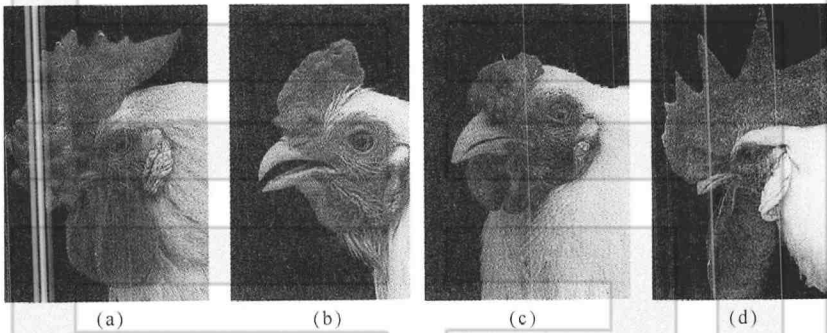


图 4-11 鸡冠的形状

a. 玫瑰冠; b. 豌豆冠; c. 胡桃冠; d. 单冠

例如鸡冠的遗传。鸡冠的形状很多，除常见的单冠外，还有玫瑰冠、豌豆冠和胡桃冠等。这些不同种类的鸡冠是品种的特征之一 (图 4-11)。如果把豌豆冠的鸡和玫瑰冠

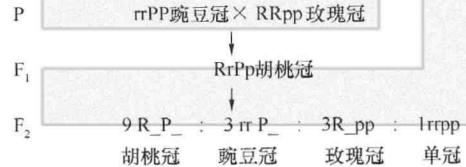


图 5-12 鸡冠性状的遗传

的鸡交配，子一代的鸡为胡桃冠，与亲本不同。子一代个体间相互交配，得到子二代，它们的鸡冠有胡桃冠、豌豆冠、玫瑰冠和单冠四种，比例大体上接近 $9:3:3:1$ 。

为什么会出这种现象呢？原因是鸡冠的形状同时由两对基因控制。

控制豌豆冠的显性基因是 P，而控制玫瑰冠的显性基因是 R（图 4-12）。豌豆冠的鸡没有基因 R，因此基因型是 rrPP；玫瑰冠的鸡没有基因 P，因此基因型是 RRpp。当两亲本杂交，子一代的基因型是 RrPp，由于 P 基因和 R 基因的相互作用，出现了胡桃冠。子一代的公鸡和母鸡都形成 RP、Rp、rP、rp 4 种配子，数目相等。根据自由组合规律，子二代出现四种表现型：胡桃冠（R_P_）、玫瑰冠（R_pp）、豌豆冠（rrP_）、和单冠（rrpp），比例为 9 : 3 : 3 : 1。

基因相互作用决定生物性状的表现情况复杂，存在多种互作方式。下面就两对独立遗传的等位基因的互作表现，举例说明各种互作方式。

(1) 互补作用 (complementary effect)

两对独立遗传基因分别处于纯合显性或杂合状态时，共同决定一种性状的发育。

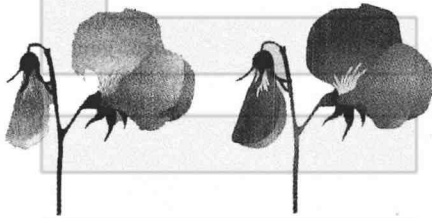


图 4-13 香豌豆的白花和紫花

当只有一对基因是显性，或两对基因都是隐性时，则表现为另一种性状。这种基因互作的类型称为互补作用。发生互补作用的基因称为互补基因 (complementary gene)。

例如，在香豌豆 (*Lathyrus odoratus*) (图 4-13) 中有两个白花品种，二者

杂交产生的 F₁ 开紫花。F₁ 植株自交，其 F₂ 群体分离为 9/16 紫花：7/16 白花。对照独立分配规律，可知该杂交组合是两对基因的分离。F₁ 和 F₂ 群体的 9/16 植株开紫花，说明两对显性基因的互补作用。如果紫花所涉及的两个显性基因为 C 和显性基因 P，就可以确定杂交亲本、F₁ 和 F₂ 各种类型的基因型如下：

P 白花 CCpp × 白花 ccPP

F₁ 紫花 CcPp

↓ ⊗

F₂ 9 紫花 (C_P_) : 白花 (3C_pp + 3ccP_ + 1ccpp)

上述试验中，F₁ 和 F₂ 的紫花植株表现其野生祖先的性状，这种现象称为返祖遗传 (atavism)。这种野生香豌豆的紫花性状取决于两种基因的互补。这两种显性基因在进化过程中，如果显性基因 C 突变成隐性基因 c，产生了一种

白花品种；如果显性基因 P 突变成隐性基因 p，又产生另一种白花品种。当这两个品种杂交后，两对显性基因重新结合，于是出现了祖先的紫花。

(2) 积加作用 (additive effect)

在有些试验中，发现两种显性基因同时存在时产生一种性状，单独存在时能分别表现相似的性状，两种显性基因均不存在时又表现第三种性状，这种基因互作称为积加作用。例如，南瓜 (Cucurbita pepo) 有不同的果形 (图4-14)，圆球形对扁盘形为隐性，长圆形对圆球形为隐性。

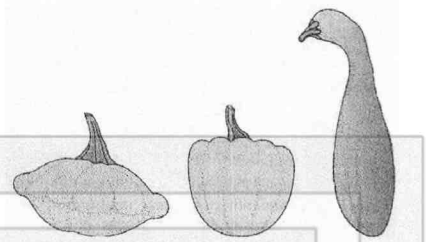


图 4-14 南瓜果形

如果用两种不同型的圆球形品种杂交，F₁ 产生扁盘形，F₂ 出现三种果形：9/16 扁盘形，6/16 圆球形，1/16 长圆形。对它们的遗传行为分析如下：

P 圆球形 AAbb × 圆球形 aaBB



F₁ 扁盘形 AaBb



F₂ 9 扁盘形 (A_B_) : 6 圆球形 (3A_bb + 3aaB_) : 1 长圆形 (aabb)

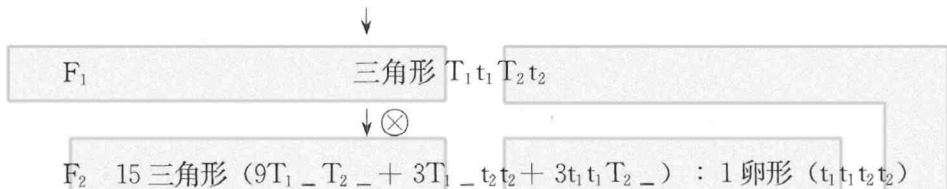
从以上分析可知，两对基因都是隐性时，形成长圆形；只有显性基因 A 或 B 存在时，形成圆球形；A 和 B 同时存在时，则形成扁盘形。

(3) 重叠作用 (duplicate effect)

不同对基因互作时，对表现型产生相同的影响，F₂ 产生 15 : 1 的比例，这种基因互作称为重叠作用。这类表现相同作用的基因，称为重叠基因 (duplicate gene)。例如，芥菜 (Bursa bursa-pastoria) 中常见的植株是三角形蒴果，极少数植株是卵形蒴果 (图 4-15)。将这两种植株杂交，F₁ 全是三角形蒴果。F₂ 分离为 15/16 三角形蒴果，1/16 卵形蒴果。卵形蒴果的后代不再分离；三角形蒴果的后代有一部分不分离，一部分分离为 3/4 三角形蒴果，1/4 卵形蒴果，还有一部分分离为 15/16 三角形蒴果和 1/16 卵形蒴果。由此可知，上述试验中 F₂ 出现 15 : 1 的比例，实际上是 9 : 3 : 3 : 1 比例的变型，只是前三种表现型没有区别。这显然是由于每对基因中的显性基因具有使蒴

果表现为三角形的相同作用。如果缺少显性基因，即表现为卵形蒴果。如用 T_1 和 T_2 表示这两个显性基因，则三角形蒴果亲本的基因型为 $T_1T_1T_2T_2$ ，卵形蒴果亲本的基因型为 $t_1t_1t_2t_2$ 。F₁ 和 F₂ 的各种基因型如下：

P 三角形 $T_1T_1T_2T_2$ × 卵形 $t_1t_1t_2t_2$



当杂交试验涉及三对重叠基因时，则 F₂ 的分离比例将相应的为 63 : 1，其余类推。在这里它们的显性基因作用虽然相同，但并不表现累积的效应。基因型内的显性基因数目不等，并不改变性状的表现，只要有一个显性基因存在，就能使显性性状得到发育。但在有些情况下，重叠基因也表现累加的效应。

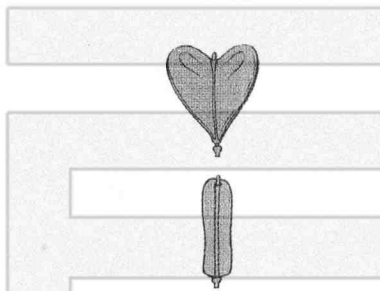


图 4-15 芥菜的三角形和卵圆形蒴果

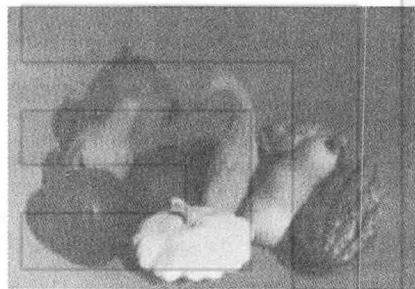


图 4-16 西葫芦果皮颜色

(4) 显性上位作用 (epistatic dominance)

两对独立遗传基因共同对一单位性状发生作用，而且其中一对基因对另一对基因的表现有遮盖作用，这种情形称为上位性 (epistasis)。反之，后者被前者所遮盖，称为下位性 (hypostasis)。起遮盖作用的基因如果是显性基因，称为上位显性基因。例如，影响西葫芦果皮 (图 4-16) 的显性白皮基因 (W) 对显性黄皮基因 (Y) 有上位性作用。当 W 基因存在时能阻碍 Y 基因的作用，表现为白色；缺少 W 时，Y 基因表现其黄色作用；如果 W 和 Y 都不存在，则表现 y 基因的绿色。

P 白皮 WWYY × 绿皮 wwyy
 ↓
 F₁ 白色 WwYy
 ↓ ⊗
 F₂ 12 白皮 (9W_Y_ + 3W_yy) : 3 黄皮 (wwY_) : 1 绿皮 (wwyy)
 (5) 隐性上位作用 (epistatic recessiveness)

在两对互作的基因中，其中一对隐性基因对另一对基因起上位性作用，称为隐性上位作用。例如，玉米胚乳蛋白质层颜色的遗传，当基本色泽基因 C 存在时，另一对基因 Prpr 都能表现各自的作用，即 Pr 表现紫色，pr 表现红色。缺 C 基因时，隐性基因 c 对 Pr 和 pr 起上位作用，使得 Pr 和 pr 都不能表现其性状而呈现白色。

P 红色蛋白质层 CCprpr × 白色蛋白质层 ccPrPr
 ↓
 F₁ 紫色 CcPrpr
 ↓ ⊗
 F₂ 9 紫色 (C_Pr_) : 3 红色 (C_prpr) : 4 白色 (3ccPr_ + 1ccprpr)

上位作用和显性作用不同，上位作用发生于两对非等位基因之间，而显性作用则发生于同一对等位基因的两个成员之间。

(6) 抑制作用 (inhibiting effect)

在两对独立基因中，其中一对显性基因，本身并不控制性状的表现，但对另一对基因的表现有抑制作用，称为抑制基因。例如，家蚕中有结黄茧和结白茧的个体。如果将结白茧和结黄茧的中国品种杂交，F₁ 全为黄茧，说明中国品种的白茧是隐性的。但将结白茧和结黄茧的欧洲品种杂交，F₁ 全为白茧，表明欧洲品种白茧是显性的。如果将上述两个白茧品种杂交，F₁ 结白茧，F₂ 表现白茧与黄茧的比例为 13 : 3。

这种遗传方式可以用下列杂交说明：黄茧基因为 Y，白茧基因为 y，另一个非等位的抑制基因 I 可以抑制黄茧基因 Y 的作用。

P 白茧 IYY × 白茧 iiyy
 ↓
 F₁ 白茧 IiYy
 ↓ ⊗
 F₂ 13 白茧 (9I_Y_ + 3I_yy + 1iiyy) : 3 黄茧 (iiY_)

当基因 I 存在时，抑制了 Y 基因的作用。只有 I 不存在时，Y 基因的作用才能表现。由此可见，有些基因本身并不能独立地表现任何可见的表型效应，但可以完全抑制其他非等位基因的作用。上位作用和抑制作用不同，抑制基因本身不能决定性状，而显性上位基因除遮盖其他基因的表现外，本身还能决定性状。

以上只讨论了两对独立基因共同决定同一性状时所表现的各种情况，但这并不是说，基因的相互作用只限于两对非等位基因。如果共同决定同一性状的基因对数更多，后代表现分离的比例将更加复杂。上述两对基因互作的关系，可归纳为模式图 4-17。

基因互作方式	基因型				基因型比例
	A-B-	A-bb	aaB-	aabb	
无互作	9	3	3	1	9:3:3:1
显性互补	9		7		9:7
抑制作用	9	3	4		13:3
隐性上位	9	3	4		9:3:4
显性上位		12	3	1	12:3:1
重叠作用		15		1	15:1
积加作用	9		6	1	9:6:1

图 4-17 两对基因互作的模式图

图 4-17 以两对基因 Aa 和 Bb 的互作为例，假定各对基因的显性作用是完全的，按自由分离和独立分配规律，F₂ 出现的 9 种基因型在基因不发生互作的情况下，四种表现型的比例为 9:3:3:1。这是一个基本类型。在此基础上，由于基因互作的情况不同，才出现六种不同方式的表现型和比例。而各种表现型的比例都是在两对独立基因分离比例 9:3:3:1 的基础上演变而来的。这里只是表现型的比例有所改变，而基因型的比例仍然与独立分配一致。由此可知，由于基因互作，杂交分离的类型和比例与典型的孟德尔遗传的比例虽然不同，但这并不能因此否定孟德尔遗传的基本规律，而应该认为这是对它进一步的深化和发展。

实际上，基因互作可以分为基因内互作 (intragenic interaction) 和基因

间互作 (intergenic interaction)。基因内互作是指等位基因间的显隐性作用。基因间互作指不同位点非等位基因相互作用, 表现为互补、抑制、上位性等。性状的表现都是在一定环境条件下, 通过这两类基因互作共同或单独发生作用的产物。上述各基因互作的实例中, 两对非等位基因各自分别都表现完全显性作用。也有少数情况, 一对等位基因表现完全显性, 另一对表现不完全显性。例如, 有一种牛的毛色, 红色 (A) 对白色 (a) 为不完全显性, 杂合型 (Aa) 表现灰色。无角 (B) 对有角 (b) 为完全显性。两对杂合基因型的 F₂ 代的分离比例为: 红色无角 (AAB₋) 3/16 : 灰色无角 (AaB₋) 6/16 : 红色有角 (AAbb) 1/16 : 灰色有角 (Aabb) 2/16 : 白色无角 (aaB₋) 3/16 : 白色有角 (aabb) 1/16。上述显性上位作用、隐性上位作用和抑制作用, 都是基因内和基因间相互共同作用的结果。

5. 多因一效和一因多效

以上基因互作的实例, 说明了一个单位性状的表现并不都受一个位点基因控制, 而经常受许多不同位点基因的影响。许多基因影响同一个性状的表现称为“多因一效”(multigenic effect)。例如, 已知玉米正常叶绿素的形成与 50 多对不同的基因有关, 其中的任何一对发生改变, 都会影响叶绿素的消失或改变。玉米籽粒胚乳蛋白质层的紫色, 已知是由 A₁、A₂、A₃、C、R、Pr 六对不同的显性基因和一对隐性抑制基因 i 共同决定的。

另一方面, 一个基因也可以影响许多性状的发育, 这称为“一因多效”(pleiotropism)。孟德尔在豌豆杂交试验中就曾发现, 红花的植株同时结灰色种皮的种子, 叶腋上有黑斑; 开白花的植株结淡色种皮的种子, 叶腋上没有黑斑。在杂交后代中, 这三种性状总是连在一起出现, 仿佛是一个遗传单位。可见决定豌豆红花或白花的基因不但影响花色, 而且也控制种子颜色和叶腋上黑斑的有无。水稻的矮生基因也常有多效性的表现, 它除表现矮化的作用以外, 一般还有提高分蘖力、增加叶绿素含量和扩大栅栏细胞的直径等作用。

从生物个体发育的整体观念出发, “多因一效”和“一因多效”现象是不难理解的。一方面, 一个性状的发育是由许多基因所控制的许多生化过程连续作用的结果。另一方面, 如果某一基因发生了改变, 其影响虽然只有一个以该基因为主的生化过程, 但也会影响与该生化过程有联系的其他生化过程,

从而影响其他性状的发育。

第二节 连锁与互换定律

一、连锁现象的发现

正如 T. H. Morgan 发现例外白眼雄蝇提出了性连锁遗传，以及 Bridges 关于果蝇白眼遗传的研究中发现了初级例外和次级例外，促使他注意到染色体不分开现象，直到最终证明了遗传的染色体学说一样，人们从两对基因的杂交实验的 F_2 分离比与预期的 $9:3:3:1$ 出现的例外，发现了连锁遗传的现象。遗传学的深入研究不仅证明了染色体带有许多基因，而且证明了这些基因在染色体上是以直线排列的。这便是遗传学的第三定律——基因的连锁与交换定律。

W. Bateson 和 R. C. Punnett 所研究的香豌豆的花有紫色和红色的，花粉粒的形状有长形和圆形的。将紫花、长花粉粒和红花、圆花粉粒的植株作亲本进行杂交， F_1 都是紫花、长花粉粒，可见紫、长为显性。但是 F_2 的 4 种表型的比率却不符合 $9:3:3:1$ ，其中紫长和红圆的比率远远超出 $9/16$ 和 $1/16$ ，而相应的紫、圆和红、长却大大少于 $3/16$ （表 4-7）。

表 4-7 香豌豆紫长×红圆交试

F_2	紫长	紫圆	红长	红圆	总数
实计数	4831	390	393	1338	6952
预计数	3910.5	1303.5	1303.5	434.5	6952

上述结果进行 χ^2 检验时， $\chi^2 = 3371.58$ 。如此可观的 χ^2 数值，无疑说明实计频数与预计频数的极其显著的差异不可能由随机原因所造成。科学的态度是重复实验，视其同类现象是否稳定再现，如果仍旧出现异常，则应提出新的理论（或假设）来加以合理的解释并设计新的实验来验证其假设的正确性。

Bateson 又把亲本的性状组合调换，进行重复实验，用紫花、圆形花粉粒

和红花、长形花粉粒的植株杂交， F_1 自然还是紫、长， F_2 仍然不符合孟德尔的自由组合定律，但这里的紫、圆和红、长都高出预期的数目（表 4-8），其中 $\chi^2 = 32.40$ ，证明它仍是显著不符合 9:3:3:1 的。

表 4-8 香豌豆紫圆×红长交试

F_2	紫长	紫圆	红长	红圆	总数
实计数	226	95	97	1	419
预计数	235.8	78.5	78.5	26.2	419

从这两个实验的数据可以看出一种共同的倾向：即与自由组合定律所预期结果相比较， F_2 中性状的亲本组合类型远远多于重组组合的类型，这等于说，在 F_1 杂种形成配子时两对基因可能发生的 4 种类型中，有更多保持亲代原来组合的倾向，而且这种倾向与显隐性无关。这是在自由组合定律方面第一次出现的显著的例外，无疑，这是一个重要的发现。然而，Bateson 等人未能从中得到启发，终于被自己所发现的新现象所迷惑。他除了给前一种实验的亲本组合（紫、长，红、圆）称为互引相（coupling phase），称后一实验的亲本组合（紫、圆，红、长）为互斥相（repulsion phase）两个名词术语外，并未提出理论上的解释。前述两个名词的确切定义为：当两个非等位基因 a 和 b 处在一个染色体上，而在其同源染色体上带有野生型 A、B 时，这些基因被称为处于互引相（AB/ab）；若每个同源染色体上各有一个突变基因和一个野生型基因，则称为互斥相（Ab/aB）。

这个迷案直到 1912 年才被摩尔根创造性的研究和思想所突破，从而揭开了遗传学发展史上新的一页。

二、完全连锁与不完全连锁

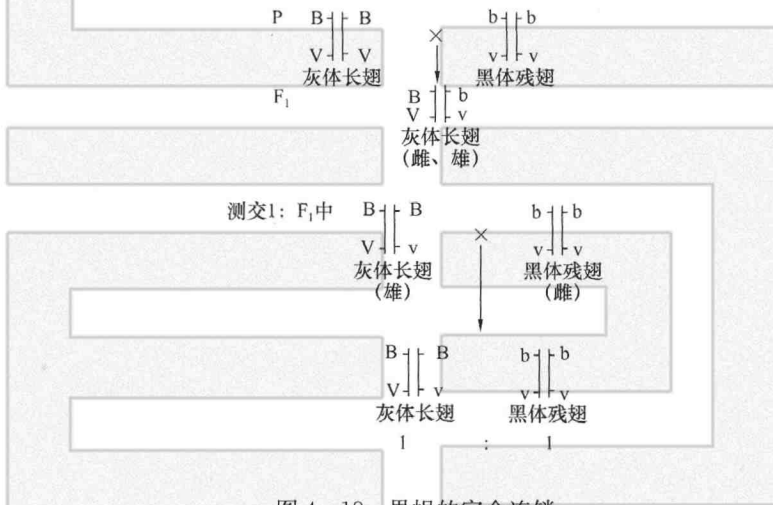
Morgan 用果蝇进行了大量的杂交试验，提出了连锁互换定律，被后人誉为遗传的第三定律。

在黑腹果蝇中，灰体（B）对黑体（b）是显性，长翅（V）对残翅（v）为显性，这两对基因在常染色体上。他用灰体长翅（BBVV）和黑体残翅（bbvv）的果蝇杂交， F_1 都是灰体长翅（BbVv）。用 F_1 的杂合体进行下列两种方式的测交，所得到的结果却完全不同：

1. 取 F_1 雄蝇和黑体残翅的雌蝇测交。按两对基因的自由组合定律来预测, 应产生灰体长翅 ($BbVv$)、灰体残翅 ($Bbvv$)、黑体长翅 ($bbVv$)、黑体残翅 ($bbvv$) 4 种表型的后代, 且比率为 $1:1:1:1$ 。可是实验的结果并非如此, 而是只有数目相同的灰体长翅 ($BbVv$) 和黑体残翅 ($bbvv$) 两种类型。

2. 如果用 F_1 雌蝇和黑体残翅的雄蝇测交, 后代中就出现了灰体长翅、灰体残翅、黑体长翅和黑体残翅 4 种类型, 但比例却是 $0.42:0.08:0.08:0.42$ 。

摩尔根的解释是: 假定基因 B 和 V 同处于一条染色体上, 基因 b 和 v 同处于同源染色体的另一条染色体上。他将这种处于同一条染色体上的基因遗传时较多地联系在一起的现象称为连锁 (linkage)。对于上述第一种测交结果摩尔根称完全连锁 (complete linkage), 如图 4-18。



由于 F_1 杂合雄蝇 ($BbVv$) 只产生两种类型的配子且数目相等, 所以用双隐性雌蝇测交时, 后代只能有两种表型而且为 $1:1$ 。在果蝇中雄蝇是完全连锁的, 即在同一染色体上的连锁基因 (如 B 、 V 基因) 100% 联系在一起传递到下一代。

第二种方式的测交, 后代出现数目不等的 4 种表型, 其中亲本的类型远远多于新类型。摩尔根称这种连锁为不完全连锁 (incomplete linkage) (图 4-19)。新类型是由于同源染色体上的不同对等位基因之间重新组合的结果, 这种现象称为重组 (recombination)。

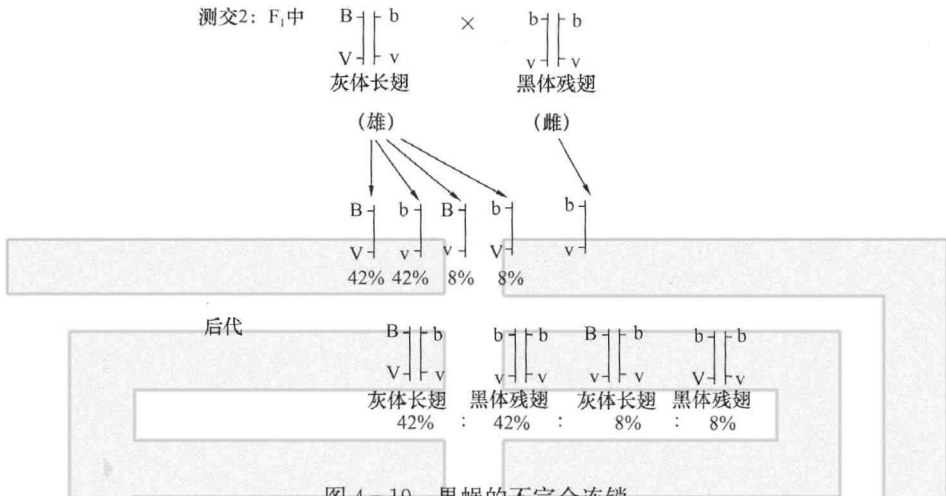


图 4-19 果蝇的不完全连锁

在本项测交实验中，出现的新类型共占 16%，该数值表示基因 Bb 和 Vv 之间的重组频率。

大量的实验资料表明，连锁现象在生物界普遍存在。凡是位于同一对染色体上的基因群，均称为一个连锁群 (linkage group)。摩尔根等人在 1914 年已发现黑腹果蝇有 4 个连锁群。到 1967 年为止，在果蝇中至少测定了将近 1000 个基因分别属于这 4 个连锁群。第 I 连锁群包括全部伴性遗传的基因，因为果蝇的第 1 染色体为性染色体；第 II 和第 III 连锁群分别相当于两对大的 V 字形的第 2 和第 3 染色体；第 IV 连锁群的基因数目最少，相当于最小的点状第 4 染色体。

在遗传学上，凡是充分研究过的生物中，连锁群的数目都应该等于单倍体染色体数 (n)。如玉米中已研究了几千个基因，确定它们有连锁关系的基因就有几百个，但它们分别属于 10 个连锁群，而玉米的染色体也正好是 10 对 ($n=10$)。链孢霉 $n=7$ ，连锁群有 7 个。水稻 $n=12$ ，连锁群就是 12。有些生物目前已发现的连锁群数少于单倍染色体数，如：家兔 $n=22$ ，连锁群是 11；而家蚕 $n=28$ ，连锁群却是 27；牵牛花 $n=15$ ，连锁群是 12 等等。这可能是由于研究尚不充分，或由于某些染色体上可供检出的基因数较少等缘故。继续深入研究，预期它们的连锁群数目将会达到各自的单倍体染色体数目 n ，但不会超过 n 。这也是遗传的染色体理论的有力证据。

三、重组频率的测定

某两个特定基因是否连锁，需要通过对其杂交 F_1 进行测交来确定，如果测交后代的比率为 $1:1:1:1$ ，则是自由组合的，如果亲组型高于预期比率，而重组型又显著地低于预期的比率，则说明它们是连锁的。遗传学以测交子代中出现的重组型频率来测定在这样的杂交中所表现出的连锁程度。求重组频率 (recombination frequency, RF) 的公式是：

$$\text{重组频率 (RF)} = \frac{\text{重组型数目}}{\text{亲组型数目} + \text{重组型数目}}$$

用玉米为材料研究基因的连锁，有如下优越性：①很多性状可以在种子上看到，种子虽然长在母本植株的果穗上，但它们已是子代产物；②同一果穗上有几百粒种子，便于计数分析；③雌、雄蕊长在不同花序上，去雄容易，便于杂交；④它是一种经济作物，某些实验结果有经济价值。

现以下列玉米的测交实验说明重组频率测定方法。玉米籽粒的糊粉层有色 (C) 对糊粉层无色 (c) 为显性；饱满种子 (Sh) 对凹陷种子 (sh) 为显性。基因 Sh 与 C 是连锁的。杂交亲本为互引相 $C\ Sh/C\ Sh \times c\ sh/c\ sh$ 时，有如下结果：

P	$C\ Sh / C\ Sh$	×	$c\ sh / c\ sh$
	有色饱满		无色凹陷
F_1	$C\ Sh / c\ Sh$	×	$c\ sh / c\ sh$
	有色饱满		无色凹陷
测交后代	$C\ Sh / c\ sh$	$C\ sh / c\ Sh$	$c\ Sh / c\ sh$
	有色饱满	有色凹陷	无色饱满
	4032	149	152
			4035
			无色凹陷
			4035
			8368
频率	$\frac{4032}{8368}$		$\frac{4035}{8368}$
		重组型	亲组型
		$\frac{301}{8368}$	$\frac{4035}{8368}$

显而易见：测交后代中亲组型的频率 = $\frac{8067}{8368} = 96.4\%$ ，重组频率 (RF) = $\frac{301}{8368} = 3.6\%$ 。另外，某次杂交在互斥相 $c\ Sh/c\ Sh \times C\ sh/C\ sh$ 中进行，得到下列结果：

在此，重组频率 (RF) = $\frac{1310}{44595} = 2.94\%$ 。

P	c Sh / c Sh	×	C Sh / C Sh
	无色饱满	↓	有色凹陷
F ₁	C Sh/c Sh	×	c Sh/c Sh
	有色饱满	↓	无色凹陷
测交后代	C Sh/c Sh	C Sh/c Sh	c Sh/c Sh
	有色饱满	有色凹陷	无色饱满
	638	21 379	21 906
			672



由此可知，不论用哪种基因组合的交配方式，测交的结果都是亲组型的频率很高，占 97% 左右，而重组型的频率很低，仅占 3% 左右。显然，测得的两个特定基因座的 RF 值越大，例如接近 50% 时，基因间的连锁关系越难以判断，因为与基因的自由组合测交比 1 : 1 : 1 : 1 很接近，在这种情况下，必须用大量的测交后代的数据方可鉴别。如 RF 值越小，基因间的连锁关系越容易察觉，因为测交比例与 1 : 1 : 1 : 1 相差甚远。

四、交换

在同一染色体上相互连锁的基因为什么会出现一定频率的重组？换言之，在两对基因连锁遗传时，形成亲代所没有的新组合的机制是什么？

1. Janssens (1909) 根据两栖类和直翅目昆虫的减数分裂的观察，在摩尔根等人确立遗传的染色体学说之前，就提出了一种交叉型假设 (chiasmatype hypothesis)，其要点是：

(1) 在减数分裂前期，尤其是双线期，配对中的同源染色体不是简单地平行靠拢，而是在非姊妹染色单体间有某些点上显示出交叉缠结的图像，每一点上这样的图像称为一个交叉 (chisma)，这是同源染色体间对应片段发生过交换 (crossing-over) 的地方。

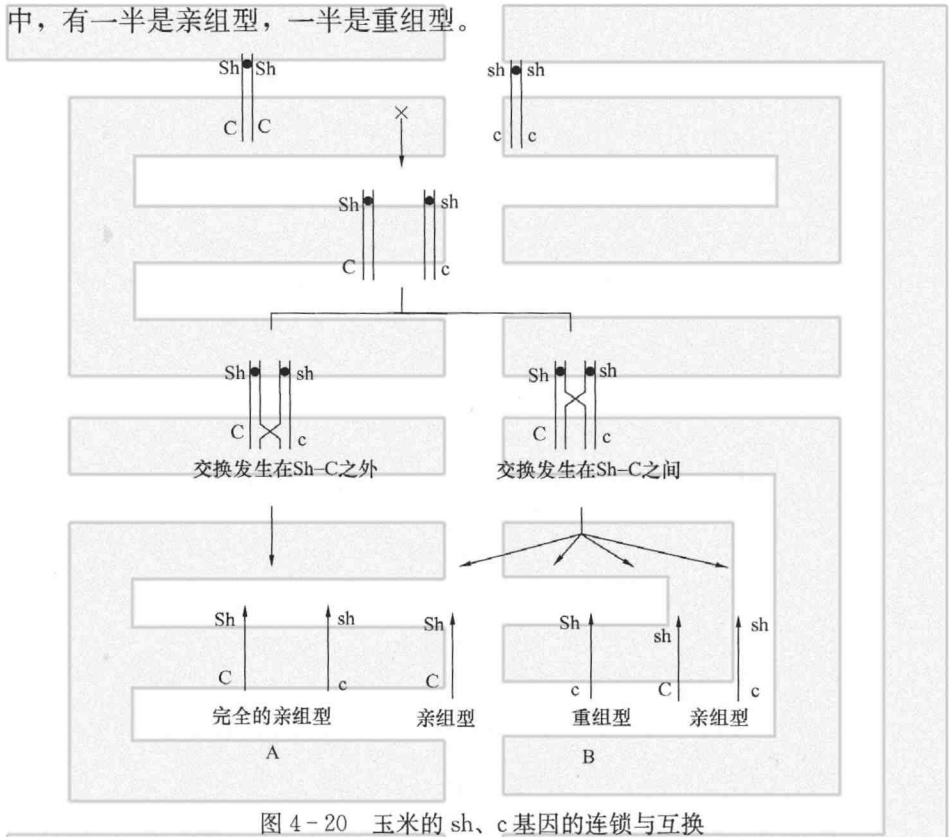
(2) 处于同源染色体的不同座位的相互连锁的两个基因之间如果发生了交换，就导致这两个连锁基因的重组 (recombination)。

这个学说的核心是：交叉是交换的结果，而不是交换的原因，也就是说，遗传学上的交换发生在细胞学上的交叉出现之前。

如果交换发生在两个特定的所研究的基因之间，则出现染色体内重组 (intrachromosomal recombination) 形成交换的产物。否则，交换发生在所研

究的基因之外，则得不到特定基因的染色体内重组的产物。

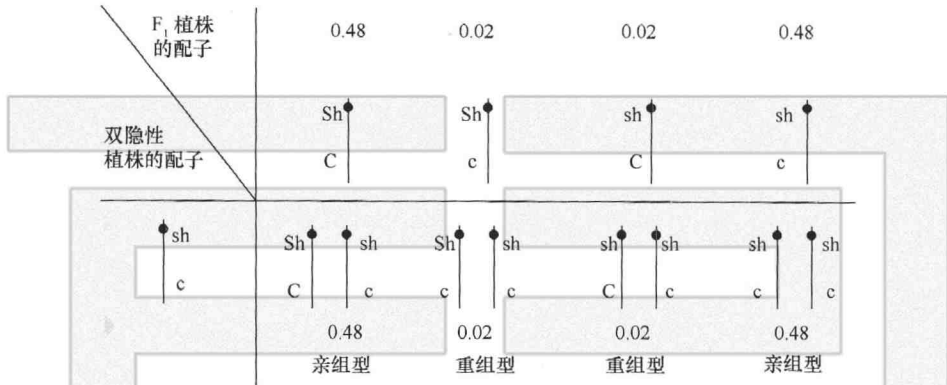
一般情况下，染色体愈长，显微镜下可以观察到的交叉数愈多。一个交叉就代表一次交换。以后的研究证实了交叉型假设。图 4-20 说明重组类型产生的细胞学机制： F_1 植株的性母细胞在减数分裂时，如交叉出现在 sh—c 之间，表示有一半的染色单体在这两对基因间产生过交换，所以形成的配子中，有一半是亲组型，一半是重组型。



如果 F_1 植株的性母细胞在减数分裂时，有 8% 的性母细胞在 sh—c 之间形成一个交叉，表明有一半的染色单体在 sh—c 间发生过交换，所以形成的配子中，有 4% 是亲代没有的重组型。sh—c 间有 4% 的重组，所以我们说，sh 与 c 间的交换值是 4%。

2. 连锁和交换这一对名词在遗传学上被沿用下来，作为与孟德尔定律中基因分离律、自由组合律并立的，由 T. H. Morgan 证明并完善的第三条遗传学上的基本规律。连锁交换定律的基本内容是指：处在同一染色体上的两个

或两个以上基因遗传时，联合在一起的频率大于重新组合的频率。重组类型的产生是由于配子形成过程中，同源染色体的非姊妹染色单体间发生了局部交换的结果。



分离律是自由组合律和连锁定律的基础，而后两者又是生物体遗传的性状发生变异的主流。自由组合与连锁交换二者的差别在于前者的基因是由不同源的染色体所传递，重组类型是由染色体间重组 (interchromosomal recombination) 造成，而后者则是由同一对源染色体所传递的，是染色体内重组所产生的重组类型。另外，自由组合受到生物染色体对数的限制，而连锁交换则受到其染色体本身长度的限制。只要我们在染色体上发现的突变愈多，由交换而产生的重组类型的数量也愈大。从这个意义上说，自由组合是有限的，连锁交换相对的限度较小。

3. 有许多因素影响交换，其中包括温度、性别、射线、化学物质等等。在果蝇中早已发现 22°C 饲养的雌蝇的交换频率最低。某些化学物质如放线菌 D、丝裂霉素 C 可以引起交换值的增加。着丝粒使附近基因的交换值明显降低，人类和许多哺乳动物例如小鼠，雌性交换值大于雄性。性别对于交换的影响最典型的情况是在果蝇中由 Morgan 所发现的雄蝇无交换 (完全连锁)。在家蚕中则恰恰相反，交换只发生在雄蚕而不发生在雌蚕中。实际上，凡是性染色体决定性别的生物，异配性别的个体中一般总是较少发生交换。英国生理学及遗传学家 J. B. S. Haldane 提出：凡是较少发生交换的个体必定是异配性别个体。这被称为霍尔丹定律。交换值虽然受许多因素的影响，但在一定条件下，相同基因之间的交换值是相对稳定的。据研究，凡无交换的类型

都没有发现联会复合体 (SC)。因为 SC 被认为是保证同源染色体准确配对和交换的一个重要结构。

4. 多线交换与最大交换值在同源染色体的两个基因之间如果出现一个交叉, 它只涉及两条非姊妹染色单体, 由此产生的 4 条染色单体有两条是已交换的, 两条是没有交换的, 重组型占 $1/2$ 。如果 100% 的性母细胞在两个基因之间都出现一个交叉, 则重组型频率为 $1/2 \times 100\% = 0.50$ 。这就是两个基因间最大的交换值。当两个基因相距较远时, 其间可以发生两个或两个以上的交叉, 即可能发生过两次或两次以上的交换, 在这种情况下, 不同交叉点上涉及的染色单体将不限于两条, 可以是多线交换。这时, 是否会改变最大交换值?

假定有一个互引相杂交组合 AC/ac, A、C 两个基因座间同时在 I 区和 II 区发生两次交换, I 区发生交换的位置不变, II 区在非姊妹染色单体之间随机地发生另一次交换, 即在 A 和 C 两基因座之间同时发生了两次交换称作双交换 (double crossingover), 就交换所涉及的染色单体来讲, 邻接的两个交换的关系可有 I—II₁ 二线双交换、I—II₂ 三线双交换、I—II₃ 三线双交换和 I—II₄ 四线双交换 (图 4-21)。

图中出现的 4 种双交换所产生的配子, 在染色体结构和遗传组成上各有不同, 但就 A、C 基因座而言, 并没有发生重组。综合全部可能的结果, 可见未发生交换的配子、单交换的配子和双交换的配子的数目之比为 1:2:1。由于 A、C 基因座之间没有遗传标记 [中央的 (B) 及 (b) 基因座暂忽略不计], 双交换与非交换产生的配子, 在表型上无法区别, 只能从单交换结果来计算交换值, 占总数的 $1/2$ 。由此可见, 尽管 A、C 基因之间发生了两次交换, 其最大交换值仍为 0.50。由此可以从理论上计算, 只要 4 条染色单体发生交换的机会相等, 在所研究的两个基因座之间, 产生重组型与亲组型的比总是 1:1, 而且可以推论偶数次交换的结果与非交换相同, 奇数次交换的结果与单交换相同, 因此最大的交换值也是 0.50。

双交换有以下两个特点:

(1) 双交换概率显著低于单交换的概率。如果两次同时发生的交换互不干扰, 各自独立, 则根据概率相乘定律, 双交换发生的概率应是两个单交换概率的乘积。

双交换方式	交换的产物	交换结果	亲组型	重组型
二线 I-II ₁ (2,3)	1 ● A (B) C	非交换	1	0
	2 ● A (b) C	双交换		
	3 ● a (B) c	双交换		
	4 ● a (b) c	非交换		
三线 I-II ₂ (2,3,4)	1 ● A (B) C	非交换	1	1
	2 ● A (b) C	单交换		
	3 ● a (B) c	双交换		
	4 ● a (b) c	单交换		
三线 I-II ₃ (1,2,3)	1 ● A (B) c	单交换	1	1
	2 ● A (b) C	双交换		
	3 ● a (b) C	单交换		
	4 ● a (b) c	非交换		
四线 I-II ₄ (1,2,3,4)	1 ● A (B) c	单交换	0	1
	2 ● A (b) c	单交换		
	3 ● a (B) C	单交换		
	4 ● a (b) C	单交换		
		非交换	4	
		单交换	8	
		双交换	4	
非交换(包括双交换)8:		交换(单交换) 8		

图 4-21 A、C 基因座之间发生各种可能双交换的结果

(2) 3 个连锁基因间发生双交换的结果，旁侧基因无重组。3 个基因中只有处在中央位置的基因改变了位置，末端两个基因 A、C 的相对位置不变。这样，只是 A—B 和 B—C 间发生重组，而 A—C 间不发生重组，但却在其间同时发生了两次交换。因此，在这里 A—C 之间的重组率低于实际的交换值。

第三节 Hardy—Weinberg 定律

一、孟德尔群体

遗传学上的群体不是一般个体的简单集合，而是指相互有交配关系的个

体所构成的有机集合体。一个群体中全部个体所包含的全部基因称为基因库 (gene pool)。在一个大群体内, 个体间随机交配, 孟德尔的遗传因子以各种不同方式从一代传递到下一代, 通常称这种群体为孟德尔群体 (Mendelian population)。最大的孟德尔群体可以是一个物种。存在交配关系意味着群体内个体间有着共同的染色体组, 染色体组内的基因可通过个体间的交配而在个体间得到相互交换。在同一群体内, 不同个体的基因型虽有不同, 而群体的总体所具有的基因则是一定的。群体中各种基因的频率, 以及由不同的交配机制所形成的各种基因型频率在数量上的分布特征称为群体的遗传结构。生物在繁殖过程中, 每个个体传递给子代的并不是其自身的基因型, 而只是不同频率的基因。

孟德尔群体与一般群体的主要区别在于群体内个体间能够随机交配。因此, 几乎所有的动物和异花授粉植物群体都属于孟德尔群体, 而自体受精动物及自花授粉植物构成的群体则只能属于一般的群体或称非孟德尔群体。

由许多群体所构成的生物集团叫群落。群落中的群体间没有交配关系, 群体间的相对独立性主要通过生殖隔离来实现。

二、群体的基因频率和基因型频率

个体性状表现的遗传基础是个体的基因型。而基因型决定于基因与基因的分离与组合, 通过追踪基因在世代间的分离与组合及其所形成的基因型, 可以推断性状表现在群体内个体间和家系水平的遗传与变异规律。群体性状表现在个体间的遗传与变异规律决定于群体的基因频率和基因型频率。

基因频率 (Alleles frequency) 是指某位点的某特定基因在其群体内占该位点基因总数的比率。例如, 某群体内某一基因位点 A 基因与 a 基因在总个体数中共计 10,000 个, 其中 A 基因 9,975 个, a 基因 25 个, 则 A 基因的频率为 0.9975, a 基因的频率为 0.0025。

基因型频率 (Genotype frequency) 是指群体内某特定基因型个体占个体总数的比率。例如, 一个群体中纯合显性基因型 AA 个体 80 个, 杂合基因型 Aa 个体 14 个, 纯合隐性基因型 aa 6 个。则 AA 、 Aa 和 aa 3 种基因型频率分别为 0.80、0.14 和 0.06。

基因频率和基因型频率一般都无法直接计算。但是表现型是由基因型决

定的，表现型是可以直接度量 and 计算的，因此可以通过表现型频率求得基因型频率，进而推知基因频率。

设由 N 个个体构成的某二倍体生物群体中，有一对等位基因 A 与 a ，其可能的基因型为 AA 、 Aa 、 aa 共 3 种，对应的个体数分别为 N_D 、 N_H 和 N_R ，相应的基因型频率为 D 、 H 、 R ，则 3 种基因型的频率各为：

$$AA:D = \frac{N_D}{N} \quad Aa:H = \frac{N_H}{N} \quad aa:R = \frac{N_R}{N}$$

显然， $N_D + N_H + N_R = N$ ， $D + H + R = 1$ 。由于每个个体含有一对等位基因，群体的总基因数为 $2N$ 。由此根据基因频率的定义可知基因 A 的频率为：

$$p = \frac{2N_D + N_H}{2N} = D + \frac{1}{2}H$$

同样，等位基因 a 的频率为

$$q = \frac{2N_R + N_H}{2N} = R + \frac{1}{2}H$$

并且， $p + q = 1$ 。基因频率与基因型频率变动在 $0 \sim 1$ 之间。

在人类群体中，常染色体上有一对等位基因 T 与 t ，决定对苯硫脲 (PTC) 的尝味能力。 T 与 t 构成的 3 种基因型与其决定的表现型有对应关系，即 TT 表现为尝味者， Tt 为味觉杂合体， tt 为味盲。因此，可以根据表现型判断基因型，并进而计算出基因频率。例如，某次抽样调查了中国汉族人群的 1,000 人，3 种基因型的分布如表 4-9。

表 4-9 中国汉族人群中 PTC 尝味能力的分布

表现型	基因型	人数	基因型频率
尝味者	TT	(N_D) 490	(D) 0.49
味觉杂合体	Tt	(N_H) 420	(H) 0.42
味盲	tt	(N_R) 90	(R) 0.09
总计		1000	1

用上面两个公式可计算出 T 与 t 的基因频率：

$$p(T) = D + \frac{1}{2}H = 0.49 + \frac{1}{2} \times 0.42 = 0.7$$

$$q(t) = R + \frac{1}{2}H = 0.09 + \frac{1}{2} \times 0.42 = 0.3$$

并且有 $p + q = 0.7 + 0.3 = 1$ 。

三、Hardy—Weinberg 定律及其应用

1. Hardy-Weinberg 定律

基因频率和基因型频率是群体遗传组成的基本特征。在一定条件下，基因频率和基因型频率在世代间可以保持不变，当各基因频率和基因型频率在上下代间保持不变或相对稳定时，群体的性状表现就会保持遗传上的稳定，这是群体遗传的重要机制和现象之一。1908年英国的数学家 G. H. Hardy 和德国医生 W. Weinberg 分别独立地发现了这种遗传机制，因此称为 Hardy-Weinberg 定律，又称为群体的遗传平衡定律。

该定律的要点是：在一个大的随机交配的群体中，如果没有改变基因频率因素（如选择、突变和迁移）的干扰，群体的基因频率和基因型频率将保持不变，这样的群体称为平衡的孟德尔群体；在任何一个大群体内，不论基因频率和基因型频率如何，只要经过一代随机交配，这个群体就可以达到平衡状态。群体处于平衡状态时，基因型频率和基因频率的关系是 $D = p^2$ ， $H = 2pq$ ， $R = q^2$ 。

设一对等位基因 A 与 a ，频率分别为 p 和 q ，构成基因型为 AA 、 Aa 、 aa 的群体，其相应基因型频率分别为 p^2 ， $2pq$ 和 q^2 ，由该群体的个体间随机交配产生子代。则随机交配前，群体产生 A 和 a 两种配子，其频率为 p 和 q ，随机交配时，两种雌配子 $(p+q)$ 与两种雄配子 $(p+q)$ 随机结合，即 $(p+q)^2$ ，雌配子 A 与雄配子 A 结合产生 AA 合子，其频率为 $p \times p = p^2$ ，雌配子 a 与雄配子 a 结合产生 aa 合子，其频率为 q^2 ，而雌配子 A 与雄配子 a 结合即 $p \times q$ 和雌配子 a 与雄配子 A 结合即 $q \times p$ 都形成相同杂合子 Aa ，因此杂合体 Aa 的频率为 $2pq$ 。即子代的各基因型及其频率是：

AA	Aa	aa
p^2	$2pq$	q^2

子代各基因型频率与亲代完全相同。这一结果也可通过亲代各基因型频率直接推导求得，亲代 3 种基因型 AA 、 Aa 和 aa ，频率分别为 p^2 、 $2pq$ 和 q^2 。相互随机交配即 $(p^2 + 2pq + q^2)^2$ ，将有 9 种交配方式共 6 种交配类型。各种交配类型的频率及所产生子代基因型的频率见表 4-10。

根据子代的基因型频率，依照最初的两个公式，可得子代的 A 基因频率为：

$$p_1 = p^2 + \frac{1}{2} \cdot 2pq = p(p+q) = p$$

a 基因的频率为：

$$q_1 = q^2 + \frac{1}{2} \cdot 2pq = q(p+q) = q$$

子代的基因频率与亲代也完全相同。

在此基础上可进一步求证群体实现遗传平衡需要的世代。设在一个群体中，一对等位基因 A 与 a 构成的三种基因型及其初始频率为：

$$\begin{array}{ccc} AA & Aa & aa \\ D_0 & H_0 & R_0 \end{array}$$

由基因型频率可得 A 与 a 的基因初始频率：

$$p_0 = D_0 + \frac{1}{2}H_0 \quad q_0 = R_0 + \frac{1}{2}H_0$$

表 4-10 单个位点随机交配群体的基因型频率

亲代交配类型	交配频率	子代基因型与频率		
		AA	Aa	aa
AA × AA	p^4	p^4	—	—
AA × Aa	$4p^3q$	$2p^2q$	$2p^3q$	—
AA × aa	$2p^2q^2$	—	$2p^2q^2$	—
Aa × Aa	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
Aa × aa	$4pq^3$	—	$2p^2q^3$	$2pq^2$
aa × aa	q^4	—	—	q^4
总计	1	p^2	$2pq$	q^2

经随机交配，下一代群体基因型频率与基因频率为：

$$D_1 = p_0^2 \quad H_1 = 2p_0q_0 \quad R_1 = q_0^2$$

$$p_1 = D_1 + \frac{1}{2}H_1 = p_0^2 + \frac{1}{2} \times 2p_0q_0 = p_0(p_0 + q_0) = p_0$$

$$q_1 = R_1 + \frac{1}{2}H_1 = q_0^2 + \frac{1}{2} \times 2p_0q_0 = q_0(p_0 + q_0) = q_0$$

再经一代随机交配，群体的基因型与基因频率为：

$$D_2 = p_1^2 = p_0^2 \quad H_2 = 2p_1q_1 = 2p_0q_0 \quad R_2 = q_1^2 = q_0^2$$

$$p_2 = D_2 + \frac{1}{2}H_2 = p_0^2 + \frac{1}{2} \times 2p_0q_0 = p_0$$

$$q_2 = R_2 + \frac{1}{2}H_2 = q_0^2 + \frac{1}{2} \times 2p_0q_0 = q_0$$

如此继续随机交配, 不难推出, $D_2 = D_3 = D_4 \cdots \cdots = D_n = p_0^2$, $H_2 = H_3 = H_4 \cdots \cdots = H_n = 2p_0q_0$, $R_2 = R_3 = R_4 \cdots \cdots = R_n = q_0^2$, $p_2 = p_3 = p_4 \cdots \cdots = p_n = p_0$, $q_2 = q_3 = q_4 \cdots \cdots = q_n = q_0$ 。

从上述推导过程可以看出, D_0 与 D_1 , H_0 与 H_1 , R_0 与 R_1 可能不等, 但 $D_1 = D_2 = D_3 \cdots \cdots = D_n$, $H_1 = H_2 = H_3 \cdots \cdots = H_n$, $R_1 = R_2 = R_3 \cdots \cdots = R_n$, $p_1 = p_2 = p_3 = \cdots \cdots = p_n$ 。所以只要经过一代随机交配, 群体就达到遗传平衡状态, 从而使各基因型频率和基因频率在上下代间保持不变。

举例: 假设在一群体中, 一对等位基因 A 与 a 构成的三种基因型能够从表现型上加以区分。亲代的基因型频率分别为:

AA	Aa	aa
D_0	H_0	R_0
0.68	0.04	0.28

根据最上面两个公式可以计算出 p_0 和 q_0 :

$$p_0 = D_0 + \frac{1}{2}H_0 = 0.68 + \frac{1}{2} \times 0.04 = 0.7$$

$$q_0 = R_0 + \frac{1}{2}H_0 = 0.28 + \frac{1}{2} \times 0.04 = 0.3$$

经一代随机交配, 子代群体的基因型频率与基因频率:

$$D_1 = p_0^2 = 0.7^2 = 0.49$$

$$H_1 = 2p_0q_0 = 2 \times 0.7 \times 0.3 = 0.42$$

$$R_1 = q_0^2 = 0.3^2 = 0.09$$

$$p_1 = D_1 + \frac{1}{2}H_1 = 0.49 + \frac{1}{2} \times 0.42 = 0.7$$

$$q_1 = R_1 + \frac{1}{2}H_1 = 0.09 + \frac{1}{2} \times 0.42 = 0.3$$

再随机交配得子二代, 以此类推, 在 A 、 a 两种配子的频率分别保持 0.7 和 0.3 时, 子二代的三种基因型频率与子一代完全相同:

AA	Aa	aa
p_0^2	$2p_0q_0$	q_0^2
0.49	0.42	0.09

因此可以得出结论，随机交配一代的群体已经达到平衡。

实际上，自然界许多群体都属于大群体，许多性状特别是那些中性性状在个体间的交配一般是接近随机的，所以 Hardy-Weinberg 定律有普遍的适用性。

处于 Hardy-Weinberg 平衡状态的群体还具有以下重要特征：

(1) 在二倍体群体中，基因型频率与基因频率之间及基因型频率与基因型频率之间有特定的数学关系

$$D = p^2, p = \sqrt{D}$$

$$R = q^2, q = \sqrt{R}$$

$H = 2pq, H = 2\sqrt{DR}$ ，于是：

$$\frac{H}{\sqrt{DR}} = 2$$

上式反映了群体的三种基因型间的关系，与群体的基因频率无关，该式可作为检验群体是否达到平衡状态的一个尺度。上式还可写成：

$$4DR = H^2$$

根据基因频率与基因型频率间的数学关系，可以用一个更简便的方法计算平衡群体的基因频率。由上面公式计算基因频率时至少要知道群体中某位点的杂合体及一种纯合体两种基因型的频率。这两种基因型有时不能从表现型上加以区分（如完全显性时的显性纯合体与杂合体表现相同），而隐性纯合体的基因型与表现型总是一致的。因此，只要计算出隐性个体比率就可得到相应隐性基因型的频率。

$$\text{隐性基因型频率 (R)} = \frac{\text{隐性个体数}}{\text{供试群体所含个体总数}}$$

隐性基因频率为构成它的隐性基因型频率的开方，所以计算隐性基因频率的公式可以写成：

$$q = \sqrt{R} = \sqrt{\text{隐性基因型频率}}$$

$$q = \sqrt{\frac{\text{隐性个体数}}{\text{供试群体所含个体总数}}}$$

因为 $p+q=1$ ，所以显性基因频率为：

$$p = 1 - q$$

举例：设一群体，经多代随机交配开白花的 25 株，开红花的 9,975 株。已知红花对白花为显性，计算该群体的基因频率与基因型频率。

$$\text{解： } q = \sqrt{\frac{25}{9975 + 25}} = 0.05$$

$$p = 1 - 0.05 = 0.95$$

$$AA \text{ 频率} = p^2 = 0.95^2 = 0.9025$$

$$Aa \text{ 频率} = 2pq = 2 \times 0.95 \times 0.05 = 0.095$$

$$aa \text{ 频率} = q^2 = 0.05^2 = 0.0025$$

(2) 在平衡的二倍体群体中，杂合体频率 H 的最大值为 0.5，只有 $p=q=0.5$ 时， H 才取得最大值，下面进行证明：

由于 $H = 2pq = 2(1-q)q = 2q - 2q^2$ ，对函数 $H = 2q - 2q^2$ 求导得：

$$\frac{dH}{dq} = d(2q - 2q^2) = 2 - 4q$$

令导数等于零，此时 q 的取值得 H 的最大值：

$$2 - 4q = 0$$

$q = \frac{1}{2}$ ，于是

$$p = 1 - q = 1 - \frac{1}{2} = \frac{1}{2}$$

$$h = 2pq = 2 \times \frac{1}{2} \times \frac{1}{2} = 0.5$$

2. 群体平衡的检验

前已提及，判断一个群体是否实现了 Hardy-Weinberg 平衡，可以直接将 3 种基因型值代入 $\frac{H}{\sqrt{DK}} = 2$ ，如果等式成立，则群体是平衡的，否则是不平衡的。然而实际对群体基因型频率的调查结果并不能使上述等式两边完全相等，在这种情况下判断群体是否平衡，需要将基因型分布的观察频率与按 Hardy-Weinberg 平衡定律预计的频率作 χ^2 检验。

举例：在我国某省汉族人群中检测了 1,050 人对苯硫脲的尝味能力，其中尝味者 (TT) 410 人，味觉杂合体 (Tt) 500 人，味盲 (tt) 140 人。判断该抽样群体中的基因型频率是否符合 Hardy-Weinberg 平衡。

先从实际观察值中计算出基因频率：

$$T \text{ 的基因频率: } p = \frac{410}{1050} + \frac{1}{2} \times \frac{500}{1050} = 0.63$$

$$t \text{ 的基因频率: } q = \frac{140}{1050} + \frac{1}{2} \times \frac{500}{1050} = 0.37$$

再将计算得到的基因频率代入表 4-11 计算 x^2 值。

表 4-11 PTC 尝味能力的 x^2 检验

	TT	Tt	tt	总计
观察值 (o)	410	500	140	1050
期望值 (e)	416.7	489.5	143.7	1050
	$p^2 N$	$2pqN$	$q^2 N$	N
$\frac{(o-e)^2}{e}$	0.108	0.225	0.095	0.428

$x^2 = 0.428 < x^2_{(df=1, p=0.05)} = 3.841$ ，表明该样本群体 PTC 尝味的 3 种基因型频率符合 Hardy-Weinberg 平衡。也就是说样本群体的 PTC 基因处于遗传平衡状态。应该指出，在此 x^2 的自由度是 $3-1-1=1$ ，而不是 $3-1=2$ ，因为在计算期望值时要使用基因频率 p 的数值，而它是从样本中估计出来的，因此 x^2 的自由度又少了一个。

3. 群体平衡定律的推广

以上讨论了二倍体群体中一对等位基因的平衡情况。如果存在复等位基因，则在同一基因位点可能有 2 个以上的等位基因。我们先考虑 3 个等位基因 A, a, a' 在群体中的遗传，设它们的频率分别为 p, q, r, 且 $p+q+r=1$ 。在一个大的随机交配群体中，3 个基因的频率与 6 种基因型的频率如果有下列关系，则认为平衡已经建立：

$$\begin{pmatrix} A & a & a' \\ p & q & r \end{pmatrix} = \begin{matrix} AA & Aa & Aa' & aa & aa' & a'a' \\ p^2 & 2pq & 2pr & q^2 & 2qr & r^2 \end{matrix}$$

平衡状态下的基因频率可以由基因型频率按下列各式求得：

$$p = p^2 + \frac{1}{2}(2pq + 2pr) = p^2 + pq + pr$$

$$q = q^2 + \frac{1}{2}(2pq + 2pr) = q^2 + pq + qr$$

$$r = r^2 + \frac{1}{2}(2pr + 2qr) = r^2 + pr + qr$$

即：某基因的频率是其纯合体的频率与含有该基因全部杂合体频率一半之和。

若有 k 个复等位基因 $A_1, A_2, A_3 \dots A_k$ ，相应的基因频率是 $p_1, p_2, p_3 \dots p_k$ ，且 $\sum p_i = 1$ ，不考虑基因的显隐性关系，基因频率与基因型频率的关系可表示为：
$$\left(\sum_{i=1}^k p_i A_i\right)^2 = \sum_i p_i^2 A_i A_i + 2 \sum_{i < j} p_i p_j A_i A_j$$

亦即上面公式的一般形式。等式右边是 $(p_1 A_1 + p_2 A_2 + p_3 A_3 + \dots + p_n A_n)^2$ 的多项式展开式，其中 $A_i A_i$ 表示纯合体，共有 k 种， $A_i A_j$ 为杂合体，有 $k(k-1)/2$ 种。

在群体的复等位基因遗传中，尽管某基因座上可能有多种基因，但就某一个二倍体个体而言，其同源染色体的相应基因位点上只有 k 个复等位基因中的任何两个等位基因。因此，如果群体最初没有处于平衡状态，只要经过一个世代的随机交配，就可实现基因型频率的平衡。

第四节 突变、自然选择、迁移和遗传漂变 对基因频率的影响

一、基因突变

基因突变对改变群体遗传组成的作用有两个方面。一是它提供遗传变异的最原始材料，二是突变本身改变基因频率。

设一对等位基因 A 与 a ，其基因频率为 p 和 q ，当 A 基因突变为 a 基因时，群体中 A 的频率就会减少， a 的频率则增加；反过来当 a 基因突变为 A 基因时，群体中 A 的频率会增加， a 的频率则减少。当 $A \rightarrow a$ 的突变率为 u ， $a \rightarrow A$ 的突变率为 v 时，每代有 $(1-u)u$ 的 A 基因突变为 a ， qv 的 a 基因突

变为 A , 若 $(1-q)u > qv$, 则 a 基因频率增加, 若 $(1-q)u < qv$, 则 A 基因的频率增加, 于是每代 a 基因频率的净改变量为:

$$\Delta q = (1-q)u - qv$$

经过足够多的世代, 这两种相反的力量相互抵消, 也就是正突变压和负突变压相等时, $\Delta q = 0$, 基因频率保持不变, 群体处于平衡状态, 则:

$$(1-q)u = qv, \text{ 或 } pu = qv, \text{ 于是}$$

$$q = \frac{u}{u+v}$$

$$p = \frac{v}{u+v}$$

上两式给出了在正反突变压的作用下, 群体达到平衡状态时的基因频率。假定 $u = 1.5 \times 10^{-5}$, $v = 1 \times 10^{-5}$, 平衡时:

$$q = \frac{1.5 \times 10^{-5}}{1.5 \times 10^{-5} + 1 \times 10^{-5}} = \frac{3}{5} = 0.6$$

容易看出, 若 $u = v$, 达到平衡状态群体的 p 和 q 值都等于 0.5。

基因突变对群体遗传组成的作用还可以由经过一定世代基因频率的改变情况来了解。设显性基因频率在某一世代是 p_0 , 群体中只发生 $A \rightarrow a$ 的突变, 经过 n 个世代群体中显性基因 A 的频率为:

$$p_n = p_0(1-u)^n$$

若突变频率 u 很低, 尽管 n 可能很大, 但 $(1-u)^n$ 的值仍然会很接近于 1, 这样, p_n 与 p_0 之间的差值也将很小。如果知道突变前后显性基因频率的变化, 可以计算出突变经历的代数, 因为:

$$p_n = p_0(1-u)^n$$

$$(1-u)^n = \frac{p_n}{p_0}$$

$$n \lg(1-u) = \lg \frac{p_n}{p_0}$$

$$n = \frac{\lg p_n - \lg p_0}{\lg(1-u)}$$

还可表示为:

$$n = -\frac{1}{u} \ln \frac{p_n}{p_0}$$

也就是说,当 $A \rightarrow a$, 并且不计反突变时,使显性基因频率降低到某一数值所经历的代数与突变后和突变前基因频率之比的对数成正比,与其突变频率成反比。在自然条件下,突变速率很小,一般都在 $10^{-4} \sim 10^{-7}$ 。因此,要想明显改变群体的基因频率,一定要经过许多世代。例如, $u = 1 \times 10^{-5}$, p 由 0.6 降到 0.5,需要近 2 万代。但某些生物(如微生物中的细菌)的世代很短,突变就可能成为改变群体基因频率的重要因素。

二、选择

选择是改变基因频率的最重要因素,也是生物进化的驱动力量。就显隐性性状而言,选择通常分为两种:一种是淘汰显性个体,使隐性基因增加的选择;另一种是淘汰隐性个体使显性基因增加的选择。前者能迅速改变群体的基因频率。例如,在一个包含开红花和开白花植株的群体中选留白花,只需经过一代就能把红花植株从群体中消灭,从而把红花基因的频率降低到 0,白花基因的频率增加到 1。

淘汰隐性个体保留显性个体的选择情况比较复杂。因为选留的显性个体可能包含两种基因型,其中有一种是杂合体。杂合体内的一半隐性基因不能被淘汰而同显性基因在杂合体内保留下来。所以,这种选择方式只能使隐性基因频率逐渐变小,但不会降到 0,显性基因频率则逐渐增加,也不会达到 1。

设红花显性基因 A 选择前的频率为 p_0 , 隐性基因 a 的频率为 q_0 , 选择前 AA , Aa , aa 3 种基因型的频率分别为 $D_0 = p^2$, $H_0 = 2p_0q_0$, $R_0 = q_0^2$ 。经过淘汰隐性个体后,群体中只留下 AA 和 Aa , 在此基础上随机交配,繁殖产生下一代群体。现在我们只考虑隐性基因的变化情况。由于 aa 已被淘汰,下一代隐性基因频率 q_1 只有从杂合子的一半占整个群体的比例中求出。

$$q_1 = \frac{\frac{H_0}{2}}{D_0 + H_0} = \frac{\frac{1}{2} \cdot 2p_0q_0}{p_0^2 + 2p_0q_0}$$

将 $p_0 = 1 - q_0$ 代入上式得:

$$q_1 = \frac{q_0}{1 + q_0}$$

下一代再淘汰白花植株并随机交配繁殖,则同理:

$$q_2 = \frac{q_1}{1+q_1} = \frac{\frac{q_0}{1+q_0}}{1+\frac{q_0}{1+q_0}} = \frac{q_0}{1+2q_0}$$

$$q_3 = \frac{q_2}{1+q_2} = \frac{\frac{q_0}{1+2q_0}}{1+\frac{q_0}{1+2q_0}} = \frac{q_0}{1+3q_0}$$

经过 n 代淘汰后，隐性基因频率为：

$$q_n = \frac{q_0}{1+nq_0}$$

q_0 为选择前隐性基因 a 的频率，它等于：

$$q_0 = \sqrt{\frac{\text{选择前隐性个体数}}{\text{选择前群体的总个体数}}}$$

可以推算出，由一群体出发，达到某一基因频率需要的世代数，推导过程如下：

$$q_n = \frac{q_0}{1+nq_0}$$

$$1+nq_0 = \frac{q_0}{q_n}$$

$$nq_0 = \frac{q_0}{q_n} - 1$$

$$n = \frac{1}{q_0} \left(\frac{q_0}{q_n} - 1 \right)$$

$$n = \frac{1}{q_n} - \frac{1}{q_0}$$

上式中 $q_n = \sqrt{\frac{\text{选择 } n \text{ 代后群体隐性个体数}}{\text{选择 } n \text{ 代后群体的总个体数}}}$

举例：一个开白花和开红花的随机交配群体，在选择前群体中白花个体 625 株，红花个体 9,375 株。经过若干代淘汰白花个体之后，群体内开白花植株只有 25 株，开红花的 9,975 株，试分析经历的选择代数。

因为：

$$q_n = \sqrt{\frac{25}{9975+25}} = 0.05 \quad q_0 = \sqrt{\frac{625}{9375+625}} = 0.25$$

所以：

$$n = \frac{1}{q_n} - \frac{1}{q_0} = \frac{1}{0.05} - \frac{1}{0.25} = 16 \text{ (代)}$$

当 $q_n = 1/2q_0$ 时, $n = 1/q_0$, 这意味着隐性基因频率减至初始频率一半时所需的世代数是初始频率的倒数。

自然界中的选择并不总是像上面介绍的情形, 完全淘汰某一类基因型, 而是表现为一些个体的生活率、繁殖率低于正常个体, 从而降低这些个体所携带基因的频率。为了说明这种选择对群体基因频率的作用, 需要引入适应值 (Fitness) 和选择系数 (Selection coefficient) 的概念。

适应值也叫适合度 (f), 是表示某种基因型个体在一定环境下相对繁殖率的度量。将具有最高繁殖率基因型的适应值定为 1, 以其他基因型与之相比较的相对值作为他们的适应值。隐性致死基因的纯合体在成熟前死亡, 不可能留下后代, 因此对后代群体没有遗传贡献, 其适应值为 0。一个群体的适应值等于群体内全部个体适应值的平均值。表 4-12 说明了计算不同基因型适合度的方法。

表 4-12 适应值的计算

	基 因 型			总 计
	AA	Aa	aa	
当代个体数 (a)	40	50	10	100
下一代个体数 (b)	80	90	10	180
繁殖率 (b/a)	$80/40=2$	$90/50=1.8$	$10/10=1$	
适应值 (f)	$2/2=1$	$1.8/2=0.9$	$1/2=0.5$	

表 4-12 中具有最高繁殖率的基因型是 AA, 将 AA 的适应值定为 1, 其他基因型的 f 值只需以其繁殖率除以 AA 的繁殖率即可求得。

在人类群体中, 曾调查出软骨发育不全的侏儒症患者 108 人, 他们共生育 27 个子女。以他们的 457 个正常同胞共生育了 582 个子女, 生育率为对照, 可计算出侏儒症的相对生育率或适应值为:

$$f = \frac{27/108}{582/457} = 0.2$$

侏儒症的适应值为 0.2, 意味着若每个正常人平均能留下 1 个子女, 侏儒

症患者只能平均留下 0.2 个后代。

选择系数 s ，是表示某一基因型在群体中不利于生存程度的度量， $s=1-f$ ，显然适合度 f 可用 $1-s$ 来表示。在表 4-12 中，基因型 AA 的 $s=1-1=0$ ，Aa 的 $s=1-0.9=0.1$ ，aa 的 $s=1-0.5=0.5$ ，隐性致死基因纯合体的 $f=0$ ，其 $s=1$ 。基因型的 $s=0$ 意味着无选择。 $s=1$ 则表示全部淘汰该种基因

型个体，其效应已在公式 $q_n = \frac{q_0}{1+nq_0}$ 中给出。下面讨论 $0 < s < 1$ 的情况。

设有一对基因 A, a, 且 A 对 a 完全显性，群体中 AA, Aa, aa 3 种基因型为 $p^2 + 2pq + q^2 = 1$ 。又设 AA 与 Aa 的适应值 $f=1$ ，隐性纯合体 aa 的适应值为 $1-s$ 。经过一代选择后基因型的频率如表 4-13。

表 4-13 基因型频率经过一代选择的变化情况

基因型	AA	Aa	aa	总计
选择前频率	p^2	$2pq$	q^2	1
适应值	1	1	$1-s$	
选择后频率	p^2	$2pq$	$q^2(1-s)$	$1-sq^2$
选择后相对基因型频率	$\frac{p^2}{1-sq^2}$	$\frac{2pq}{1-sq^2}$	$\frac{q^2(1-s)}{1-sq^2}$	1

由表 4-13 可求得经过一代选择后的基因频率 q_1 及其基因频率的改变 Δq ：

$$q_1 = \frac{1}{2} \cdot \frac{2pq}{1-sq^2} + \frac{q^2(1-s)}{1-sq^2} = \frac{q(1-sq)}{1-sq^2}$$

$$\Delta q = q_1 - q = \frac{q(1-sq)}{1-sq^2} - q = \frac{-sq^2(1-q)}{1-sq^2}$$

从上式可以看出，选择对改变基因频率的作用不仅与选择系数 s 有关，还随初始基因频率的大小而不同。如果选择系数不变，基因的初始频率越接近 0.5，选择造成的基因频率改变量越大，而当 q 很小时，分母 $1-sq^2$ 可视为 1，这样选择引起的 q 的改变可以近似地表示为：

$$\Delta q = -sq^2(1-q)$$

可见 q 值很小时， Δq 很小，因此难以将隐性基因淘汰干净。

三、遗传漂变

在一个小群体内由于抽样误差造成的群体基因频率的随机波动现象称作随机遗传漂移 (Random genetic drift), 也叫作遗传漂变 (Genetic drift)。遗传漂变也是影响群体平衡的重要因素, 但与其他影响群体平衡因素如突变、选择和迁移相比, 其不同之处在于它改变群体基因频率的作用方向是完全随机的。

在小群体内所包含的各种基因型个体的频率不会刚好和原来的群体一样, 因此基因频率势必要改变, 现举例分析如下:

设: 每代只由 4 个雌配子和 4 个雄配子随机结合, 产生 4 个个体组成的小群体, 其中等位基因 A, a 的频率 $p=q=0.5$, 则形成下一代的雌、雄各 4 个配子中所含 A 基因数有 0~4 共 5 个等级, 其频率由二项式 $(1/2+1/2)^4$ 展开式的各项给出:

A 基因数目	0	1	2	3	4
频率	1/16	4/16	6/16	4/16	1/16

在雌、雄各 4 个配子随机结合而形成的子代 4 个个体包含的 8 个基因中, A 基因的数目有可能是由 0~8 个共 9 种情况。4 个个体含有 A 基因数的 9 种情况各自所占频率为 $(1/2+1/2)^4$ 展开式的各项, 各项的频率如表 4-14。

表 4-14 9 种情况下 A 基因频率与发生概率

A 基因数目	A 基因频率	发生概率
0	0.000	1/256=0.004
1	0.125	8/256=0.031
2	0.250	28/256=0.109
3	0.375	56/256=0.219
4	0.500	70/256=0.273
5	0.625	56/256=0.219
6	0.750	28/256=0.109
7	0.875	8/256=0.031
8	1.00	1/256=0.004

可见由这 4 个个体构成的群体中, A 基因频率与亲代群体相同 ($p=0.5$)

的概率仅为 27.3%，与亲代不同的概率为 72.7%。在各个不同于亲代基因频率的数值中，A 基因频率或增或减，增减的多少都是完全随机的。其中还有 A 基因消失 $p=0$ 的可能性，概率为 0.4%，A 基因固定的可能性也是 0.4%。由于随机漂移的原因一旦使某基因在一小群体中消失，除非发生相同突变，否则该基因就会永远消失；相反，由于同样的原因也可使群体中某基因被完全固定。

遗传漂变的作用大小因样本群体的个体数不同而异。样本越小，基因频率的随机波动越大；样本越大，基因频率改变的幅度越小。用基因频率的标准差与样本大小的关系可以定量描述遗传漂变作用。如果在等位基因 A 与 a 的频率分别为 p 和 q 的群体中每次取 N 个个体作为繁殖下一代的样本亲本，

则样本基因频率的标准差为：
$$\sigma = \sqrt{\frac{pq}{2N}}$$

式中： $2N$ 为样本群体中的基因数。

如果设 $p=q=0.5$ ，则：

$$N = 5 \text{ 时, } \sigma = \sqrt{\frac{0.5 \times 0.5}{2 \times 5}} = 0.158$$

$$N = 100 \text{ 时, } \sigma = \sqrt{\frac{0.5 \times 0.5}{2 \times 100}} = 0.035$$

上述计算结果表明如果样本仅包含 5 个个体，下一代群体基因频率变化幅度为 $0.5+0.158 \sim 0.5-0.158$ ，若样本容量增加到 100，则基因频率的变幅为 $0.5+0.035 \sim 0.5-0.035$ 。可见样本越小，抽样误差越大。

利用计算机模拟实验可证明随机遗传漂移效应在不同规模群体中的大小。图 4-22 显示了群体大小与遗传漂变的关系，三种群体的个体数分别是 50、500 和 5000，初始等位基因频率约为 0.5。漂变使随机交配小群体（50 个个体）的等位基因在 30~60 代就被固定。群体增大至 500 个个体以后，经过 100 代的随机交配等位基因频率已经逐渐偏离 0.5。个体数达 5000 的群体至 100 代时，等位基因频率仍然接近初始值 0.5。

遗传漂移在小群体中的作用很强，它可以掩盖甚至违背选择所起的作用。无适应意义的中性突变基因，或选择与之不利但尚未达到携带者致死程度的基因，都有机会因漂移作用而被固定。遗传漂移作用可以解释在远离大陆的

孤岛上常出现大陆上所没有的物种的现象。这是因为，在被隔离的小群体里，一旦发生了某个位点的突变，经过若干代小群体繁殖，新产生的基因可以由漂移作用使其频率增加以致被固定下来，逐渐形成一个新类型或物种，而在随机交配的大群体里，尽管这种突变同样可以发生，但由于大群体随机交配产生的遗传平衡作用使新生成基因的频率没有机会被固定，也就难有机会被发现。

需要指出，遗传漂变来自小样本留种的随机抽样误差，而不是群体内个体间的非随机交配作用。

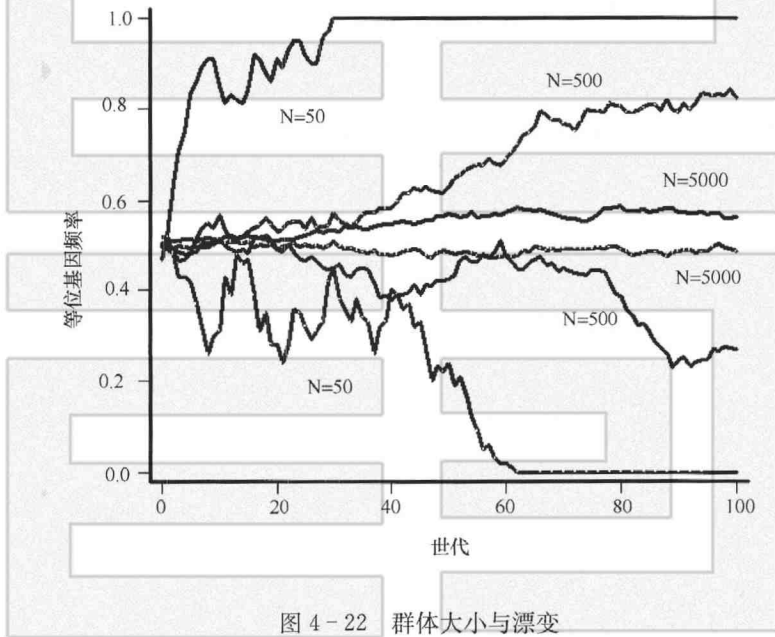


图 4-22 群体大小与漂变

四、迁移

个体在群体间的迁移 (migration) 同样也是影响群体等位基因频率改变的一个因素。设在一个大的群体内，每代有一部分个体新迁入，其迁入率为 m ，则 $1-m$ 是原有个体的比率。令迁入个体某一等位基因的频率是 q_m ，原来个体所具同一等位基因的频率是 q_0 ，二者混杂后群体内等位基因的频率 q_1 将是：

$$q_1 = mq_m + (1 - m)q_0 = m(q_m - q_0) + q_0$$

一代迁入所引起的等位基因频率的变化 Δq 则为：

$$\Delta q = q_1 - q_0 = m(q_m - q_0)$$

可见，在有迁入个体的群体里等位基因频率的变化率等于迁入率同迁入个体等位基因频率与本群体等位基因频率差异的乘积。

第五节 近亲繁殖和杂种优势

大多数动植物的繁殖方式是有性繁殖，由于亲本来源和交配方式不同，它们的后代遗传动态有着明显的差异。早在 19 世纪 60 年代，达尔文就提出了“异花授粉一般对后代有利，而自花授粉对后代是有害”的结论。孟德尔遗传规律发现以后，近亲繁殖和杂种优势一直是遗传学研究的一个重要方面。

一、近亲繁殖及其遗传效应

1. 近交的概念

有性生殖是动、植物繁殖的普遍形式，由于动、植物的交配方式不同，群体结构的相应变化也不相同，性状的遗传效应也有明显差异。深入了解在不同交配方式下动、植物的遗传规律，可更有效地开展动、植物育种工作。

根据亲缘关系的远近，可把一些交配方式图示为图 4-23。

由图 4-23 所示，以品种内交配为起点，愈上则亲缘关系愈近，属于近亲繁殖 (inbreeding)，也称近亲交配或简称近交，是指血统或亲缘关系相近的两个个体间的交配，其极端类型的为自交；愈下则亲缘关系愈远，属于异交，而以远缘杂交为极点。

植物群体或个体近亲交配的程度，常是根据天然杂交率的高低划分的，一般可分为自花授粉植物 (self-pollinated plant)、常异花授粉植物 (often cross-pollinated plant) 和异花授粉植物 (cross-pollinated plant) 三种类型。栽培作物中约有 1/3 是自花授粉植物，如小麦、水稻、大豆等，不过它们也不是绝对自交繁殖，由于遗传基础和环境条件的影响，常发生少量的天然杂交 (1%~4%)。常异花授粉植物，如棉花、高粱等，其天然杂交率较高 (5%~20%)。自花授粉和常异花授粉植物绝大多数是雌雄同花，在自然状态

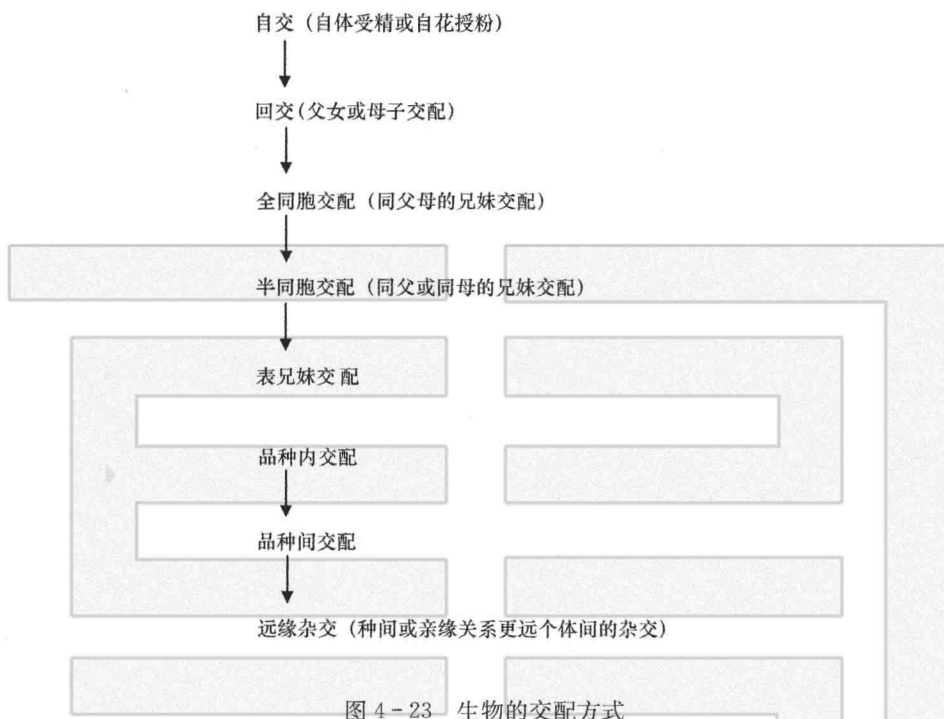


图 4-23 生物的交配方式

下大多能够实现自交繁殖。异花授粉植物天然杂交率高（20%~50%），如玉米、白菜型油菜等，在自然状态下是自由传粉的。

近亲繁殖的后代，特别是异花授粉植物的自交后代，一般表现生活力衰退，产量和品质下降，出现退化现象。但是遗传研究和育种工作中却十分强调自交或近亲繁殖。这是因为只有在自交或近亲繁殖的前提下，才能使供试材料具有纯合的遗传组成，从而才能更确切地分析和比较其杂种后代的遗传差异，研究性状的遗传规律，更有效地开展育种工作。

2. 自交的遗传效应

杂合体通过自交，主要表现为三方面的遗传效应。

(1) 导致杂合基因型的纯合

以一对基因为例，分析其自交后代群体的遗传组成。两个基因型纯合的亲本杂交 ($AA \times aa$)，其 F_1 是 100% 的杂合体 (Aa)。 F_1 自交产生 F_2 ， F_2 基因型的分离比例为 $1/4AA : 1/2Aa : 1/4aa$ ，其中纯合体 (AA, aa) 占 $1/2$ ，杂合体 (Aa) 也占 $1/2$ 。若继续自交，杂合的个体又产生 $1/2$ 纯合的后

代，而纯合体的个体只能产生纯合的后代。这样，每自交一代，杂合体减少 1/2，纯合体增加 1/2。连续自交 n 代（即 F_{n+1} 代），其后代杂合体逐步减少，而纯合体相应的逐步增加（表 4-15）。

若以 x 、 y 分别表示其纯合体和杂合体的频率，则

$$y = \left(\frac{1}{2}\right)^n$$

相应的

$$x = 1 - y = 1 - \left(\frac{1}{2}\right)^n = \frac{2^n - 1}{2^n}$$

其中， AA 、 aa 各占 1/2，即 AA 、 aa 的频率都为

$$\frac{1}{2} \left(\frac{2^n - 1}{2^n}\right)$$

表 4-15 一对杂合基因 (Aa) 连续自交的后代基因型比例的变化

世代	自交代数	基因型的比数	杂合体 (Aa)		纯合体 ($AA+aa$)	
			比数	%	比数	%
F_1	0	Aa	1	100%	0	0
F_2	1	1/4 AA 1/2 Aa 1/4 aa	1/2	1/2 ¹ =50%	1/2	1-1/2 ¹ =50%
F_3	2	3/8 AA 1/4 Aa 3/8 aa	1/4	1/2 ² =25%	3/4	1-1/2 ² =75%
F_4	3	7/16 AA 1/8 Aa 7/16 aa	1/8	1/2 ³ =12.5%	7/8	1-1/2 ³ =87.5%
F_5	4	15/32 AA 1/16 Aa 15/32 aa	1/16	1/2 ⁴ =6.25%	15/16	1-1/2 ⁴ =93.75%
...
F_{n+1}	n			1/2 ^{n} →0		1-1/2 ^{n} →100%

若有 m 对独立遗传基因，自交 n 代 (F_{n+1})，则杂合个体的频率为

$$y = \left(\frac{1}{2}\right)^{mn}$$

纯合个体的频率为

$$x = \left(\frac{2^n - 1}{2^n}\right)^m$$

约翰森 (Johannsen) 提出自交 n 代纯合个体数计算公式，可以计算纯合体的频率，并知道其基因组合中多少对为纯合的：

$$x = [1 + (2^n - 1)]^m \times (1/2)^{mn}$$

例如, $n=5, m=3$ 时

$$x = [1 + (2^5 - 1)]^3 \times (1/2)^{5 \times 3}$$

$$= (1 + 93 + 2883 + 29791) \times (1/32768)$$

其中, 三对基因纯合: $29291/32768=0.9091$; 三对基因杂合: $1/32768$;
 两对基因杂合 $93/32768$; 一对基因杂合: $2883/32768$ 。

应用以上公式必须具备两个条件: 一是各对基因为独立遗传; 二是各种基因型后代繁殖能力相同。按上式分列求出 1、5、10 和 15 对独立遗传基因自交 1~10 代的纯合率, 以此绘成曲线图 (图 4-24)。此曲线图表明, 在同一自交世代中, 等位基因对数越少, 纯合体占的比例就越大; 等位基因对数越多, 纯合体占的比例则越小。随自交世代的增加, 纯合体逐渐趋近于 100%。

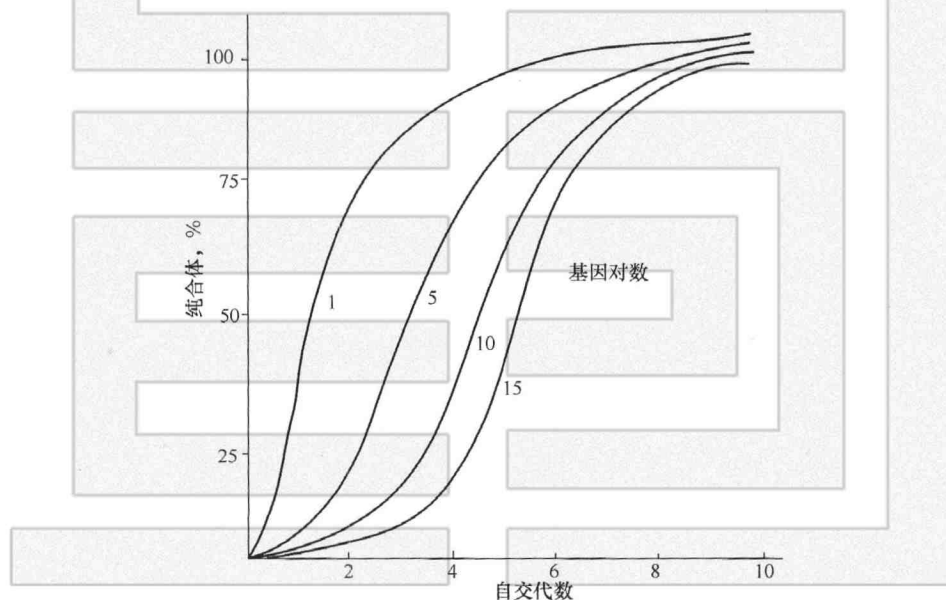


图 4-24 杂种所涉及的基因对数与自交后代纯合的关系

由上述分析可见, 杂合体通过自交可以导致后代基因分离, 并使后代群体的遗传组成迅速地趋于纯合。

(2) 淘汰有害隐性纯合体

杂合体通过自交, 必然导致等位基因的纯合而使隐性有害性状表现出来,

从而可以淘汰隐性有害个体，改良群体的遗传组成。在杂合状态下，隐性基因常被显性基因掩盖而不能表现出来。自花授粉作物由于长期自交，隐性性状可以表现，其有害的隐性性状已被自然选择和人工选择所淘汰。但是，异花授粉作物是杂合体，有害隐性基因常被显性的等位基因掩盖而不能表现。一经自交，由于成对基因的分离和重组，有害的隐性性状得以表现，如玉米自交后代出现白苗、黄苗、花苗、矮生等畸形性状，引起后代的严重衰退。通过对畸形植株的淘汰，控制畸形性状的隐性基因也随之清除了。

(3) 获取不同纯合基因型

杂合体通过自交遗传性状分离和重组，使同一群体内出现多个不同的纯合基因型。例如，两对基因的杂种 $AaBb$ ，通过长期自交，会出现 $AABB$ 、 $Aabb$ 、 $aaBB$ 、 $aabb$ 四种纯合基因型，逐代趋于稳定。这对于品种的保纯和物种的相对稳定都具有重要意义。

3. 回交遗传效应

回交 (back cross) 是指杂种后代与其两个亲本之一再次交配。如 $A \times B \rightarrow F_1$ ， $F_1 \times B \rightarrow BC_1$ ， $BC_1 \times B \rightarrow BC_2 \dots\dots$ 或 $F_1 \times A \rightarrow BC_1$ ， $BC_1 \times A \rightarrow BC_2 \dots\dots$ BC_1 表示回交一代， BC_2 表示回交二代，其余类推。被用来连续回交的亲本，称为轮回亲本 (recurrent parent)；相对地，未被用来回交的亲本，称为非轮回亲本 (non-recurrent parent)。

设两亲本的基因型为 AA 、 aa ， F_1 为 Aa ，则回交后代的遗传组成如表 4-16。可见， F_1 与 AA 回交， BC_1 基因型的分离比例为 $\frac{1}{2}AA : \frac{1}{2}Aa$ ，如继续与 AA 回交，纯合体仍形成纯合体，而杂合体又产生 $1/2$ 纯合体的后代和 $1/2$ 杂合体的后代。因此，回交 n 代后，杂合体 (y) 和纯合体 (x) 的频率分别为：

$$x = \frac{2^n - 1}{2^n}$$

$$y = \left(\frac{1}{2}\right)^n$$

上面两式中， n 指回交代数，即 BC_n 纯合体和杂合体的频率。

表 4-16 回交后代的遗传组成

世 代	交配方式	基因型频率	
		AA	Aa
P	AA × aa		
F ₁	Aa × Aa		
BC ₁	(1/2 AA + 1/2 Aa) × AA	1/2	1/2
BC ₂	(3/4 AA + 1/4 Aa) × AA	3/4	1/4
BC ₃	(7/8 AA + 1/8 Aa) × AA	7/8	1/8
BC ₄	(15/16 AA + 1/16 Aa) × AA	15/16	1/16
...	
BC _n		1 - (1/2) ⁿ	(1/2) ⁿ

若涉及 m 对基因，则

$$x = \left(\frac{2^n - 1}{2^n}\right)^m \times 100\%$$

$$y = \left(\frac{1}{2}\right)^m \times 100\%$$

回交后代纯合率和自交后代纯合率估算公式是一样的，但两者包含的内容却不同，自交后代纯合率是各种基因型纯合率的累加值，即自交后代将分离为多种纯合基因型。由于各对基因分离总数为 2^m ，所以每种基因型的纯合率为 $\left(\frac{2^n - 1}{2^n}\right)^m \times \left(\frac{1}{2}\right)^m$ ，累加为 $\left(\frac{2^n - 1}{2^n}\right)^m$ ；而回交后代只是轮回亲本一种纯合基因型，其纯合率为 $\left(\frac{2^n - 1}{2^n}\right)^m$ 。由此可见，在基因型纯合的进度上，回交显然大于自交。

回交的遗传解释，还有一种核置换理论（图 4-25）。两个亲本杂交后，F₁ 的核基因组成各占双亲的 1/2。经一次回交后，BC₁ 中所含轮回亲本的基因组成中，除了由轮回亲本直接提供 1/2 外，还由 F₁ 间接提供 (1/2) / 2 = 1/4，两者合起来为 1/2 + 1/4 = 3/4。同理，BC₂ 中由轮回亲本直接提供 1/2，由 BC₁ 间接提供 (3/4) / 2 = 3/8，两者合起来为 1/2 + 3/8 = 7/8。余类推。概括地说，一个杂种与其轮回亲本每回交一次，将使后代增加轮回亲本的 1/2 的基因组成；多次连续回交后，其后代基本上回复为轮回亲本的核基因组成。但细胞质仍为母本的基因组成。

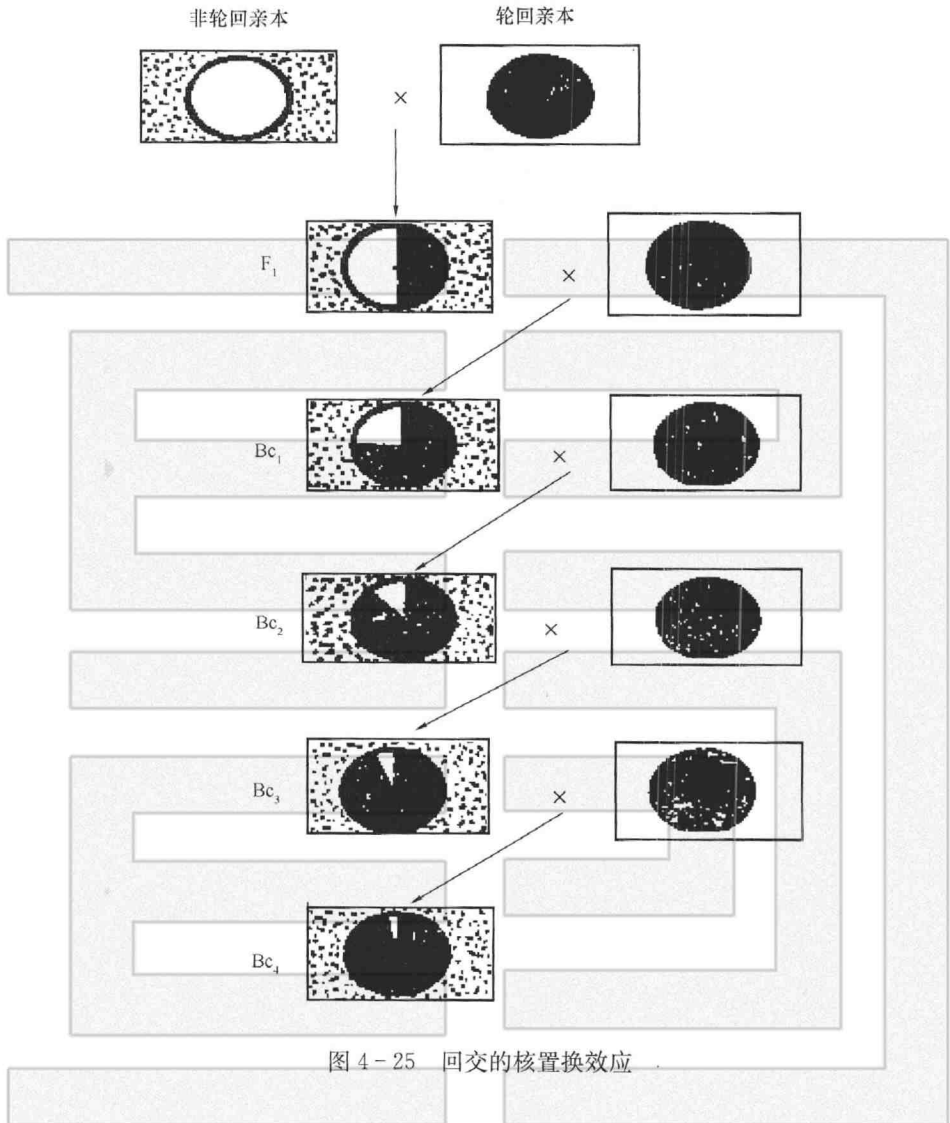


图 4-25 回交的核置换效应

二、杂种优势

1. 杂种优势的表现

杂种优势 (heterosis) 是指两个遗传组成不同的亲本杂交产生的杂种一代, 在生长势、生活力、繁殖力、产量和品质上比其双亲优越的现象。杂种优势所涉及的性状大都为数量性状, 故必须以具体的数值来衡量和表示其优势表现的程度。就某一性状而言, 通常以 F_1 超过双亲平均数即中亲值的百分

率来表示其优势程度，称为中亲优势。 F_1 超过双亲中最优亲本的杂种优势，为超亲优势。

$$\text{亲本本优} = \frac{F_1 - MP}{MP} \times 100\%$$

$$\text{超亲优势} = \frac{F_1 - HP}{HP} \times 100\%$$

式中， MP 为中亲值， HP 为较好亲本。

杂种优势是生物界既普遍又复杂的一种现象，凡是能进行有性生殖的生物，都可见到杂种优势现象，但并不是任何两个亲本杂交产生的杂种或者杂种的所有性状都表现优势。有些杂种与亲本水平相当，无明显的优势；有些不但没有优势，甚至还表现劣势，如高粱的产量性状优势很大，而品质性状却表现劣势。因此杂种优势也是一种很复杂的生物学现象。

杂种优势不是某一两个性状，而是许多性状综合表现突出。许多禾谷类作物的杂种一代，在产量和品质上表现为穗多、穗大、粒多、粒大、蛋白质含量高；生长势上表现为株高、茎粗、叶大、干物质积累快等；在抗逆性上表现为抗病、抗虫、抗寒、抗旱等杂种优势。

杂种优势的大小与诸多因素有关，一般来说，A. 异花授粉植物如玉米要比常异花授粉植物和自花授粉植物如棉花、小麦的杂种优势强。B. 在一定范围内，亲缘关系远、遗传差异大、双亲优缺点互补的组合，其杂种优势强；反之，就较弱。如原产我国的高粱品种与原产西非或南非的高粱品种间的杂种，其优势一般高于同一地区原产的品种间杂交；玉米马齿型与硬粒型自交系间杂交比同类型自交系间杂交表现较强的优势。C. 双亲基因型纯合程度高的杂种优势较高，如玉米自交系间杂种优势比品种间杂种优势为高。D. 在适宜的环境条件下种植比在不适宜的环境条件下种植的优势大。

根据遗传的基本规律， F_2 群体内必出现性状分离和重组。因此， F_2 和 F_1 相比，其生长势、生活力、抗逆性和产量等方面都显著地下降，即所谓衰退 (depression) 现象。并且 F_1 优势愈大，则其 F_2 表现衰退现象愈加明显。中国农科院作物所 (1958) 进行玉米杂种优势试验的结果表明，品种间杂种 F_2 比 F_1 减产 11.8%；双交种减产 16.2%；而单交种减产最多，为 34.1%。所以，在杂种优势利用上，一般不再利用 F_2 ，必须重新配制杂种，才能满足生

产的需要。

2. 杂种优势的遗传假说

迄今为止，杂种优势产生的原因尚无一致结论。目前主要有两种解释，即显性假说和超显性假说。近期，人们探讨了杂种优势的遗传学基础，并利用现代分子遗传学理论和技术，对杂种优势进行了研究。

(1) 显性假说

显性假说 (dominance hypothesis) 首先由 Bruce (1910) 提出，其基本论点是：杂交亲本的有利性状大都由显性基因控制，不利性状大都由隐性基因控制。通过杂交，使双亲的显性基因全部集中在杂种里，杂种优势是由于双亲的有利显性基因全部聚集在杂种里所引起的互补作用的结果。

显性假说的实验证据最早由两个豌豆品种杂交而获得。两个株高均为 5~6 英尺的豌豆品种，一个茎秆是节多而节间短，另一个的茎秆是节少而节间长，其 F_1 聚集了双亲的节多和节间长的显性基因，株高达 7~8 英尺，表现出明显的杂种优势。

根据显性假说，按独立分配规律，如所涉及的显隐性基因只是少数几对时，其 F_2 的理论次数应为 $(3/4+1/4)^n$ 的展开，表现为偏态分布。但事实上 F_2 一般仍表现正态分布。另外， F_2 以后虽然优势显著降低，但理论上应该能从其后代中选出具有与 F_1 同样优势，而且把全部纯合显性基因聚合起来的个体。然而，事实上很难选出这种后代。为此，Jones (1917) 又提出了显性连锁基因假说作了补充解释，认为一些显性基因与一些隐性基因位于各个同源染色体上，形成一定的连锁关系。而且控制某些有利性状的显性基因是非常多的，即 n 很大时，则 F_2 将不是偏态分布而是正态分布了。同时，在这样非常大的分离群体中，选出显性基因完全纯合的个体几乎是不可能的。

现以两个玉米自交系为例，说明显性假说。假定它们有 5 对基因互为显隐性关系，分别位于两对染色体上。同时设定各隐性纯合基因对性状发育的作用为 1 个单位，而各显性纯合基因和杂合基因的作用为 2 个单位。这两个自交系杂交产生的杂种优势可图示如下：

$$\begin{array}{c}
 P \quad \frac{A}{A} \frac{b}{b} \frac{C}{C} \frac{D}{D} \frac{e}{e} \quad \times \quad \frac{a}{a} \frac{B}{B} \frac{c}{c} \frac{d}{d} \frac{E}{E} \\
 (2+1+2+2+1=8) \quad \downarrow \quad (1+2+1+1+2=7) \\
 F_1 \quad \frac{A}{a} \frac{b}{B} \frac{C}{c} \frac{D}{d} \frac{e}{E} \\
 (2+2+2+2+2=10)
 \end{array}$$

可见，由于显性基因的互补作用， F_1 比双亲表现了显著的优势。

显性假说虽然得到了一些实验结果的直接证明，但也存在着一些缺点。如果杂种优势大小完全决定于有利显性基因的累加效应，即完全符合显性假说，那么两自交系杂交产生单交种 (F_1) 的产量就不可能超过两个亲本产量的总和。但事实上好的玉米单交种，其产量却大大超过双亲自交系之和。所以有利显性基因的累加效应，不能说是产生杂种优势的唯一原因，还应考虑到非等位基因间的相互作用等。

(2) 超显性假说

超显性假说 (overdominance hypothesis 或 superdominance hypothesis)，也称等位基因异质结合假说，由 Shull (1908) 首先提出。这个假说的基本论点是：杂种优势是由于双亲基因型的异质结合所引起的等位基因间的相互作用的结果。等位基因间没有显隐性关系，杂合的等位基因间相互作用大于纯合等位基因间的作用。

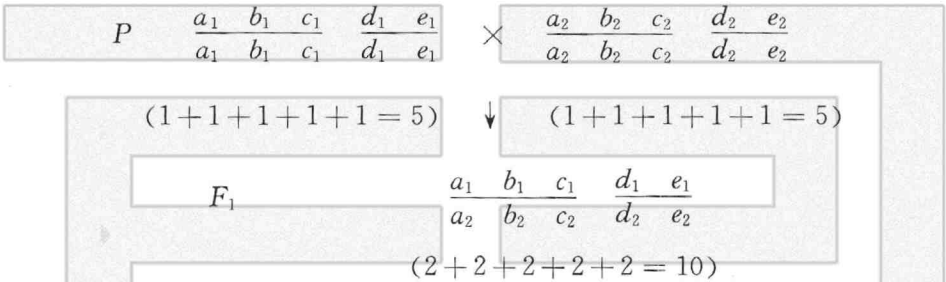
假定一对纯合等位基因 $a_1 a_1$ 能支配一种代谢功能，生长量为 10 个单位；另一对纯合等位基因 $a_2 a_2$ 具有另一种代谢功能，生长量为 4 个单位。杂种为杂合等位基因 $a_1 a_2$ ，它必将同时支配 a_1 和 a_2 所支配的两种代谢功能，于是生长量超过最优亲本而达到 10 个单位以上，即 $a_1 a_2 > a_1 a_1$ ， $a_1 a_2 > a_2 a_2$ 。由于这一假说解释了杂种超过最优等位基因纯合亲本的现象，故称为超显性假说。

两个亲本只有一对等位基因差异，杂交就能出现明显的杂种优势，这是超显性假说的直接证据。某些植物花色遗传有一对基因差别，但杂种植株的花色往往比其任一纯合亲本的花色都要深。如粉红色 \times 白色， F_1 表现红色，而 F_2 表现粉红色：红色：白色为 1：2：1 的简单比例。

许多生化遗传学的试验结果，有力地阐明了由异质等位基因所表现的杂种优势。例如，把一个不稳定而活泼的酶与一个稳定而不活泼的酶的异质等

位基因结合起来，获得的杂种酶表现既稳定又活泼。

为了说明超显性假说，现同样假定玉米的两个自交系各有 5 对基因与生长势有关，各对等位基因均无显隐性关系。同时设同质的等位基因（如 $a_1 a_1$ 、 $a_2 a_2$ ）的生长量为 1 个单位，而异质的等位基因（如 $a_1 a_2$ ）的生长量为 2 个单位。两个自交系杂交产生的杂种优势可图示如下：



由此可见，由于异质基因的互动， F_1 的表现则明显超过双亲。

越来越多的试验资料支持超显性假说。但是这一假说也存在着局限，它完全排斥了等位基因间的显隐性差别，排斥了显性基因在杂种优势表现中的作用。许多事实证明，杂种优势并不都是与等位基因的异质结合相一致的。例如，在自花授粉作物中，有一些杂种并不一定比其纯合亲本表现优势，甚至还有不如亲本的现象。

(3) 基因效应与杂种优势

数量遗传学把基因的作用区分为三种效应，即基因的加性效应、显性效应和互作效应（上位性效应）。超显性假说实际上是对显性效应所致的杂种优势而言的。显性假说主要指非等位基因的加性效应。而近来研究表明，非等位基因间的相互作用即上位性效应对杂种优势具有明显影响，即上位性假说。

综观生物界杂种优势的种种表现，三种假说解释的情况都存在。所以概括地说，杂种优势可能是由于上述某一个或某几个遗传机制造成的。既可能是由于双亲显性基因互补、异质等位基因互作和非等位基因互作的单一作用，也可能是由于这些因素的综合作用和累加作用所引起的。

随着分子标记技术的发展，利用分子标记遗传图谱进行 QTL 分析，可以获得 QTL 的作用方式、效应大小和位置。不少学者研究 QTL 与杂种优势的关系，认为 QTL 的作用方式与杂种优势有关，分别得到了支持显性假说、超显性假说和上位性假说的结论。例如，Xiao 等（1995）通过对水稻 QTL 分析

发现, 水稻杂种优势与整个基因组的杂合性不相关, 认为显性互补是水稻杂种优势产生的遗传基础, 支持显性假说; Stuber 等 (1992) 发现玉米中决定产量性状的 QTL 位点杂合性与杂种优势呈正相关, 认为超显性效应是杂种优势的重要遗传基础, 支持超显性假说; 余四斌等 (1998) 利用覆盖整个水稻基因组的分子连锁图进行 QTL 分析发现大量显著上位性效应广泛存在于基因组, 认为上位性效应在杂种优势形成中起决定性作用。因此, 杂种优势这一复杂的遗传现象可能是 QTL 多种作用方式的综合作用结果。因作物种类、生长时期、生长部位的不同, 可能有某种方式起关键作用, 但不能否认各种作用方式的相互联系与共同作用。

此外, 关于杂种优势的解释还有质核互补假说、遗传平衡假说等。在生理生化水平上, 发现线粒体互补和叶绿体互补以及杂种酶与杂种优势有关。

(4) 杂种优势的分子机理

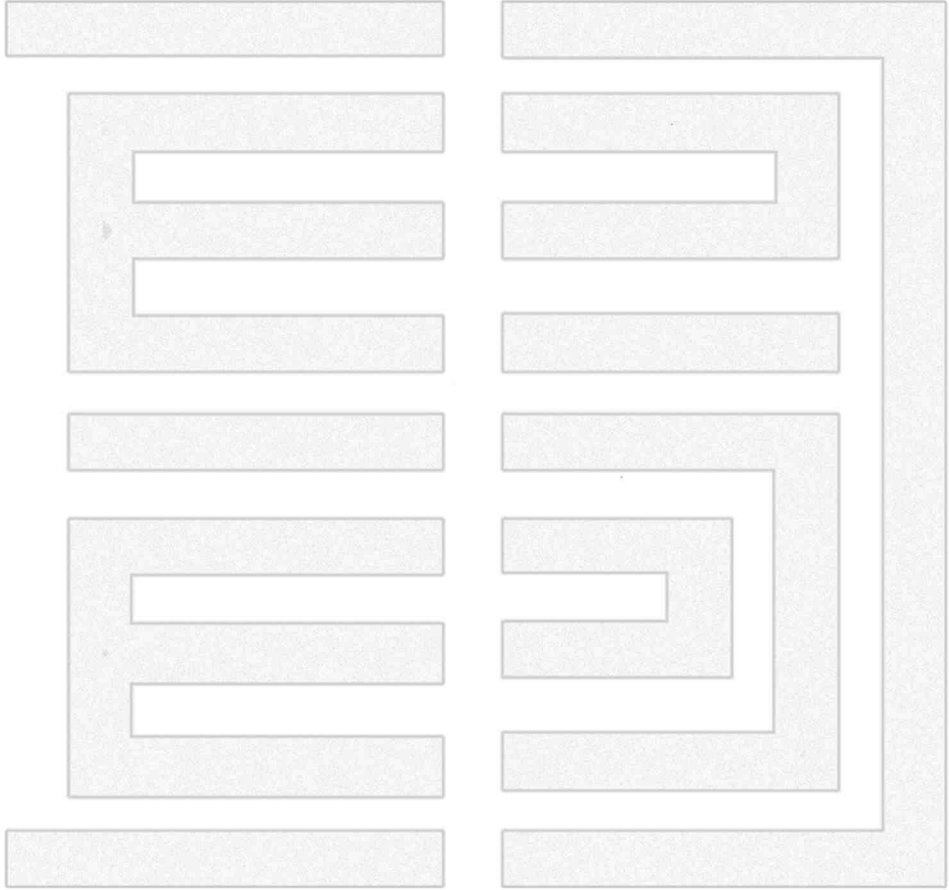
杂种优势的形成机理非常复杂, 从基因组水平来说, 杂种的全部基因均来自双亲, 杂种中并未产生新的基因, 而基因型不同的两个亲本杂交却可产生杂种优势。我们知道, 从基因组到性状表现, 基因表达和调控起了重要的作用。研究杂种与其亲本的基因表达差异, 从而在基因表达调控水平上解释杂种优势形成的机理显得更为直接和科学。

Romagnoli 等 (1990) 研究了基因表达与玉米杂种优势的关系。他们从杂交种 cDNA 文库中筛选出了 3 个在杂交种和亲本自交系中差异表达的克隆, 其中 1 个克隆在 F_1 的表达介于双亲之间, 另外 2 个克隆在 F_1 的表达接近高效表达亲本。他们还对比杂交种和亲本的 PolyA RNA 进行了离体翻译, 发现有 33% 的差异表达产物在杂交种中更丰富或特异表达。此后, 许多学者利用玉米、水稻和小麦等证明了杂交种和亲本之间存在明显的基因表达差异。

由于转录调控是基因表达调控的一种重要机制, 转录因子又是参与调控的重要反式作用因子, 而基因家族则是编码转录因子的一系列功能和结构相似的基因。因此, 研究基因家族在亲本与杂交种间的差异, 有可能从基因表达方面揭示杂种优势的分子机理。各性状杂种优势与不同基因家族不同差异表达模式表现出不同程度的显著相关, 总趋势是杂种优势和双亲共沉默类型呈显著负相关, 和杂种特异表达类型呈显著正相关。这意味着杂种优势的形成是因为对杂种优势具有正效应基因的表达和具有负效应基因的抑制所造

成的。

此外，从分子水平提出了基因网络系统、杂种的自组织理论、遗传振动合成学说和双向基因重组方式下再选择等来解释杂种优势，但缺乏实验依据。随着分子生物学的发展，从分子水平来探讨杂种优势，可能使人们更好地理解杂种优势的遗传机理。



第五章 中国少数民族遗传学 的数据处理

孟德尔对他自己所得的分离比并未作过统计学分析，而且也未认识到这种处理的必要性。但他也看到分离比，如 1 : 1、3 : 1 等都是要子代个体数较多时才比较接近；子代个体数不多时，其实际所得比例与理论比例常表现明显的波动。到 20 世纪初期，孟德尔的遗传规律被重新发现以后，通过对大量的遗传试验资料的统计分析，才从逻辑上明确和证实了统计学分析在遗传研究中的必要性和重要性。

第一节 概率原理

一、概率原理

1. 概率的概念

概率 (probability) 是指一定事件总体中某一事件可能出现的几率。例如，在孟德尔的豌豆试验中，圆粒 × 皱粒的 F_1 为杂合基因型 Rr 。当 F_1 植株的花粉母细胞进行减数分裂时，其中 R 和 r 基因分配到每个雄配子的机会是均等的。这就是说，在所形成的雄配子总体中带有 R 或 r 基因的雄配子概率各为 $\frac{1}{2}$ 。又如抗病植株 RR 和感病植株 rr 杂交的 F_1Rr 进行自交，预计 F_2 代将出现 RR 、 Rr 、 Rr 、 rr 4 种组合的个体，比例为 1 : 1 : 1 : 1。因而子代中 RR 基因型个体出现的概率为 $\frac{1}{4}$ 。 RR 、 Rr 都是抗病的，所以 F_2 代中抗病植株出现的概率为 $\frac{3}{4}$ 。

因此，在遗传研究中可以通过概率来推算遗传比率，从而分析和判断该

比率发生的真实性和可靠性。

2. 概率的基本定理

在遗传学研究中，主要应用概率的两个基本定理，即乘法定理和加法定理。

(1) 乘法定理

两个独立事件同时发生的概率等于各个事件发生的概率的乘积。例如，豌豆黄子叶、圆粒绿子叶、皱粒产生的杂种基因型为 $YyRr$ ，由于这两对性状是受两对独立的基因所控制，表明它们是独立事件。所以，当植株在减数分裂形成配子时，两个非等位基因同时进入某一雄配子的概率将是各基因概率的乘积，即 $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ 。同理，两个非等位基因同时进入某一雌配子的概率将是 $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ 。虽然高等动植物在减数分裂时所形成的四个大孢子，只有一个能继续发育为雌配子，但其两个非等位基因所形成的各种基因型的配子概率仍然是 $\frac{1}{4}$ 。在杂种 $YyRr$ 中，杂合基因对数 $n=2$ ，故该杂种可形成 $2^n = 2^2 = 4$ 种配子。根据概率的乘法定理，四种配子（雌配子或雄配子）中的基因组合及其出现的概率是： $YR = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ ， $Yr = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ ， $yR = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ ， $yr = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ 。

2. 加法定理

加法定律适合于互斥事件。所谓互斥事件是某一事件出现，另一事件即被排斥。两个互斥事件任意发生的概率是各个事件各自发生的概率之和。例如，豌豆子叶不是黄色，就是绿色，二者只居其一。因此，如果问豌豆子叶黄色或绿色的概率是多少，则是二者概率之和，即 $\frac{1}{2} + \frac{1}{2} = 1$ 。

根据上述概率的两个定理，现将豌豆杂种 $YyRr$ 的雌雄配子发生的概率，通过受精的随机结合所形成的合子基因型及其概率可按棋盘方法表示为：

	♂ 配子	$\frac{1}{4} YR$	$\frac{1}{4} Yr$	$\frac{1}{4} yR$	$\frac{1}{4} yr$
♀ 配子	$\frac{1}{4} YR$	$\frac{1}{16} YYRR$	$\frac{1}{16} YYRr$	$\frac{1}{16} YyRR$	$\frac{1}{16} YyRr$

续表

♂配子 ♀配子	$\frac{1}{4} YR$	$\frac{1}{4} Yr$	$\frac{1}{4} yR$	$\frac{1}{4} yr$
$\frac{1}{4} Yr$	$\frac{1}{16} YYRr$	$\frac{1}{16} YYrr$	$\frac{1}{16} YyRr$	$\frac{1}{16} Yyrr$
$\frac{1}{4} yR$	$\frac{1}{16} YyRR$	$\frac{1}{16} YyRr$	$\frac{1}{16} yyRR$	$\frac{1}{16} yyRr$
$\frac{1}{4} yr$	$\frac{1}{16} YyRr$	$\frac{1}{16} Yyrr$	$\frac{1}{16} yyRr$	$\frac{1}{16} yyrr$

由此可见，同一配子中具有互斥性质的等位基因不可能同时存在，故形成了YR、Yr、yR、yr 4种配子，且其概率各为 $\frac{1}{4}$ 。它们的雌雄配子受精结合成为16种合子，各个雌配子和雄配子受精结合为一种基因型的合子以后，它就不可能再同时形成另一种基因型的合子。也就是说通过受精形成的组合彼此是互斥事件。因此，可以把上述F₂群体的表现型和基因型进一步归纳成表5-1。

表5-1 豌豆杂种YyRr自交产生的F₂群体中各基因组合的概率

配子	♀	♂	概率	♀	♂	概率	♀	♂	概率	♀	♂	概率	♀	♂	概率
	YR	YR		YR	Yr		YR	Yr		yR	YR		yR	Yr	
子代基因型的排列			$\frac{1}{16}$	YR	Yr	$\frac{1}{16}$	YR	yr	$\frac{1}{16}$	Yr	yr	$\frac{1}{16}$	yr	yr	$\frac{1}{16}$
				YR	yR	$\frac{1}{16}$	yr	YR	$\frac{1}{16}$	yR	yr	$\frac{1}{16}$			
				Yr	YR	$\frac{1}{16}$	Yr	Yr	$\frac{1}{16}$	yr	Yr	$\frac{1}{16}$			
				yR	YR	$\frac{1}{16}$	yR	yR	$\frac{1}{16}$	yr	yR	$\frac{1}{16}$			
							yR	Yr	$\frac{1}{16}$						
组合							Yr	yR	$\frac{1}{16}$						
	4显性基因		$\frac{1}{16}$	3显性基因 1隐性基因		$\frac{4}{16}$	2显性基因 2隐性基因		$\frac{6}{16}$	1显性基因 3隐性基因		$\frac{4}{16}$	4隐性基因		$\frac{1}{16}$

在表5-1中F₂的群体里，纯合基因型个体的概率较低，只有 $2^n = 2^2 = 4$ 种；而杂合基因型个体的概率较高，共有 $3^n - 2^n = 3^2 - 2^2 = 9 - 4 = 5$ 种。如杂合基因的对数增多，其差异将更大。

二、二项式展开

在遗传研究中,有时需要推算一定个体数目的某种组合事件出现的概率。例如 $Aa \times Aa$ 的子代中,表现 A 性状的个体概率是 $3/4$,表现 a 性状的个体的概率是 $1/4$ 。现在问,如果 $Aa \times Aa$ 只产生两个子代个体,那么这两个个体都表现 A 性状的概率是多少?两个个体都表现 a 性状的概率是多少?两个个体一个表现 A 性状、另一个表现 a 性状的概率是多少?

类似上述的遗传问题可以采用二项式公式进行分析:

设 p = 某一事件出现的概率, q = 另一事件出现的概率, $p + q = 1$ 。 n = 为估测其出现概率的事件数。二项式展开的公式为:

$$(p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{2!}p^{n-2}q^2 + \frac{n(n-1)(n-2)}{3!}p^{n-3}q^3 + \dots + q^n$$

当 n 较大时,二项式展开的公式过长。为了方便,如仅推算其中某一项事件出现的概率,可用以下通式:

$$\frac{n!}{r!(n-r)!}p^r q^{n-r}$$

r 代表某事件(基因型或表现型)出现的次数; $n-r$ 代表另一事件(基因型或表现型)出现的次数;!代表阶乘符号,如 $4!$,即表示 $4 \times 3 \times 2 \times 1 = 24$ 。应该注意: $0!$ 或任何数的 0 次方均等于 1 。

上述 $Aa \times Aa$ 子代的问题可以这样来推算:设 p 为表现 A 性状个体的概率, q 为表现 a 性状个体的概率, n 为子代个体数。则 $p = \frac{3}{4}$, $q = \frac{1}{4}$, $p + q = 1$ 。二项式展开的公式为:

$$\begin{aligned} (p+q)^n &= \left(\frac{3}{4} + \frac{1}{4}\right)^2 \\ &= \left(\frac{3}{4}\right)^2 + 2\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)^2 \\ &= \frac{9}{16} + \frac{6}{16} + \frac{1}{16} \end{aligned}$$

上述数字表明,两个个体都表现 A 性状的概率是 $\frac{9}{16}$;两个个体一个表现 A 性状、另一个表现 a 性状的概率是 $\frac{6}{16}$;两个个体都表现 a 性状的概率是 $\frac{1}{16}$ 。

利用二项式展开也可以分析推算杂种后代群体的基因型或表现型的分布。

以杂种 $YyRr$ 为例，用二项式展开分析其后代群体的基因型分布。

显性基因 Y 或 R 出现的概率 $p = \frac{1}{2}$ ，隐性基因 y 或 r 出现概率 $q = \frac{1}{2}$ ，

$p+q = \frac{1}{2} + \frac{1}{2} = 1$ ， $n=4$ 为杂合基因个数。则：

$$\begin{aligned} (p+q)^n &= \left(\frac{1}{2} + \frac{1}{2}\right)^4 \\ &= \left(\frac{1}{2}\right)^4 + 4\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + \frac{4 \times 3}{2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + \frac{4 \times 3 \times 2}{3!}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 \\ &= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} \end{aligned}$$

这样计算所得的各项概率，即如表 4-6 所列的结果：4 显性基因为 $\frac{1}{16}$ ，3 显性和 1 隐性基因为 $\frac{4}{16}$ ，2 显性和 2 隐性基因为 $\frac{6}{16}$ ，1 显性和 3 隐性基因为 $\frac{4}{16}$ ，4 隐性基因为 1。

如果只需了解 3 显性和 1 隐性基因个体出现的概率，即 $n=4$ ， $r=3$ ， $n-r=4-3=1$ ，则可采用单项事件概率的通式进行推算，获得同样结果：

$$\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{4!}{3!(4-3)!} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) = \frac{4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 1} \left(\frac{1}{8}\right) \left(\frac{1}{2}\right) = \frac{4}{16}$$

上述二项式展开不但可以应用于杂种后代 F_2 群体基因型的排列和分析，同样可以应用于测交后代 F_1 群体中表现型的分布。因为测交后代，显性个体和隐性个体出现的概率也都分别是 $\frac{1}{2}$ ($p = \frac{1}{2}$ ， $q = \frac{1}{2}$)。

此外，如果推算杂种自交的 F_2 群体中表现型分布，同样可以采用二项式进行分析。根据孟德尔的遗传规律，任何一对完全显隐性的杂合基因型，其自交的 F_2 群体中，显性性状出现的概率 $p = \frac{3}{4}$ ，隐性性状出现的概率 $q = \frac{1}{4}$ ，

$$p+q = \frac{3}{4} + \frac{1}{4} = 1。$$

n 代表杂合基因对数。

例如，两对基因杂种 $YyRr$ 自交产生的 F_2 群体，其表现型个体的概率按

上述 $\frac{3}{4} : \frac{1}{4}$ 的概率代入二项式展开为:

$$\begin{aligned}(p+q)^n &= \left(\frac{3}{4} + \frac{1}{4}\right)^2 \\ &= \left(\frac{3}{4}\right)^2 + 2\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)^2 \\ &= \frac{9}{16} + \frac{6}{16} + \frac{1}{16}\end{aligned}$$

这表明具有两个显性性状 ($Y_R_$) 的个体概率为 $\frac{9}{16}$, 一个显性性状和一个隐性性状 (Y_rr 和 $yyR_$) 的个体概率为 $\frac{6}{16}$, 两个隐性性状 ($yyrr$) 的个体概率为 $\frac{1}{16}$, 即表现型的遗传比率为 $9 : 3 : 3 : 1$ 。

同理, 如果是三对基因杂种 $YyRrCc$, 其自交的 F_2 群体的表现型分布, 可按二项式展开求得:

$$\begin{aligned}(p+q)^n &= (p+q)^3 \\ &= \left(\frac{3}{4}\right)^3 + 3\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + 3\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 \\ &= \frac{27}{64} + \frac{27}{64} + \frac{9}{64} + \frac{1}{64}\end{aligned}$$

这表明具有三个显性性状 ($Y_R_C_$) 的个体概率为 $\frac{27}{64}$, 两个显性性状和一个隐性性状 (Y_R_cc 、 $Y_rrC_$ 和 $yyR_C_$ 各占 $\frac{9}{64}$) 的个体概率为 $\frac{27}{64}$, 一个显性性状和两个隐性性状 (Y_rrcc 、 yyR_cc 和 $yyrrC_$ 各占 $\frac{3}{64}$) 的个体概率为 $\frac{9}{64}$, 三个隐性性状 ($yyrrcc$) 的个体概率为 $\frac{1}{64}$ 。即表现型的遗传比率为 $27 : 9 : 9 : 9 : 3 : 3 : 3 : 1$ 。

如果需要了解 F_2 群体中某种表现型个体出现的概率, 也同样可用上述单项事件概率的通式进行推算。例如, 在三对基因杂种 $YyRrCc$ 的 F_2 群体中, 计算两显性性状和一隐性性状个体出现的概率时, $n=3$, $r=2$, $n-r=3-2=1$, 则可按上述通式求得:

$$\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{3!}{2!(3-2)!} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right) = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} \left(\frac{9}{16}\right) \left(\frac{1}{4}\right) = \frac{27}{64}$$

第二节 卡方检验

在遗传学试验中，由于各种因素的干扰，实际获得的各项观察值与其理论上按概率估算的期望数值常具有一定的偏差。这种偏差主要是由两种因素造成：一是由于群体小等原因产生的随机误差，另一是由遗传因素造成的内在本质的差异。如果要判断这种偏差究竟是属于试验误差造成的，还是本质的差异，通常需要适合度测验。适合度测验是比较试验数据与理论假设是否符合的假设测验。统计学中通常用 χ^2 检验来进行。 χ^2 值是经过统计学处理后计算出的用以度量观察数与理论数相差程度的一个指数。根据 χ^2 值查知误差概率的大小，从而可以判断偏差的性质。

进行 χ^2 测验时可利用以下公式计算 χ^2 值：

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

其中， O 是实测值， E 是理论值， Σ (sigma) 是总和的符号，代表许多上述比值的总和。从以上公式可以说明，所谓 χ^2 值代表理论值与实际值之间变异的总和。

算出 χ^2 值，根据自由度（用 df 表示），就可以查出概率 P 值。自由度一般为子代分离类型的数目减 1，若用 k 表示类型数，则 $df = k - 1$ 。例如，子代为 1:1 或 3:1 的两种分离类型，自由度是 1；9:3:3:1 的四种分离类型，自由度为 3。

例如，用 χ^2 检验验证孟德尔两对相对性状的杂交试验结果（表 5-2）。

表 5-2 孟德尔两对基因杂种自交结果的测验

	圆、黄	圆、绿	皱、黄	皱、绿	总 数
实测值 (O)	315	108	101	32	556
理论值 (E)	312.75	104.25	104.25	34.75	556
($O-E$)	2.25	3.75	-3.25	-2.75	

续表

	圆、黄	圆、绿	皱、黄	皱、绿	总数
$(O-E)^2$	5.06	14.06	10.56	7.56	
$\frac{(O-E)^2}{E}$	0.016	0.135	0.101	0.218	
$\chi^2 = \sum \frac{(O-E)^2}{E}$	$X^2 = 0.016 + 0.135 + 0.101 + 0.218 = 0.47$				

注：理论值是由总数 556 粒种子按 9 : 3 : 3 : 1 分配求得的。

由表 5-2 求得所谓 χ^2 值为 0.47，自由度为 3，查表 5-3 得 P 值为 0.90 ~ 0.99 之间，说明实际值与理论值差异由随机误差引起的概率在 90% 以上，因而样本的表现型比例符合 9 : 3 : 3 : 1。要指出的是，在遗传学实验中 P 值常以 5% (0.05) 为标准， $P > 0.05$ 说明“差异无统计学意义”，实际值与理论值相符合； $P < 0.05$ 说明“差异有统计学意义”，实际值与理论值不符合；如果 $P < 0.01$ 说明“有更大的把握认为差异有统计学意义”。

χ^2 检验法不能用于百分数资料的适合度检验。如果遇到百分比应根据总数把他们化成频数，然后计算差数，例如，在一个实验中得到雌果蝇 44%，雄果蝇 56%，总数是 50 只，现在要测验一下这个实际数值与理论数值是否相符，这就需要首先把百分比根据总数化成频数，即 $50 \times 44\% = 22$ 只， $50 \times 56\% = 28$ 只，然后按照 χ^2 检验公式求 χ^2 值。

表 5-3 χ^2 表

P	0.99	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
df											
1	0.00016	0.04	0.016	0.064	0.148	0.455	1.074	1.642	2.706	3.841	6.635
2	0.0201	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	9.210
3	0.115	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	11.345
4	0.297	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	13.277
5	0.554	1.145	1.610	2.343	3.000	4.351	6.064	7.269	9.236	11.070	15.086
6	0.872	1.635	2.204	3.070	3.828	5.345	7.231	8.588	10.645	12.592	16.812
7	1.239	2.167	2.833	3.822	4.671	6.346	8.783	9.803	12.017	14.067	18.475
8	1.646	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	20.090
9	2.088	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	21.666
10	2.558	3.940	4.865	6.179	7.227	9.342	11.781	13.442	15.987	18.307	23.209

表内数字是各种 χ^2 值, df 为自由度, P 是在一定自由度下 χ^2 大于表中数值的概率。

第三节 遗传度、杂合度、多态信息量和吻合度测验

一、遗传度计算

在多基因病中, 易患性的高低受遗传基础和环境因素双重影响。其中遗传基础所起作用的大小称为遗传率或遗传度。遗传度一般用百分率 (%) 来表示。一种多基因病如果完全由遗传基础决定其易患性变异和发病, 遗传率就是 100%。对于疾病, 遗传率高, 表明遗传基础决定疾病的决定性作用大。

1. 先天性髋关节脱位的遗传度计算

对先天性髋关节脱位的调查发现, 该病的群体发病率为 0.1%, 在先证者的一级亲属 1777 人中, 有 35 人患该病, 二级亲属 4746 人中, 有 16 人患该病, 三级亲属 4220 人中, 有 8 人患该病, 请根据上述数据, 按所给公式估算先天性髋关节脱位的遗传度, 并将结果填入表 5-4。

$$h^2 = \frac{b}{r}, b = \frac{X_g - X_r}{a_g}, V_b = \left(\frac{1}{a_g}\right)^2 W_r = \left(\frac{1}{a_g}\right)^2 \left(\frac{1 - q_r}{a_r^2 A_r}\right)$$

$$S_{h^2} = \frac{\sqrt{V_b}}{r}, t = \frac{h^2}{S_{h^2}}$$

$$h^2 \text{ 的加权平均值} = \frac{\frac{h_1^2}{S_1^2} + \frac{h_2^2}{S_2^2} + \frac{h_3^2}{S_3^2} + \Lambda + \frac{h_n^2}{S_n^2}}{\frac{1}{S_1^2} + \frac{1}{S_2^2} + \frac{1}{S_3^2} + \Lambda + \frac{1}{S_n^2}}$$

$$S_{h^2} \text{ 的加权平均值} = \frac{1}{\sqrt{\frac{1}{S_1^2} + \frac{1}{S_2^2} + \frac{1}{S_3^2} + \Lambda + \frac{1}{S_n^2}}}$$

其中 b — 亲属易患性对先证者易患性的回归系数; r — 亲属系数; X_g — 一般群体易患性平均值与阈值之间的标准差数; X_r — 先证者亲属易患性平均值与阈值之间的标准差数; a_g — 一般群体易患性平均值与一般群体中患者易患性平均值之间的标准差数; a_r — 先证者亲属易患性平均值与先证者亲属中患者易

患性平均值之间的标准差数； q_g ——一般群体发病率； q_r ——先证者亲属发病率； A_r ——先证者亲属中的病例数； V_b —— b 的方差； S_{h^2} —— h^2 的标准误； $h_1^2, h_2^2, h_3^2, \dots, h_n^2$ ——分别为各级亲属的遗传度； $S_1, S_2, S_3, \dots, S_n$ ——分别为各级亲属遗传度的标准误。

X_g, X_r 和 a_g, a_r 均可由一般群体发病率和先证者亲属发病率查 Falconer 表得到。

表 5-4 先天性髋关节脱位的遗传度估算

	A_r	N	$q(\#)$	X	a	b	V_b	$S_{h^2}(\#)$	$h^2(\#)$	t	P
一般群体											
一级亲属											
二级亲属											
三级亲属											
加权均值											

二、杂合度和多态信息量的计算

杂合度是指由所有标记检测到的两亲本间的差异程度，杂合度用于群体遗传学上对群体多态性的描述；而多态信息量用于连锁分析时对标志基因（或标志序列）的多态性估计。

从某汉族人群随机抽取 171 人组成样本，对位于第 8 号染色体的 N-乙酰化转移酶 2 (N-acetyltransferase2, NAT2) 位点进行基因分型，结果见表 5-5。请按所给公式计算该人群 NAT2 位点的杂合度 (heterozygosity, h) 和多态信息容量 (polymorphic information content, PIC)。

$$h = 1 - \sum p_i^2, PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

式中 p_i ：第 i 个等位基因的频率； p_j ：第 j 个等位基因的频率； n ：该位点的等位基因数。

表 5-5 某人群 NAT2 位点的等位基因频率

等位基因	等位基因数	基因频率	等位基因	等位基因数	基因频率	等位基因	等位基因数	基因频率
F1	138		S3	15		S8	14	
S1a	1		S4	23		S9	11	
S1b	14		S5	1		S10	3	
S1c	2		S6	2		S11	19	
S2	43		S7	41		S12	15	

三、基因的吻合度测验

检测 106 例健康汉族人的 NAT2 基因型，发现基因型 F_1/F_1 为 24 例， F_1/S 为 52 例， S/S 为 30 例，根据该检测结果，就 NAT2 基因而言，基因频率估计的可靠性和样本的代表性如何（结果填入表 5-6）。

等位基因 F1 的频率 $p=$ 等位基因 S 的频率 $q=$

表 5-6 NAT2 基因的吻合度测验

基因型	期望频率	期望数	观察数	χ^2
F_1/F_1				
F_1/S				
S/S				
合计				

$$\chi^2 = \sum \frac{(\text{期望数} - \text{观察数})^2}{\text{期望数}}$$

$$\begin{aligned} \text{自由度 } df &= \text{基因型类型的自由度} - \text{等位基因数的自由度} \\ &= (\text{基因型数} - 1) - (\text{等位基因数} - 1) \end{aligned}$$

第四节 主成分分析

所谓主成分分析 (principal component Analysis) 是根据原始变量间的相互关系，寻找一组由原变量组成，而彼此不相关的综合变量，从而浓缩原数

据信息、简化数据结构、压缩数据规模的一种统计方法。为方便说明,假定有 n 个个体的两指标(如身高 x_1 , 体重 x_2), x_1 与 x_2 有显著性的相关。当 n 很大时, n 对观测值在平面坐标系中形成椭圆形的散点分布。由于身高与体重的相关性, n 个点分布成椭圆形。如果把通过该椭圆长轴的直线取作新坐标系的横轴 Z_1 , 在长轴原点取一条垂直于 Z_1 的直线作为新坐标系的纵轴 Z_2 。于是这 n 个点在新坐标系 Z_1 和 Z_2 中的坐标发生了改变, 但其点间的相对位置、点群的形态均未改变。 n 点在这两个坐标系中的坐标有如下关系:

$$\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2 = \sum_{i=1}^n (z_{i1} - \bar{z}_1)^2 + \sum_{i=1}^n (z_{i2} - \bar{z}_2)^2$$

新坐标轴 Z_1 与 n 个点分布的长轴一致, 反映了 n 个个体间离差的最大部分信息。若 Z_1 反映了原数据总信息的 80%, 则 Z_2 只反映总信息的 20%。另一方面, 这 n 个点在新坐标系中的分布是不相关的。这样研究 n 个点的性质就可以在新坐标系上进行。新指标 z_1 称为原指标的第一主成分, z_2 称为第二主成分。当椭圆的短轴相对于长轴很小时, 像一长条带状分布, 第一主成分 z_1 指标便解释了数据变异的绝大部分, 即上式右边第一项很大, 而第二主成分 z_2 , 即上式右第二项很小, 则可以忽略不计。

若原变量有 x_1 、 x_2 和 x_3 三个, 且彼此相关, 则 n 个样本点在三维空间呈椭球形分布。同样取椭球的长轴为第一主成分, 取较短轴为第二主成分, 而取最小的短轴为第三主成分, 并且这三个主成分正交。当第三主成分解释的变异很小时, 即其分布呈一压扁的椭圆烧饼状时, 只用前面两个主成分便能反映原始数据的绝大部分信息。若第二主成分的变异也很小时, 即 n 个点在三维空间分布呈一长杆带状时, 则只需用一个综合指标(第一主成分)便能反映原始数据中三个变量的基本特征。

拓展到更一般情况, 假定原数据有 p 个指标, x_j ($j=1, 2, \dots, p$), 同理可以寻找出由这 P 个原变量的线性组合所产生相互独立(正交)的 p 个新指标, 并且按其变异由大到小排列, 记为 z_1, z_2, \dots, z_p 。 z_1 显示最大变异, z_2 次之, 如此类推, z_p 的变异最小, 即

$$\text{Var}(z_1) \geq \text{Var}(z_2) \geq \dots \geq \text{Var}(z_p)$$

其中 $\text{Var}(z_i)$ 为 z_i 的方差

z_i 称为第 i 个主成分。

在进行主成分分析时，总希望用少数几个主成分反映原数据绝大部分信息，使得变异小的几个主成分可以忽略不计，从而简化数据结构。

从数学的角度看主成分分析，实际上就是对 p 个原始变量 x_j ($j=1, 2, \dots, p$) 进行一次线性变换，产生 m 个 ($m \leq p$) 新的综合变量 (z_1, z_2, \dots, z_m) 的过程，使新变量 z_i ($i=1, 2, \dots, m$) 与原变量 x_j ($j=1, 2, \dots, p$) 有如下关系：

$$\begin{cases} z_1 = w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p \\ z_2 = w_{21}x_1 + w_{22}x_2 + \dots + w_{2p}x_p \\ \dots \\ z_m = w_{m1}x_1 + w_{m2}x_2 + \dots + w_{mp}x_p \end{cases}$$

其中 z_i 为第 i 个主成分 ($i=1, 2, \dots, m$)，

x_j 为第 j 变量 x_j ，

w_{ij} 为第 j 变量 x_j 在第 i 主成分 z_i 上的权重系数，又称载荷。

z_1 是所有 m 个主成分中方差最大者，而且与其他主成分都无关，服从条件

$$w_{211} + w_{221} + \dots + w_{2m1} + \dots + w_{2p1} = 1$$

于是 z_1 的方差 $\text{Var}(z_1)$ 是在这一给定条件下的最大。因为如果没有该约束条件，则 $\text{Var}(z_1)$ 可能通过简单地增加任何一个 w_{1j} 而增大。第二主成分 z_2 是服从

$$w_{221} + w_{222} + \dots + w_{2m2} + \dots + w_{2p2} = 1$$

约束条件下 m 个主成分中第二个方差大者，而且与其他主成分无关。其他主成分用相同的方法加以定义。

一、主成分的求解

主成分分析实质上是实施对原变量向新变量的数据转换过程，其关键在于依次找出样本中变异最大的方向作为 m 维空间的新坐标轴方向，即各主成分的方向。由上面的公式可见，各主成分 z_i ($i=1, 2, \dots, m$) 都是原变量 x_j ($j=1, 2, \dots, p$) 的线性组合。欲求主成分 z_i 实际上就是求上式中各 x_j 所对应的系数，称为第 j 个原变量对第 i 个主成分的权重系数。其求解过程为：

1. 由原始数据计算 p 个 x_j 间的相关矩阵 R 。该矩阵为 $p \times p$ 的对称矩阵，

其元素 r_{jk} 表示原变量 x_j 与 x_k 的相关系数, 当 $j=k$ 时, 则为 1, 而且 $r_{jk} = r_{kj}$ 。其中 r_{jk} 的定义见多元相关。

2. 求相关矩阵 R 的特征值 (Eigenvalue): 由 R 矩阵建立特征矩阵: $R - \lambda I$

其中 R 为相关矩阵;

λ 为相关矩阵 R 的特征值;

I 为单位矩阵。

令该特征矩阵行列式为零, 即得到含未知数 λ 的 p 次方程, 称为特征方程, 即

$$\begin{vmatrix} 1-\lambda r_{11} & r_{12} & \cdots & r_{1p} \\ r_{211} & 1-\lambda & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & 1-\lambda \end{vmatrix} = 0$$

解此一元 p 次方程, 便获得 p 个特征值, 表示为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$$

3. 求特征向量 (Eigenvector) 令 λ_i 为第 i 特征值。于是可构造如下列齐次方程组:

$$\begin{cases} (1-\lambda_i) y_1 + r_{12} y_2 + \cdots + r_{1p} y_p = 0 \\ r_{21} y_1 + (1-\lambda_i) y_2 + \cdots + r_{2p} y_p = 0 \\ \cdots \\ r_{p1} y_1 + r_{p2} y_2 + \cdots + (1-\lambda_i) y_p = 0 \end{cases}$$

解此方程组得到 y_1, y_2, \cdots, y_p 的非零解, 记为 $(l_{i1} \ l_{i2} \ l_{i3} \ \cdots \ l_{ip})$, 该向量称为相关矩阵 R 对于特征值 λ_i 的特征向量。

4. 求各主成分的权重系数: 用 L 矩阵元素 (l_{ij}) 乘以相应特征根的平方根, 便得第一式中的主成分系数, 即用第 i 个特征值的平方根 $(\lambda_i)^{1/2}$ 乘以上式的第 i 列的各元素。

$$w_{ij} = l_{ij} (\lambda_i)^{1/2}$$

由此, 便完成了由原始变量到新综合变量 (主成分) 的数据转换。如果取 m ($m \leq p$) 个主成分, 依次将所有样品的原始变量数据代入方程组, 就得到各样品的各主成分值, 即主成分得分。

二、主成分的计算步骤

例如冠心病研究中考虑了三个变量 ($p=3$)， x_1 为收缩压 (Kpa)， x_2 为血浆胆固醇浓度 (mmol/L)， x_3 为研究对象的年龄。样本包含 15 个正常人及 10 个冠心病患者，其资料示于表 5-7 中。

表 5-7 冠心病研究数据

代码	类别	血压	胆固醇	年龄
NAME	GRP	X1	X2	X3
A	N	18.0	5.90	45
B	N	16.2	5.93	41
C	N	17.3	5.69	49
D	N	19.7	6.37	52
E	N	19.4	5.80	54
F	N	17.2	5.59	47
G	N	18.9	6.37	60
H	N	19.6	5.98	48
I	N	19.1	5.98	44
J	N	18.0	5.85	64
K	N	18.4	5.77	59
L	N	18.6	6.24	55
M	N	18.4	6.40	54
N	N	18.6	6.24	56
O	N	17.8	5.72	50
P	Y	20.7	6.60	60
Q	Y	21.5	6.84	64
R	Y	19.4	6.42	54
S	Y	19.8	6.50	48
T	Y	23.9	6.97	43
U	Y	20.0	6.45	43
V	Y	21.4	7.12	63
W	Y	22.6	7.28	64
X	Y	20.2	7.07	62
Y	Y	21.8	6.86	65

计算步骤：

1. 数据标准化

计算各变量的均值、标准差，并将原始数据进行标准化处理。本例标准

化后的数据见表 5-8:

表 5-8 标准化的冠心病研究数据

代码	类别	x1	x2	x3
A	N	-0.83344	-0.83767	-1.13702
B	N	-1.79058	-0.78532	-1.65621
C	N	-1.20157	-1.25650	-0.61783
D	N	0.12369	0.10471	-0.22844
E	N	-0.02356	-1.04709	0.03115
F	N	-1.27520	-1.46592	-0.87743
G	N	-0.31806	0.10471	0.80993
H	N	0.05007	-0.68061	-0.74763
I	N	-0.17081	-0.68061	-1.26682
J	N	-0.83344	-0.94238	1.32912
K	N	-0.61257	-1.09944	0.68014
L	N	-0.46531	-0.15706	0.16095
M	N	-0.61257	0.15706	0.03115
N	N	-0.46531	-0.15706	0.29075
O	N	-0.90707	-1.20415	-0.48804
P	Y	0.71270	0.57590	0.80993
Q	Y	1.15445	1.04709	1.32912
R	Y	-0.02356	0.20942	0.03115
S	Y	0.19732	0.36648	-0.74763
T	Y	2.47971	1.30886	-1.39662
U	Y	0.27094	0.26177	-1.39662
V	Y	1.08082	1.62298	1.19933
W	Y	1.74346	1.93711	1.32912
X	Y	0.41819	1.51828	1.06953
Y	Y	1.30170	1.09944	1.45892

2. 求相关矩阵 R

按下式计算各变量间的相关系数

$$r_{kj} = \frac{1}{(n-1)} \sum_{s=1}^n x_{sk} x_{sj} \quad k, j=1, 2, \dots, p$$

$$s=1$$

其中 n 为样本含量, 本例 $n=25$

x_{sk} , x_{sj} 分别为第 s 个样本点的第 k 变量与第 j 变量标准化后的观测值。于是可得到相关矩阵。

3. 求相关矩阵 R 的特征值

解特征方程 $|R-\lambda I|=0$ 的计算过程复杂, 一般由计算机程序算得。本例所求矩阵 R 的特征值、贡献率及累积贡献率于表 5-9。

表 5-9 特征值、贡献率及累积贡献率

i	特征值	贡献率	累积贡献率
λ_i	p_i	C_m	
1	2.147076	0.7157	0.7157
2	0.713138	0.2377	0.9534
3	0.139786	0.0466	1.0000

4. 选取主成分个数 (m)

由表 5-9 看出, 三个特征值中 $\lambda_1=2.147076$, 其贡献率 $p_1=2.147076/3=0.7157$, 说明第一主成分反映了原数 71.57% 的信息。第二主成分虽然小于 1, 但结合前两个主成分的累积贡献率达 0.9534%, 说明前两个主成分已经反映了原数据的绝大部分信息, 故此确定取两个主成分为宜。

5. 计算相应特征值的特征向量矩阵 (L)

计算得相应特征值的标准化特征向量矩阵。

6. 求因子载荷矩阵 (A)

求各主成分的权重系数或因子载荷矩阵, 第一主成分与变量 x_1 (血压)、 x_2 (胆固醇水平) 相关密切; 而第二主成分与 x_3 (年龄) 有关。由此, 我们可以得到本例由原始变量 x_1 、 x_2 和 x_3 到两个综合指标 z_1 与 z_2 的转换式。

7. 标准化主成分得分 (standard principal component score) 计算

经过以上分析已完成由三个原始变量到二个新的综合变量的转换, z_1 和 z_2 的均值为零, 但方差并不为 1。为了统一标度, 对右侧系数按下式进行一次标准化转换

$$b_{ij}=a_{ij}/(\lambda_i)^{1/2}$$

其中 b_{ij} 为第 i 主成分第 j 变量的标准化得分系数。获得 25 个对象的两个主成分的得分值, 见表 5-10:

表 5-10 标准化主成分得分

代码 ID	组别 GRP	血压 x1	胆固醇 x2	年龄 x3	主成分 1 z1	主成分 2 z2
A	N	-0.83344	-0.83767	-1.13702	-1.07097	-0.55285
B	N	-1.79058	-0.78532	-1.65621	-1.61123	-0.63634
C	N	-1.20157	-1.25650	-0.61783	-1.24845	0.28157
D	N	0.12369	0.10471	-0.22844	0.02706	-0.32714
E	N	-0.02356	-1.04709	0.03115	-0.46013	0.32205
F	N	-1.27520	-1.46592	-0.87743	-1.45193	0.10263
G	N	-0.31806	0.10471	0.80993	0.16373	0.97411
H	N	0.05007	-0.68061	-0.74763	-0.50994	-0.62406
I	N	-0.17081	-0.68061	-1.26682	-0.76387	-1.05680
J	N	-0.83344	-0.94238	1.32912	-0.35156	2.04792
K	N	-0.61257	-1.09944	0.68014	-0.52916	1.30357
L	N	-0.46531	-0.15706	0.16095	-0.21453	0.43912
M	N	-0.61257	0.15706	0.03115	-0.17871	0.29295
N	N	-0.46531	-0.15706	0.29075	-0.17424	0.57454
O	N	-0.90707	-1.20415	-0.48804	-1.06144	0.25784
P	Y	0.71270	0.57590	0.80993	0.80375	0.34063
Q	Y	1.15445	1.04709	1.32912	1.35743	0.53936
R	Y	-0.02356	0.20942	0.03115	0.09175	-0.01146
S	Y	0.19732	0.36648	-0.74763	0.01183	-0.97462
T	Y	2.47971	1.30886	-1.39662	1.18323	-3.02762
U	Y	0.27094	-0.26177	-1.39662	-0.20465	-1.66024
V	Y	1.08082	1.62298	1.19933	1.53916	0.28739
W	Y	1.74346	1.93711	1.32912	1.99581	0.01260
X	Y	0.41819	1.51828	1.06953	1.17449	0.50660
Y	Y	1.30170	1.09944	1.45892	1.48258	0.58825

附主成分分析的 SAS 程序

```
DATA DONG.BOOK;
```

```
INPUT NAME $ GRP $ SBP CHOL AGE @@;
```

```
CARDS;
```

```
A N 135 227 45 N N 140 240 56 G N 142 245 60 T Y 180 268 43
```

```
B N 122 228 41 O N 134 220 50 H N 147 230 48 U Y 150 248 43
```

```
C N 130 219 49 P Y 156 254 60 I N 144 230 44 V Y 161 274 63
```

```
D N 148 245 52 Q Y 162 263 64 J N 135 225 64 W Y 170 280 64
```

```
E N 146 223 54 R Y 146 247 54 K N 138 222 59 X Y 152 272 62  
F N 129 215 47 S Y 149 250 48 L N 140 240 55 Y Y 164 264 65  
M N 138 246 54
```

;

```
PROC PRINCOMP DATA=DONG. BOOK PREFIX=PRIN COV;
```

```
VAR AGE SBP CHOL;
```

```
RUN;
```

第五节 聚类分析

人类认识自然的一种很重要的方法是对事物进行分类，有分类才有科学。随着科学技术的发展分类越来越细，而且单凭经验与专业知识有时也不能确切地将事物分类。近二十年来计算机技术和多元统计的发展，数值分类学逐渐形成一个新的分支，称为聚类分析（clustering analysis），又称群落分析、群分析、点群分析、分支分析。与其他多元统计方法相比，聚类分析的方法比较粗糙，理论尚不完善，但由于它的应用取得了很大的成功，故此与回归分析、判别分析称为多元分析的三大方法。

所谓类是指具有相似性的物体的集合。聚类就是以研究对象（样品或变量）间存在着不同程度的相似性（亲疏关系）为依据，把一些相似程度较大的样品（或指标）聚合为一类，再把另一些彼此相似程度较小的样品（或指标）又聚合为另一类，如此反复，使亲近的聚合到一个小的分类单位，疏远的聚合到一个大的分类单位，直到把所有样品（或指标）都聚合完毕为止，形成一个由亲近到疏远，由小到大的分类系统，最后把整个分类用一张聚类树图（又称分群图、谱系图）把样品（或指标）间的亲疏关系表达出来。

应用聚类分析方法进行分类与判别分析的分类不同。判别分析必须事先知道待判别的类型，并且要有一批来自已知类别的样品，根据这批已知类别的样品来建立判别函数式，然后用此判别函数式对未知类别的新样品进行判别归类。聚类分析对其研究对象的类别是未知的，甚至分类个数也一无所知，而需要根据事物间的各种数量结构来寻找合适的分类。所以聚类分析一般是

寻找处理客观事物分类的方法，而判别分析是在分类方法已知的条件下，确定某一样品属于何类。

聚类分析根据分类对象不同有 Q-型聚类分析（即对样品分类）和 R-型聚类分析（对指标分类）。聚类方法更是丰富多彩，大致可归纳为：系统聚类法（谱系聚类法）、动态聚类法、模糊聚类法、图论聚类法、聚类预报法等。因为少数民族遗传学亲缘关系分析大多应用系统聚类法，故这里仅简单介绍系统聚类法。

一、聚类统计量

在进行聚类分析时，首先要有一个能表示样品或变量间相似程度的统计量。聚类的统计量可多达数百种，常用的聚类统计量大致分三类：距离、相似系数与关联系数。

1. 距离 (distance)

设作聚类分析的多变量数据有 n 个样品（个体）及 p 个变量。把每个样品看作 p 维空间的一个点，聚类就是根据这 n 个点在此空间的距离把样品聚合为若干类。这样自然想到二维平面中两个点的距离拓展到 P 维空间的距离。给定两个个体 i 和 j ，最常用的距离有：

(1) 明考斯基距离 (Minkowski' s distance, 简称明氏距离)：

$$d_{ij}(q) = \left(\sum_{k=1}^p |A_{ik} - A_{jk}|^q \right)^{\frac{1}{q}}$$

① 当 $q=1$ 时，

$$d_{ij}(1) = \sum_{k=1}^p |x_{ik} - x_{jk}| \text{ 即为绝对值距离 (absolute distance);}$$

② 当 $q=2$ 时，

$$d_{ij}(2) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \text{ 即为欧几里得距离 (Euclidian' s distance, 简称欧氏距离);}$$

③ 当 $q=3$ 时，

$$d_{ij}(\infty) = \max_{1 \leq k < \pi} |x_{ik} - x_{jk}| \text{ 即为切比雪夫距离 (Chebyshev' s distance)。$$

这些距离指标以欧氏距离最为通用。明氏距离在应用上存在着不足之处：其一，P个变量的量纲不同，即测量单位不一样，使得计算距离时对每个变量的权重不等。克服其不足的办法是在计算距离之前，将数据进行标准化或正规化处理，其处理方法将在后面讨论；其二是没有考虑变量之间的相关性，使得所计算的距离可能出现超度或低估。为了克服这一问题，可采用马氏距离。

(2) 马氏距离 (Mahalanobis' distance)

$$d^2_{ij} (m) = (X_i - X_j)' \text{COV}^{-1} (X_i - X_j)$$

其中：

X_i 为第 i 样品的观测值向量， X_j 为第 j 样品的观测值向量

COV^{-1} 为向量 X_i 与向量 X_j 的协方差矩阵的逆矩阵。

马氏距离可以克服各指标间的相关性，而且不受各指标量纲的影响，但计算复杂，手算很难实现，多采用计算机程序计算。

2. 相似系数

相似性系数属于另一类聚类统计量，一般用 C_{ij} 表示第 i 与第 j 样品的相似程度。常用的相似系数有三种：

(1) 夹角余弦

把任何两个样品 x_i 、 x_j 看成是 P 维空间的两个向量，其间存在一个夹角 (α)，而该夹角的余弦定义为：

$$C_{ij} (1) = \text{COS} (\alpha_{ij})$$

$$\sum_{k=1}^P x_{ik} \times x_{jk}$$

其中 $C_{ij} (1)$ 为第 i 个体与第 j 个体间的夹角余弦；

x_{ik} 与 x_{jk} 分别为第 i 个体与第 j 个体在第 k 变量上的取值；

$\sum_{k=1}^P$ 表示对变量 k 求和。

(2) 相关系数

相关系数通常用 r_{ij} 表示，但实际上是将数据标准化后的角余弦。在此为和其他系数比较，用 $C_{ij} (2)$ 表示。其定义为：

$$r_{ij} = c_{ij}(2) = \frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^p (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^p (x_{jk} - \bar{x}_j)^2}}$$

其中 $C_{ij}(2)$ 为第 i 个体与第 j 个体间的相关系数:

$$x_i = (1/n) \sum_{k=1}^p x_{ik}$$

$$x_j = (1/n) \sum_{k=1}^p x_{jk}$$

x_{ik} 与 x_{jk} 分别为第 i 个体与第 j 个体在第 k 变量上的取值;

$\sum_{k=1}^p$ 表示对变量 k 求和。

$C_{ij}(2)$ 实际上是两向量 $(x_{ik} - x_i)$ 、 $(x_{jk} - x_j)$ 的夹角余弦。其取值范围为 $(-1 \leq C_{ij}(2) \leq 1)$ 。因聚类分析只关注其相关的强弱,即事物间的相似性,而不注意相关的方向。故此只取相关的绝对值作为聚类统计量。但是,相关系数亦存在不足之处。当某一样品的观测指标值与另一样品的观测指标值成比例时,则无论它们的差异有多大,其相关系数将接近于 1,显而易见是不合理的。因此,通常只用它作为 R 型(指标)聚类统计量。

(3) 指数相似系数

由于各指标变量间的量纲差异,可以定义第 i 样品与第 j 样品的指数相似性系数为:

$$C_{ij}(3) = (1/p) \sum_{k=1}^p \exp \left[(-3/4) (x_{ik} - x_{jk})^2 / S_{2k} \right]$$

其中 $C_{ij}(3)$ 为第 i 个体与第 j 个体间的指数相似性系数

x_{ik} 与 x_{jk} 分别为第 i 个体与第 j 个体在第 k 变量上的取值:

$\sum_{k=1}^p$ 表示对变量 k 求和:

S_{2k} 为第 k 个指标变量的方差。

这三种相似性系数具有如下特点:

①任意点间相似系数均满足 $C_{ij} \leq 1$ 的特征。

②相似性系数和距离相反, C_{ij} 愈大表示二者愈相近。实际上距离与相似性系数是可以进行相互转换的,若 d_{ij} 为标准化距离,则有关系 $d_{ij} = 1 - C_{ij}$ 。

③如有两个样品（或指标）完全相等，则有 $C_{ij}=1$ ，自身相似性为 1，即 $C_{ii}=1$ 。

④与距离相似亦有 $C_{ij}=C_{ji}$ 。

3. 关联系数

上述距离与相似性系数可用于连续变量资料的聚类分析指标，既可用于样品聚类，也可用于指标变量的聚类。当变量只能定性描述时，要用距离，相似系数就有一定的困难，而采用关联系数作为聚类指标。

广义关联性的含义是用来描述定性数据之间的关系，而关联系数是关联强度的指标。多态变量之间的关联常用列联表来表示，而其关联度的测量指标有：

$$G_{ij} (1) = \chi^2$$

即 χ^2 统计量本身就是一个关联系数。其他一些与 χ^2 统计量有关的关联系数有：

$$G_{ij} (2) = [\chi^2 / (\chi^2 + n..)]^{1/2}$$

$$G_{ij} (3) = [\chi^2 / (n.. \times \max (R-1, C-1))]^{1/2}$$

$$G_{ij} (4) = [\chi^2 / (n.. \times \min (R-1, C-1))]^{1/2}$$

$$G_{ij} (5) = \{\chi^2 / [n.. \times [(R-1)(C-1)]^{1/2}\}^2$$

其中 χ^2 为卡方统计量；
 $n..$ 为行列表的总样本数；
 R 为行列表的行数；
 C 为行列表的列数；

除了以上三类用作聚类分析的指标外尚有很多聚类指标。一般而言，对于样品（Q 型）聚类多采用距离，而对变量（R 型）聚类多采用相似性系数，更多采用相关系数。

二、数据变换

对于多变量计量数据，由于所研究的各指标变量的测量单位与量级可能不同，即便有些变量的度量一样，但各变量的绝对值大小也不一样。因此，直接用原始数据计算距离就会突出那些绝对值大的变量而压低那些绝对值小的变量的作用。因此，一般在计算距离前需对原始数据进行预处理，即对数

据进行变换。

1. 标准差标准化

假定对第 k 个变量进行标准化, 即用如下公式变换:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{s_k}$$

$$i=1, 2, \dots, n \quad k=1, 2, \dots, p$$

其中 x_{ik}^* 为第 k 变量第 i 个观测值的标准化得分, 即标准离差值;

x_{ik} 为原始数据中第 k 变量第 i 个观测值;

\bar{x}_k 为第 k 变量的均值;

s_k 为第 k 变量的标准差。

通过标准差标准化后的数据 (X_{ik}^*) 中每个变量的均值为 0, 标准差为 1, 从而消除了原变量量纲的影响。

2. 极差标准化

将变量 x_{ik} 作如下变换:

$$x_{ik}^{\#} = \frac{x_{ik} - \bar{x}_k}{R_k}$$

其中 $x_{ik}^{\#}$ 为第 i 变量第 k 个观测值的极差标准化得分;

R_k 为第 k 变量的全距 (极差)。即

$$R_k = \max(x_k) - \min(x_k)$$

经变换后的 $x_{ik}^{\#}$ 的极差为 1, 而且变换后的取值范围为 $-1 \leq x_{ik}^{\#} \leq 1$ 。从而消除其量纲的差异影响。

3. 极差正规化

将变量 x_{ik} 作如下变换:

$$x_{ik}^{\#} = \frac{x_{ik} - \min(x_k)}{R_k}$$

其中 $x_{ik}^{\#}$ 为第 k 变量第 i 个观测值的极差正规化得分

R_k 为第 k 变量的全距 (极差)。即

$$R_k = \max(x_k) - \min(x_k)$$

经变换后的 $x_{ik}^{\#}$ 的极差为 1, 而且变换后的取值范围为 $0 \leq x_{ik}^{\#} \leq 1$ 。从而消除其量纲的差异影响。

三、系统聚类法

1. 系统聚类法的基本思想与合并归类原则

(1) 基本思想

系统聚类法 (systematic cluster method), 又称谱系聚类法 (hierarchical cluster method), 是实际工作中使用最多的一种方法。其基本思想是: 首先将每个样品各看成一类, 这时类与类间的距离即是样品间的距离。然后将类间距离最近的两类合并成一类, 重新计算该新类与其他各类之间的距离; 再将类间距离最近的两类又合并成另一新类, 又重新计算该新类与其他各类之间的距离; 如此反复进行, 每次合并便减少一类, 直到所有样品都合并成一个大类为止。最后将整个聚类过程做成聚类树图, 按聚类树图特征选恰当的分类。

实际上聚类只有二个基本过程: 首先合并归类, 然后重新计算类间距离。

(2) 合并归类遵循的原则

①若选出一对样品在已经分好的类中都未出现过, 则把它们形成一个独立的新类;

②若选出的两个样品中, 有一个是在已经分好的类中出现过, 则把另一样品合并到该类中;

③若选出的两个样品, 都分别出现在已经分好的两类中, 则把这两类合并成一新的大类;

④若选出一对样品都出现在已分好的同一类中, 则这对样品就不用再归并了。

2. 最短距离法 (nearest neighbor method)

用 d_{ij} 表示样品 X_i 与 X_j 的距离, 用 G_1, G_2, \dots 表示各类, 最短距离法定义两类之间距离为两类间所有样品中最近的两个样品距离。用 d_{pq} 来表示 G_p 类和 G_q 类的距离, 则:

$$d_{pq} = \min (d_{ij})$$

其中 i 为属于 G_p 类的样品, j 为属于 G_q 类的样品, $\min (d_{ij})$ 表示所有样品间的最小距离。最短距离聚类法的基本步骤:

(1) 有必要则对原始数据进行适当的转换。

(2) 确定聚类统计量。可以是欧氏距离、绝对距离、切比雪夫距离，也可以用马氏距离，也可以用 $1 - c_{ij}$ ，其中 c_{ij} 表示前面所述的各种相似系数；自然也可以用各种关联系数。总之，获得一个描述样品（指标）间的亲疏关系的对称矩阵。

(3) 按系统聚类思想，首先将各样品看成自成一类，即每一样品便为一个类。假定有 6 个样品，使它们各成一类分别用 $G_1, G_2 \dots G_6$ 表示，类与类间距离示于 $D(0)$ 表的下三角部分，其中， (0) 表示第 0 次合类的距离表。

D (0)					
类别	G1	G2	G3	G4	G5
G2	1				
G3	4	3			
G4	6	5	2		
G5	8	7	4	2	
G6	9	8	5	3	1

(4) 寻找 $D(0)$ 中的最小距离进行合并。因 G_1 与 G_2 的距离为 1， G_5 与 G_6 的距离为 1，是最小的两个距离。合并 G_1, G_2 为一新类 G_7 ，即 $G_7 = (G_1, G_2)$ ； G_5 与 G_6 为新类 G_8 ，即 $G_8 = (G_5, G_6)$ 。

(5) 按最短距离计算新类 G_7, G_8 与其他各类的距离：

$$d_{7,3} = \min(d_{1,3}, d_{2,3}) = \min(4, 3) = 3$$

$$d_{7,4} = \min(d_{1,4}, d_{2,4}) = \min(6, 5) = 5$$

$$d_{7,8} = \min(d_{1,8}, d_{2,8})$$

$$= \min[\min(d_{1,5}, d_{1,6}), \min(d_{2,5}, d_{2,6})] = \min(8, 7)$$

$$= 7$$

$$d_{8,3} = \min(d_{5,3}, d_{6,3}) = \min(4, 5) = 4$$

$$d_{8,4} = \min(d_{5,4}, d_{6,4}) = \min(2, 3) = 2$$

于是得到距离表 $D(1)$

D (1)

类别	G7	G3	G4
G3	3		
G4	5	2	
G8	7	4	2

(6) 对表 D (1) 重复以上 (4) 和 (5) 的步骤。合并 G3、G4、G8 为新类 G9:

$$G9 = (G3, G4, G8) = (G3, G4, G5, G6)$$

计算 G9 与 G7 的距离

$$d_{9,7} = \min(d_{3,7}, d_{4,7}, d_{8,7}) = 3$$

得新的距离表 D (2)

D (2)

类别	G7
G9	3

(7) 合并 G9、G7 为新类 G10 聚类完毕。

(8) 综合以上过程，做聚类树图如下:

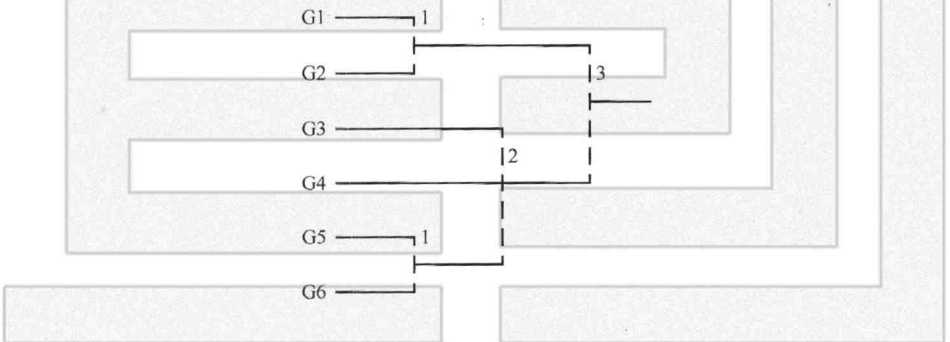


图 5-1 最短距离聚类法树图

3. 最长距离法 (farthers tneighbor method)

如果在合并新类后，定义新类与其他类间距离用两类内成员间的最长距离，即

$$D_{pq} = \max(d_{ij})$$

其中 i 为属于 G_p 类的样品，j 为属于 G_q 类的样品，max (d_{ij}) 表示所有样品间的最大距离。

最长距离法的聚类过程与原则和最短距离法基本相同，唯一的不同在于定义各类间距离时，一个选择最短距离，一个选择最长距离。

仍以前 D (0) 数据为例，将两个最小距离合并归类为：G7 = (G1, G2), G8 = (G5, G6)。按最长距离定义新类与其他类间距离为：

$$d_{7,3} = \max(d_{1,3}, d_{2,3}) = \max(4, 3) = 4$$

$$d_{7,4} = \max(d_{1,4}, d_{2,4}) = \max(6, 5) = 6$$

$$d_{7,8} = \max(d_{1,8}, d_{2,8})$$

$$= \max[\max(d_{1,5}, d_{1,6}), \max(d_{2,5}, d_{2,6})]$$

$$= \max[\max(8, 9), \max(7, 8)] = \max(9, 8) = 9$$

$$d_{8,3} = \max(d_{5,3}, d_{6,3}) = \max(4, 5) = 5$$

$$d_{8,4} = \max(d_{5,4}, d_{6,4}) = \max(2, 3) = 3$$

归纳各新的类间距离 D (1) 为：

D (1)

类别	G7	G3	G4
G3	4		
G4	6	2	
G8	9	5	3

合并 G3, G4 为 G9, 即 G9 = (G3, G4)。

按同样方式用最长距离定义新的类间距离：

$$d_{9,7} = \max(d_{3,7}, d_{4,7}) = \max(4, 6) = 6$$

$$d_{9,8} = \max(d_{3,8}, d_{4,8}) = \max(5, 3) = 5$$

得新的距离 D (2) 为：

D (2)

类别	G9	G7
G7	6	
G8	5	9

合并 G8, G9 为新类 G10, 即 G10 = (G8, G9)。计算新的类间距离：

$$d_{10,7} = \max(d_{8,7}, G_9, 7) \\ = \max(9, 6) = 9$$

得新的距离阵 D (3):

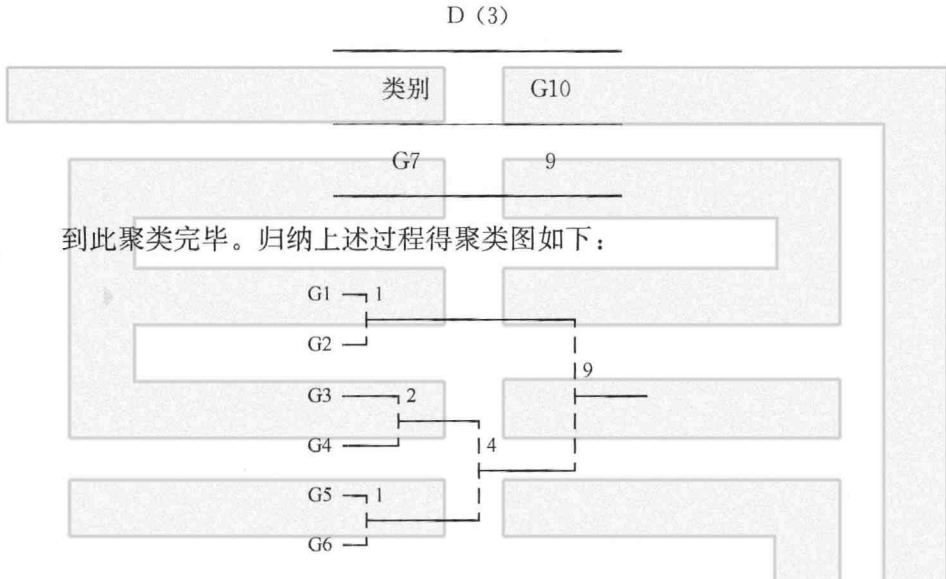


图 5-2 最长距离法聚类树图

4. 其他系统聚类法

除了常用的最短距离法、最长距离法之外，还有以下几种定义合并新类与其他类间距离的方法。这些距离的计算都较前二者复杂。在每一种聚类过程中，除了类间距离的定义不同外，其他方法与步骤与前二者完全一致。

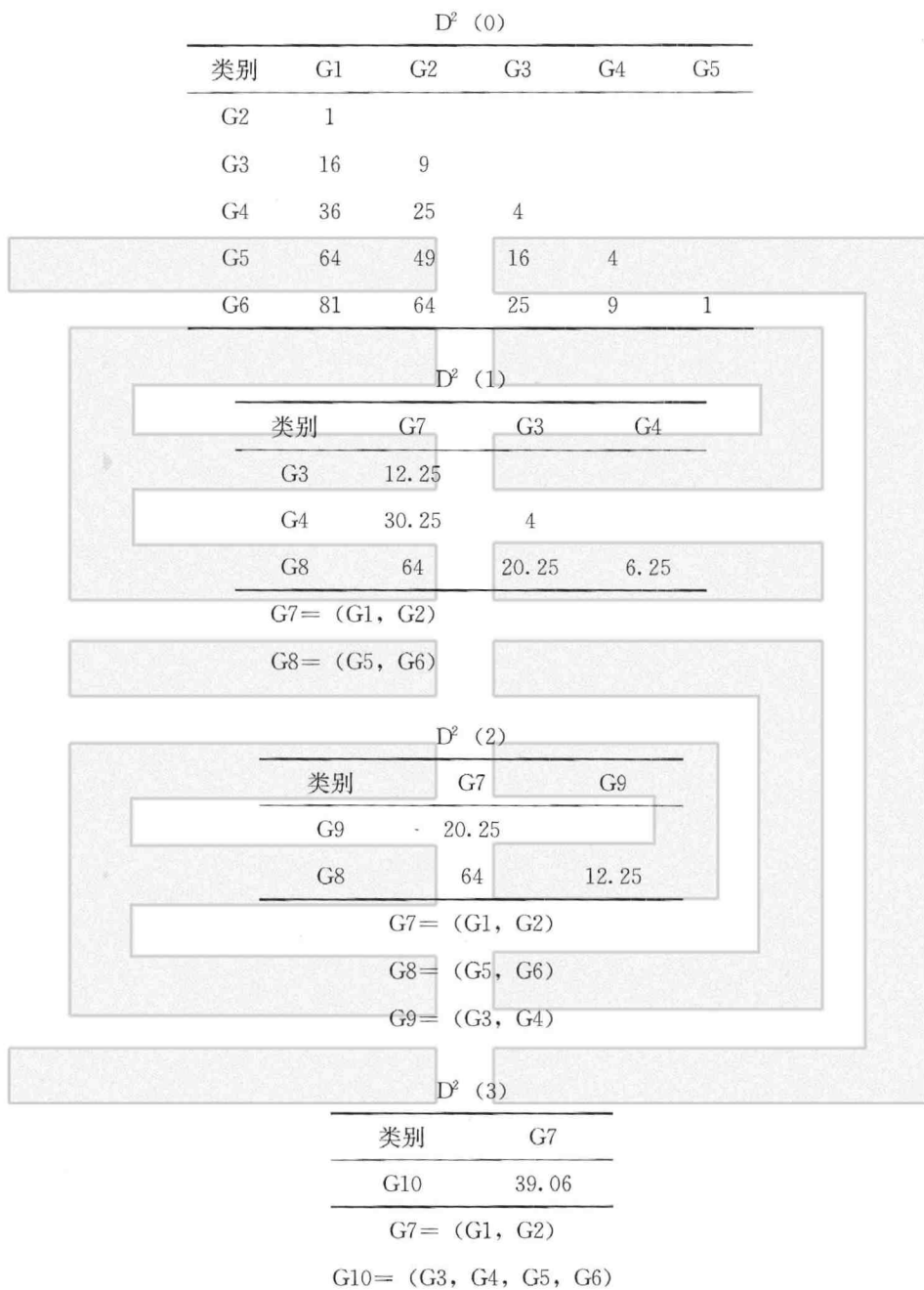
(1) 中间距离法

设 G_p 和 G_q 两类为最近合并为一新类 G_r ，即 $G_r = (G_p, G_q)$ ，用中间距离法定义 G_r 与其他类 G_k 的距离为：

$$d^2_{rk} = (1/2) d^2_{kp} + (1/2) d^2_{kq} - (1/4) d^2_{pq}$$

注意：中间距离法定义公式用的是距离的平方，因此，聚类过程中产生的距离矩阵表相应为距离平方表 $d^2(0)$ ， $d^2(1)$ ， $d^2(2)$ … 仍以前 $D(0)$ 距离表为例，用中间距离法计算出各步的 $d^2(i)$ 如下表。

按中间距离法获得各次 $D(i)$ 矩阵如下，在此省略了其计算过程。



综合各表得聚类树图 (图 5-3)。

(2) 重心法

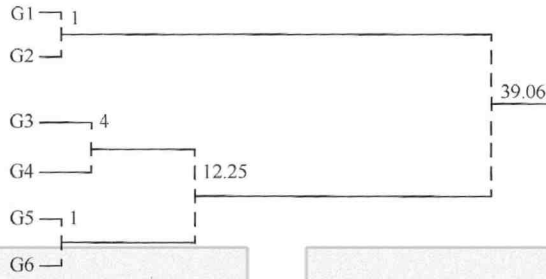


图 5-3 中间距离法的聚类树图

从物理学的角度，两类之间的距离用重心代表比较合理。若 G_p, G_q 两类距离最近，则合并为一新类 $G_r, G_r = (G_p, G_q)$ 。则 G_r 与其他类 G_k 的距离按重心法定义为：

$$d^2_{rk} = (n_p/n_r) d^2_{kp} + (n_q/n_r) d^2_{kq} - (n_p/n_r)(n_q/n_r) d^2_{pq}$$

其中 n_p, n_q 代表 G_p 与 G_q 类的样品数， $n_r = n_p + n_q$ 。

仍以前 $D(0)$ 距离矩阵为例，用重心法聚类所产生各步的 $D(i)$ 表如下：

$D^2(1)$

类别	G7	G3	G4
G3	12.25		
G4	30.25	4.00	
G8	64.00	20.25	6.25

$G7 = (G1, G2)$
 $G8 = (G5, G6)$

$D^2(2)$

类别	G7	G9
G9	20.25	
G8	64.00	12.25

$G7 = (G1, G2)$
 $G8 = (G5, G6)$
 $G9 = (G3, G4)$

$D^2(3)$

类别	G7
G1	039.06

$G7 = (G1, G2)$

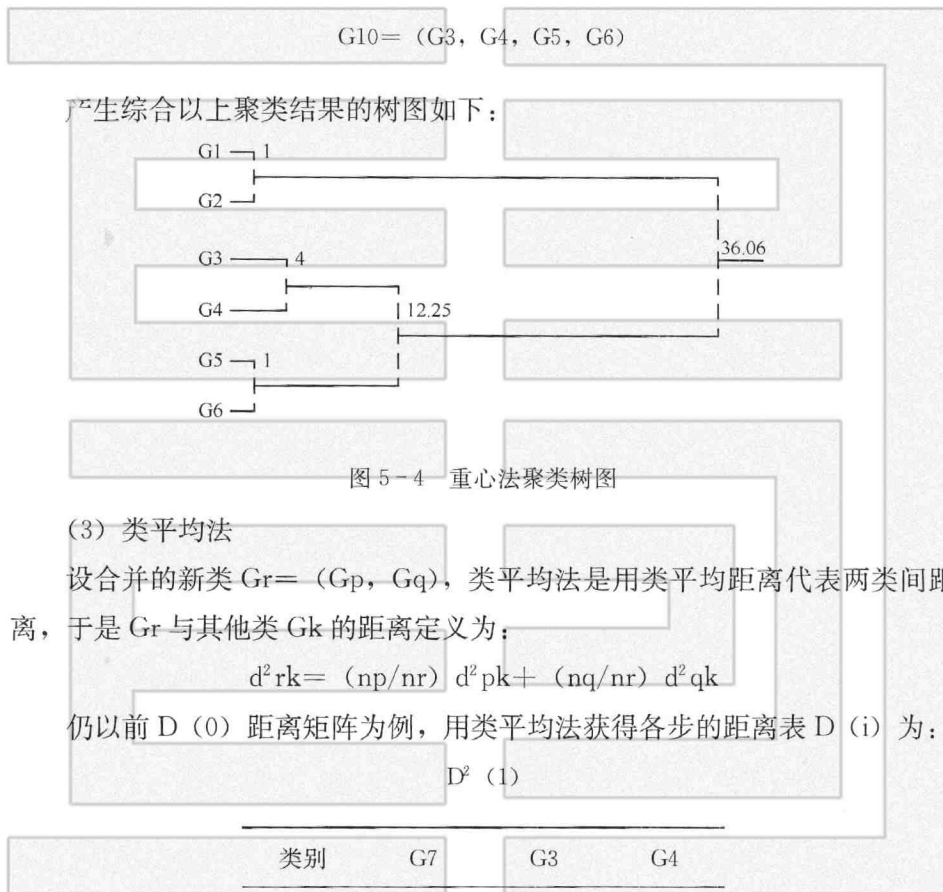


图 5-4 重心法聚类树图

(3) 类平均法

设合并的新类 $G_r = (G_p, G_q)$ ，类平均法是用类平均距离代表两类间距离，于是 G_r 与其他类 G_k 的距离定义为：

$$d^2_{rk} = (n_p/n_r) d^2_{pk} + (n_q/n_r) d^2_{qk}$$

仍以前 $D(0)$ 距离矩阵为例，用类平均法获得各步的距离表 $D(i)$ 为：

$D^2(1)$

类别	G7	G3	G4
G3	12.5		
G4	30.5	4	
G8	64.5	20.5	6.5

$G7 = (G1, G2)$

$G8 = (G5, G6)$

$$D^2 (2)$$

类别	G7	G9
G9	21.5	
G8	64.5	13.5

$$G7 = (G1, G2)$$

$$G8 = (G5, G6)$$

$$G9 = (G3, G4)$$

$$D^2 (3)$$

类别	G7
G10	43

$$G7 = (G1, G2)$$

$$G10 = (G3, G4, G5, G6)$$

综合以上过程的聚类图如下：

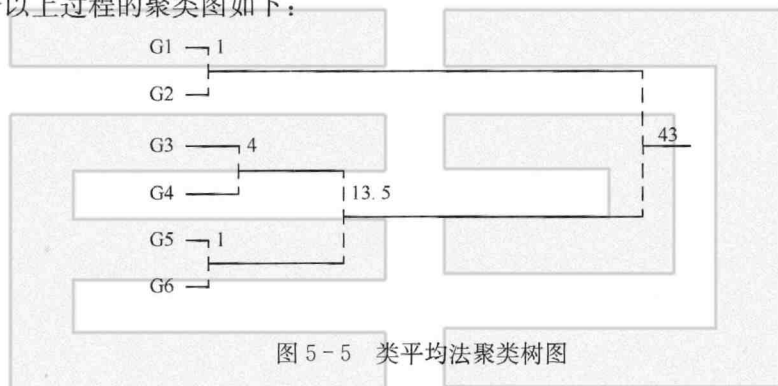


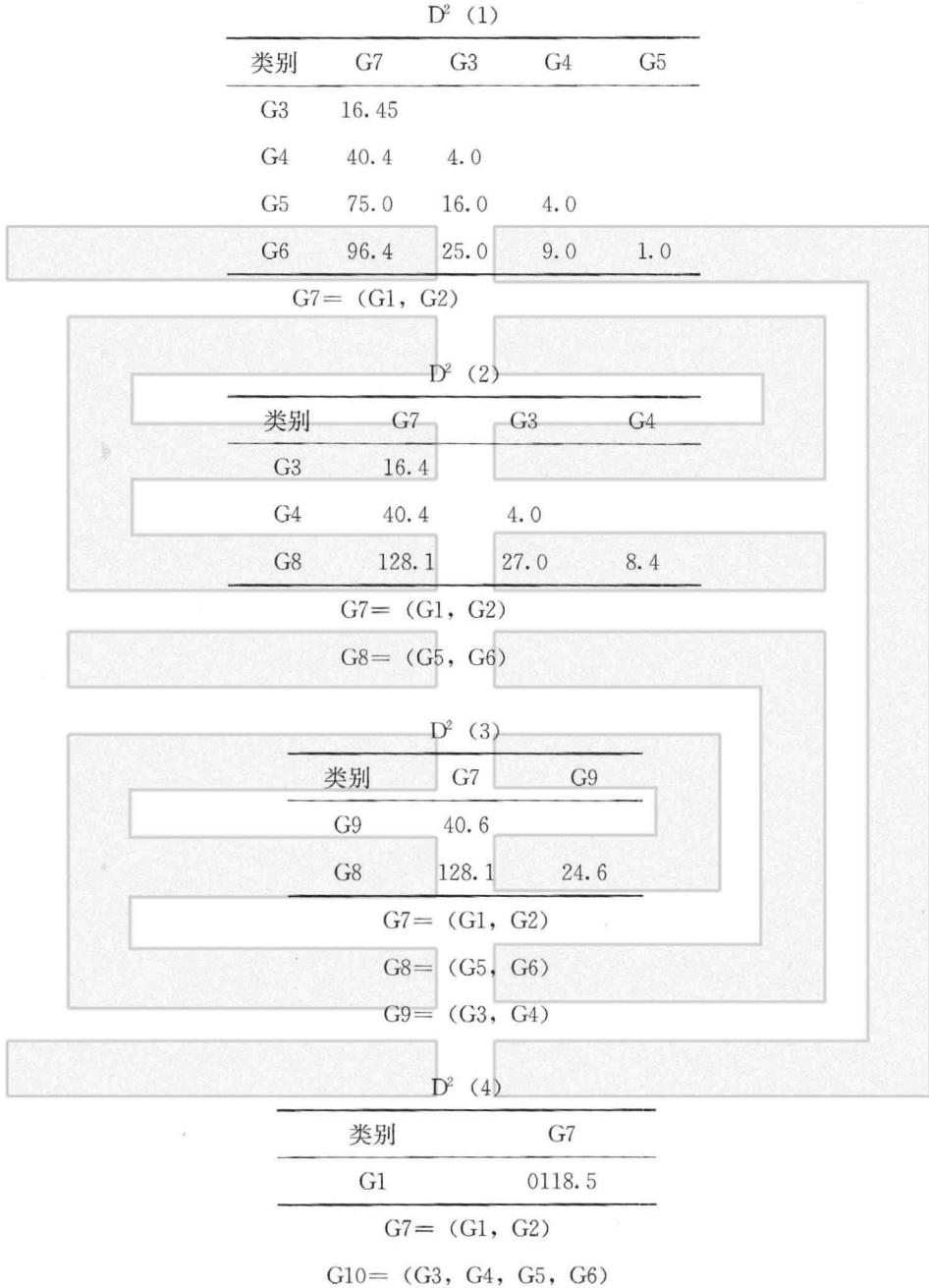
图 5-5 类平均法聚类树图

(4) 离差平方和法

本法把方差分析的思想用于定义类间距离，它采用同类样品的离差平方和最小，而类与类之间的离差平方和最大的原则。当 $G_r = (G_p, G_q)$ 为新合并类， G_k 表示其他类，则用离差平方和法定义类间距离为：

$$d^2_{rk} = \left[\frac{nk + np}{nr + nk} \right] d^2_{pk} + \left[\frac{nk + nq}{nr + nk} \right] d^2_{qk} - \left[\frac{nk}{nr + nk} \right] d^2_{pq}$$

仍以以前 $D(0)$ 距离矩阵为例，用离差平方和法获得各步的距离表 $D(i)$ 为：



综合以上过程的聚类图如下：

比较以上用同一距离矩阵数据 $D(0)$ 表，采用不同方法的聚类结果，最

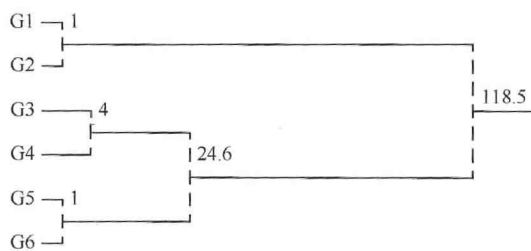


图 5-6 离差平方和法聚类树图

短距离法与其他五种方法的结果不一致，而后五种方法对于本例的结果基本一致。这并不意味后五种方法对任何数据的聚类结果都是一致的。

除以上介绍的六种系统聚类法外，尚有可变法、可变类平均法等。实际上，以上类间距离的不同定义方法可用如下统一的公式来表达：

$$D^2_{rk} = \alpha_p D^2_{pk} + \alpha_q D^2_{qk} + \beta D^2_{pq} + \gamma | D^2_{pk} - D^2_{qk} |$$

其中 D^2_{rk} 为合并的新类 $G_r = (G_p, G_q)$ 与其他类 G_k 的距离。

α_p 、 α_q 、 β 、 γ 分别为不同距离定义待定系数，这些系数的取值见表 5-11：

表 5-11 不同聚类方法的系数值

方 法	α_p	α_q	β	γ
最短距离法	0.5	0.5	0	-0.5
最长距离法	0.5	0.5	0	0.5
中间距离法	0.5	0.5	-0.25	0
重心法	np/nr	nq/nr	$-\alpha_p \alpha_q$	0
离差平方和法	$(nk+np)/(nk+nr)$	$(nk+nq)/(nk+nr)$	$-nk/(nr+nk)$	0
类平均法	np/nr	nq/nr	0	0
可变法	$(1-\beta)/2$	$(1-\beta)/2$	<1	0
可变类平均法	$(1-\beta) np/nr$	$(1-\beta) nq/nr$	<1	0

例如最短距离法用统一公式推算为：

$$D_{rk} = (0.5D^2_{pk} + 0.5D^2_{qk} - 0.5 | D^2_{pk} - D^2_{qk} |)^{1/2}$$

若 $D_{pk} > D_{qk}$ 则：

$$D_{rk} = [0.5D^2_{pk} + 0.5D^2_{qk} - 0.5 (D^2_{pk} - D^2_{qk})]^{1/2}$$

$$= (D^2 qk) 1/2 = Dqk \quad \text{这就是取最短距离；}$$

若 $Dpk < Dqk$ 则：

$$Drk = [0.5D^2 pk + 0.5D^2 qk - 0.5(D^2 qk - D^2 pk)] 1/2$$

$$= (D^2 pk) 1/2 = Dpk \quad \text{这也是取最短距离。}$$

同样可以证明其他方法均可由统一公式计算。

六、Q型(样品)聚类实例

现以我国南方几个少数民族 11 个体形指标均值资料(表 5-12)为例,说明对样品进行聚类的过程。

表 5-12 我国部分少数民族体形指标的测定均值

代码	民族	头长	头宽	额宽	面宽	下颌角	面高	形面高	鼻高	鼻宽	颞宽	身高
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
A	广西瑶族	180.7	148.8	100.3	138.2	102.6	189.3	119.2	52.4	37.7	56.7	1568.3
B	湖南侗族	184.3	151.9	117.9	127.9	113.3	184.7	110.7	52.5	38.3	51.4	1594.0
C	广西侗族	182.5	150.1	101.9	139.2	105.1	194.2	120.5	52.5	37.8	56.3	1579.3
D	土家族	183.9	151.9	105.4	141.5	110.2	186.4	119.2	48.9	45.4	48.5	1592.6
E	崩龙族	189.5	148.1	104.2	139.5	111.6	189.1	123.7	51.8	38.9	53.3	1607.3
F	基诺族	183.5	144.6	108.5	137.1	106.2	185.6	112.3	48.0	39.5	55.3	1599.7
G	白族	187.5	191.8	151.5	106.6	142.1	108.3	184.7	117.2	51.1	35.3	1551.7
H	贵州苗族	184.9	151.1	104.6	143.4	104.1	188.2	116.1	48.6	37.8	56.1	1558.0

为了消除 11 个指标的不同测量单位对聚类结果的影响,首先对原始数据进行标准差标准化处理,即使每一指标都具有均数为零、标准差为 1 的特性。本例标准化后的数据见表 5-13。

表 5-13 标准化数据

代码	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
A	-1.204	-0.377	-1.091	-0.086	-1.335	0.479	0.429	0.875	-0.389	1.033	-0.932
B	-0.227	0.854	2.178	-2.228	1.480	-0.965	-1.556	0.927	-0.184	-0.769	0.041
C	-0.716	0.139	-0.794	0.122	-0.677	2.016	0.733	0.927	-0.355	0.897	-0.515
D	-0.336	0.854	-0.144	0.601	0.664	-0.432	0.429	-0.953	2.246	-1.755	-0.012
E	1.184	-0.655	-0.367	0.185	1.033	0.416	1.480	0.561	0.021	-0.123	0.545
F	-0.444	-2.045	0.432	-0.315	-0.388	-0.683	-1.182	-1.423	0.227	0.557	0.257
G	1.808	0.695	0.079	0.725	0.164	-0.965	-0.038	0.196	-1.211	-0.667	1.938
H	-0.064	0.536	-0.293	0.996	-0.940	0.133	-0.295	-1.110	-0.355	0.829	-1.322

在标准化数据的基础上按公式计算 8 个少数民族间的欧氏距离 (见表 5-14),

表 5-14 八个少数民族间的欧氏距离

代码	A	B	C	D	E	F	G	H
A	0.000							
B	1.807	0.000						
C	0.582	1.834	0.000					
D	1.577	1.616	1.560	0.000				
E	1.239	1.616	1.083	1.242	0.000			
F	1.275	1.548	1.481	1.448	1.359	0.000		
G	1.625	1.553	1.588	1.468	1.025	1.492	0.000	
H	0.898	1.755	1.003	1.313	1.282	1.120	1.399	0.000

然后按系统聚类法过程首先选取最小距离的 A 和 C 类 ($D_{AC}=0.582$) 合并为一类, 取最短距离法作为计算合并新类与其他类间距离的方法。如此反复, 直到把所有八个类合并为一大类为止。最后综合上述过程如表 5-15 所示; 绘制其聚类树图如图 5-7 所示。

表 5-15 民族数据的 Q 型 (样品) 聚类过程

类数	合并类	新类名	新类频数	距离
7	A C	G7	2	0.582
6	G7 H	G8	3	0.898
5	E G	G9	2	1.025
4	G8 G9	G10	5	1.083
3	G10 F	G11	6	1.120
2	G11 D	G12	7	1.242
1	G12 B	G13	8	1.548

值得注意的是: 用欧氏距离定义类间相似性时, 要求参与运算的指标间互不相关。民族体型数据中的 11 个变量存在不同程度的相关性。故此, 可先行主成分得分计算, 然后进行聚类分析。本例采用主成分得分进行聚类的结

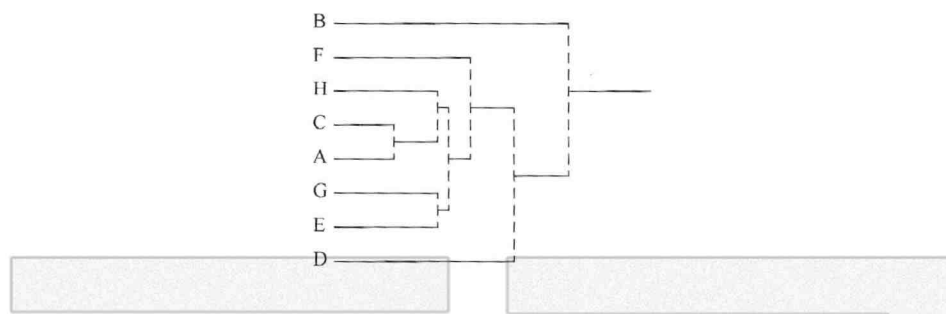


图 5-7 Q 型 (样品) 聚类树图

果如表 5-16 及图 5-8 所示。

表 5-16 主成分聚类过程

类数	合并类	新类名	新类频数	距离
7	C H	G9	2	0.285965
6	A G9	G10	3	0.297997
5	D E	G11	2	0.407777
4	G11 G	G12	3	0.436841
3	G10 F	G13	4	0.621112
2	G13 G12	G14	7	0.825451
1	G14 B	G15	8	0.863826

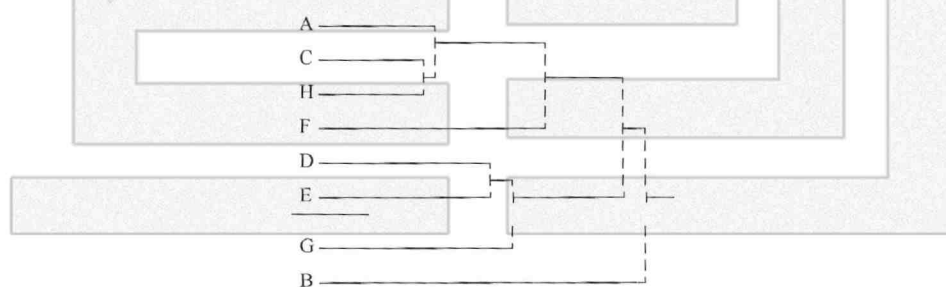


图 5-8 主成分聚类树图

最后，我们可以在不同的类间距离水平上将 8 个民族分成不同的类。例如在类间距离为 0.65 的水平上分为三大类，第一类为 (A, C, H, F) 即广西瑶族、广西侗族、贵州苗族与基诺族；第二类为 (D, E, G) 即土家族、

崩龙族与白族；第三类为(B) 湖南侗族。在聚类分析的应用过程中应结合客观实际情况选择一种适合的归类。

七、R 型 (指标) 聚类实例

仍以表 5-12 数据为例，用 11 个体形指标间的相关系数作为指标间相似性测度，对 11 个指标变量进行聚类分析。其分析步骤为：

1. 计算 11 个指标间的相关系数 (见表 5-17)。

表 5-17 11 个体形指标间的相关系数 (* 取其绝对值)

变量	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
X1	1.000										
X2	0.175	1.000									
X3	0.139	0.185	1.000								
X4	0.275	0.031	0.769	1.000							
X5	0.469	0.281	0.695	0.482	1.000						
X6	0.373	0.081	0.675	0.260	0.473	1.000					
X7	0.199	0.035	0.781	0.586	0.118	0.631	1.000				
X8	0.016	0.266	0.002	0.430	0.164	0.382	0.291	1.000			
X9	0.319	0.050	0.012	0.064	0.293	0.103	0.105	0.448	1.000		
X10	0.396	0.500	0.467	0.118	0.789	0.589	0.069	0.112	0.555	1.000	
X11	0.801	0.009	0.281	0.004	0.524	0.502	0.004	0.078	0.185	0.535	1.000

2. 首先各指标自成一类，即 G_1, G_2, \dots, G_{11} 类。然后从相关系数表中挑选出一个绝对值最大相关系数 $r_{1, 11} = 0.801$ ，合并 G_1 与 G_{11} 为一新类 G_{12} 。再计算各类与该新类的相关系数。在此采用最短距离法定义新类与其他任何一类 (设类 k) 的相关系数为合并前 G_1 和 G_{11} 与 G_k 类相关系数最大者，即：

$$r_{12, k} = \max (r_{1, k}, r_{11, k}) \text{ 如}$$

$$r_{12, 2} = \max (0.176, 0.009) = 0.176$$

$$r_{12, 3} = \max (0.139, 0.281) = 0.281$$

如此获得所有各类与新类的相关系数，得到新的相关系数矩阵。

3. 重复以上过程, 挑选另一个绝对值最大的相关 $r_{5, 10}=0.789$, 合并 G5、G10 为一新类 G13, 再计算 G13 其他各类的相关系数, 获得新的相关系数矩阵。如此反复, 循环一次便合并一类, 直到把所有各类合并为一大类为止。

综合其聚类过程如表 5-18 与图 5-9。

表 5-18 R 型 (指标) 聚类过程

类数	合并类	新类名	新类频数	相关系数
10	G1 G11	G12	2	0.801
9	G5 G10	G13	2	0.789
8	G3 G7	G14	2	0.781
7	G4 G14	G15	3	0.760
6	G13 G15	G16	5	0.695
5	G6 G16	G17	6	0.675
4	G9 G17	G18	7	0.555
3	G12 G18	G19	9	0.535
2	G2 G19	G20	10	0.500
1	G8 G20	G21	11	0.448

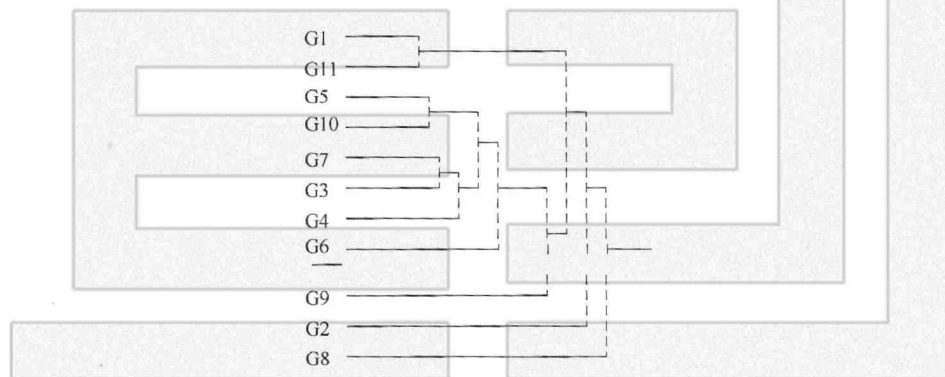


图 5-9 R 型 (指标) 聚类树图

一般而言, 不同的系统聚类方法的效果是不完全相同的。由于客观事物的千差万别, 在不同的问题中类的含义是不尽相同的, 要想给类一个严格定义, 决非一种简单的事情。在系统聚类中, 没有一种方法能把前面步骤中也许“不正确地”聚类的对象重新进行分类, 所以, 聚类的结果应当仔细考查, 看它是否符合实际。一般而言, 最短距离法适用于条形、S 形的类; 最长距离

法、重心法、类平均法及离差平方和法适用于椭圆形的类。对于特殊的问题，最好试用几种聚类法，并且在一个确定的方法内给出距离（或相似系数）的两种不同的方法。若来自各种方法的结果基本一致，则在很大程度上说明聚类确实反映了事物的规律性。

附：SAS 运算程序

```
data popul;
  input name $ x1-x11;
  cards;
A 180.7 148.8 100.3 138.2 102.6 189.3 119.2 52.4 37.7 56.7 1568.3
B 184.3 151.9 117.9 127.9 113.3 184.7 110.7 52.5 38.3 51.4 1594.0
C 182.5 150.1 101.9 139.2 105.1 194.2 120.5 52.5 37.8 56.3 1579.3
D 183.9 151.9 105.4 141.5 110.2 186.4 119.2 48.9 45.4 48.5 1592.6
E 189.5 148.1 104.2 139.5 111.6 189.1 123.7 51.8 38.9 53.3 1607.3
F 183.5 144.6 108.5 137.1 106.2 185.6 112.3 48.0 39.5 55.3 1599.7
G 191.8 151.5 106.6 142.1 108.3 184.7 117.2 51.1 35.3 51.7 1644.1
H 184.9 151.1 104.6 143.4 104.1 188.2 116.1 48.6 37.8 56.1 1558.0
;
PROC CLUSTER DATA=DONG.POPUL (TYPE=DATA) METHOD=
  AVERAGEOUTTREE=A
  SIMPLE STD NOSQUARE RMSSTD RSQ PSEUDO CCC NON-
  ORM;
  VAR X1-X11;
  ID NAME;
PROC TREE HORIZONTAL PAGES=1;
  IDNAME;
RUN;
DATA POPUL;
  INPUT NAME $ X1-X11;
  CARDS;
```

A 180.7 148.8 100.3 138.2 102.6 189.3 119.2 52.4 37.7 56.7 1568.3
 B 184.3 151.9 117.9 127.9 113.3 184.7 110.7 52.5 38.3 51.4 1594.0
 C 182.5 150.1 101.9 139.2 105.1 194.2 120.5 52.5 37.8 56.3 1579.3
 D 183.9 151.9 105.4 141.5 110.2 186.4 119.2 48.9 45.4 48.5 1592.6
 E 189.5 148.1 104.2 139.5 111.6 189.1 123.7 51.8 38.9 53.3 1607.3
 F 183.5 144.6 108.5 137.1 106.2 185.6 112.3 48.0 39.5 55.3 1599.7
 G 191.8 151.5 106.6 142.1 108.3 184.7 117.2 51.1 35.3 51.7 1644.1
 H 184.9 151.1 104.6 143.4 104.1 188.2 116.1 48.6 37.8 56.1 1558.0

```

;
PROC VARCLUS DATA = POPUL OUTSTAT = A OUTTREE = B
    SUMMARY HIERARCHY;
    VAR X1—X11;
PROC TREE DATA = B H = NCL H SORT DESCENDING;
FUN;
    
```

第六章 中国少数民族遗传多态性

第一节 遗传多态性概述

人类基因组 (human genome) 是由一套单倍体的染色体 (22 条常染色体及 X 和 Y 染色体) 所负载的 DNA 总和。大约由 3×10^9 个碱基对组成。人类基因组的 DNA 序列中不仅包含着整套基因的编码序列, 而且拥有庞大数量的非编码序列和重复序列。

人类基因组的组织结构非常复杂, 其 DNA 序列组成可根据结构和功能从不同角度分成下列不同的类别。

一、基因序列与基因相关序列

基因序列是指那些编码蛋白质的 DNA 序列, 只占基因组的 1.1% ~ 1.4%。人类基因乃至整个高等生物的基因是由内含子 (intron) 和外显子 (exon) 组成, 编码蛋白质的遗传信息存在于一系列的外显子中, 外显子由非编码的内含子隔开。

基因相关的 DNA 序列, 其中包括大量的内含子序列, 约占基因组的 24%, 以及分别存在于基因的 5' 端 UTR 和 3' 端 UTR 的 DNA 序列, 这些 DNA 序列只被转录而不被翻译; 第三种与组成基因相关的 DNA 序列是调节基因转录的启动子序列 (promoter sequence), 它们位于基因上游。此外, 还有一些假基因 (pseudogene) 和一些基因片段 (gene fragment)。所有这些大量与基因相关的序列总称为非编码序列。基因的编码序列与非编码序列在整个人类基因组 DNA 中只占 20% ~ 30%。

基因外 DNA (extragenic DNA) 序列是指除基因、基因相关序列、假基

因和基因片段以外的 DNA 序列，占基因组 DNA 总数的绝大部分（70%~80%）。

绝大多数基因外的 DNA 序列是单拷贝或低拷贝的序列，占基因外 DNA 的 70%~80%，其余的 20%~30%是中度重复序列和高度重复序列。这些重复序列以串联排列或以分散形式存在于基因组中。其中，中度重复序列的拷贝数有 10 至数百份，重复单元的平均长度约为 300bp。高度重复序列的重复数目可达数百到数百万份拷贝。

1. 重复序列

重复序列主要是基因外 DNA 序列。顾名思义，重复序列是指在基因组中重复出现的 DNA 序列。重复单元的长度相差悬殊，短的只有一个核苷酸，长的重复单元可长到整个基因（如 rDNA）。重复序列在基因组中的分布有的是散在的，有的是成簇存在的。根据重复序列在基因组中的组织形式，分为分散重复序列（dispersed repetitive sequence）和串联重复序列（tandem repeat sequence）两大类。

(1) 分散重复序列

这类 DNA 序列一般属于中度重复序列，以散在方式分布于基因组。依重复单元的长度又可分为长分散重复序列（long interspersed repeated sequence 或 long interspersed nuclear element, LINE）和短分散重复序列（short interspersed repeated sequence 或 short interspersed nuclear element, SINE）。SINE 是基因组中较短的一类分散重复序列，重复单元长度在 100~300bp 之间，重复拷贝数可达 10 万份以上。人类基因组中最典型的 SINE 代表是 Alu 家族（Alu family），由于这种 DNA 序列中有限制性内切核酸酶 Alu I 的识别序列 ACCT，故称 Alu 重复序列。Alu 家族是人体基因组中含量最丰富的中度重复序列，大约有 100 万份拷贝，平均每 4~6kb 就有一个 Alu 序列，每个序列长度为 282bp。Alu 序列一般散在分布于整个人体基因组中，少数呈簇状分布。从细胞遗传学水平上观察，Alu 重复序列集中在染色体上基因转录最活跃的区段。在所有已知的基因内含子中几乎都发现有 Alu 序列。Alu 序列的功能可能与转录调节以及 DNA 复制的起始有关。

LINE 重复单元的长度在 1000bp 以上。LINE 在人类基因组中有上万份拷贝，它与 SINE 相反，是可以自主转座的一类逆转录转座子（retroposon），

来源于 RNA 聚合酶 II 的转录产物。LINE-I (LI) 是人类基因组及其他哺乳类基因组中的一种反转录转座子, 它的总长度为 6500bp, 存在 60000 份拷贝, 这种 DNA 序列能在基因组中通过转座 (transposition) 过程而移动位置。

(2) 串联重复序列

人类基因组含有大量的正向 (首一尾) 相连或反向 (尾一尾、头一头) 相连成长串的重复序列, 称为串联重复序列 (tandemly repeat sequence) 或簇状重复序列。根据重复单元的大小或重复序列簇的长度分为下列 3 类:

① 卫星 DNA (satellite DNA), 是一类重复单元很大的高度重复序列, 其长度一般为 100~5000kb, 多分布在染色体的异染色质区 (端粒、着丝粒), 难以采用分子杂交或 PCR 技术揭示其多态性 (polymorphism)。

② 小卫星 DNA (minisatellite DNA), 长度通常在 9~70bp 之间, 重复单元的数目 (n) 为几个至几百个。小卫星 DNA 主要存在于染色体末端靠近端粒区域。同一物种不同个体间, 在给定的基因座 (locus) 上, 重复的数目是可变的, 所以小卫星 DNA 串联重复又称数目可变串联重复 (variable number of tandem repeat, VNTR)。由于不同个体间 VNTR 重复数目有差异, 因而在群体中具有丰富的多态性。不过, 重复单元的碱基组成可能有变异, 例如发生单个碱基的置换会造成长度上的差异。

③ 微卫星 DNA (microsatellite DNA)。这一类重复序列分布于染色体的任何区域, 由 1~4 个核苷酸的重复单元串联而成, 长达几十个碱基对到几百个碱基对, 在基因的间隔区和内含子等非编码区内广泛存在。人类基因组中至少有 30000 个微卫星序列基因座。基因组中两个核苷酸的重复非常普遍, 平均每 30~60kb 就存在一个这类的重复序列, 其重复数目 (n) 约为 15~60, 不言而喻, 重复次数越多, 所提供的信息量越大。有资料表明, CA 重复和单核苷酸重复如 (A)_n 两者共占整个基因组的 0.8%。这类短序列的微卫星 DNA 又称短串联重复 (short tandem repeat, STR)。

二、DNA 多态性标记

1. 遗传标记

凡用于遗传分析的生物体中可识别的等位基因、可追踪的染色体或染色体的某一节段、或在家系中传递的任何一种遗传特征, 均可称为遗传标记

(genetic marker)。

20 世纪末, 由于遗传学在基因定位、遗传作图、基因组研究以及群体遗传学与进化遗传学、法医学等领域的扩展与应用, 经典的、基于表型的形态学、细胞学水平上的遗传标记已远远不能满足其要求, 一批新型的遗传标记, 如蛋白质标记、同工酶标记、DNA 标记等分子标记都纷纷被开发利用。

从遗传分析的角度来看, 有价值的遗传标记, 不仅要求取材方便, 检测方法和实验技术简便且灵敏度高, 还要按简单的孟德尔方式遗传, 同时又必须是高度多态性和高水平的遗传杂合性。也就是说, 某一遗传标记在群体的不同个体中应存在若干种不连续的变异类型, 每种类型发生的频率大于 1%, 该标记称为遗传多态性标记 (genetic polymorphic marker)。

用于人类基因组研究和人类群体遗传学研究的人类遗传标记 (human genetic marker) 如下:

人类红细胞血型系统, 诸如 ABO、Rh、MN、Kell、Duffy、P 等, 是最早研究的遗传标记, 其中 ABO 血型系统是早在 1900 年由 Karl Landsteiner 发现的多态系统。随后, 发展利用了人类的白细胞抗原 HLA-A、-B、-C、-D、-DP、-DQ、-DR 等; 血清蛋白型 HP、GC、BF、TF、a-AT、G、C1 等以及同工酶 PMG、ESD、ACP 等多态性遗传标记。上述这些被视为第一代人类遗传标记的应用是基于人们不断发展的检测技术, 尤其是电泳技术。先前应用的是纸上电泳, 随后发明了淀粉凝胶电泳, 直到后来聚丙烯酰胺凝胶电泳、双向电泳、等电聚焦电泳新技术和免疫技术的应用大大提高了对酶和蛋白质多态现象的检出率和准确性。

虽然迄今发现的人类红细胞表面抗原、红细胞酶、血清蛋白以及人类白细胞抗原等多态基因座已达数百种, 但是, 由于已知多态的蛋白质的种类很少、等位基因 (或复等位基因) 的数目有限而且又无法获得足够多的信息量、检测技术烦琐等因素限制了该类遗传标记在人类群体遗传学和人类基因组学的广泛应用。因而促使人们从 DNA 分子上寻找新的遗传标记, 而 DNA 技术的建立确实为人类提供了大量新的遗传标记。

2. DNA 分子标记

20 世纪 80 年代初, 限制性内切核酸酶技术和基因扩增技术以及随之而来的 DNA 直接测序, 引发了对 DNA 多态性的研究。DNA 序列具有比蛋白质

序列或蛋白质电泳变化多得多的信息。DNA 序列改变所造成的 DNA 分子结构的变异称为 DNA 多态性 (DNA polymorphism)。显然, DNA 多态性本质上是核苷酸排列顺序的差异性。正因为存在这类差异, 造成了群体中每一个个体在遗传上是不同的。有数据说明, 人类基因组 DNA 平均异质性为每 200 ~ 1000 个核苷酸对就有一个是不相同的。

核内 DNA 序列变异的大量信息一方面储存在基因内的非编码区 (即基因的内含子、启动子和侧翼序列), 另一方面大量存在于基因外 DNA 的各种重复序列中。此外, 核外的线粒体 DNA 的多态性也同样提供了重要信息。

3. DNA 多态性标记的种类及其特性

(1) 限制性片段长度多态性

当同一物种不同个体的 DNA (核 DNA) 或同一基因用同一种限制性内切核酸酶切割后, 由于内切核酸酶的识别序列有差异, 不仅会出现分子质量不同的同源等位片段, 使这些 DNA 片段在长度上存在差异, 而且这种差异也反映在酶切片段的数目上, 这些差异的出现即所谓的限制性片段长度多态性 (restriction fragment length polymorphism, RFLP)。其多态性基于下列原因: 如果当 DNA 分子中的核苷酸序列发生了微小改变, 甚至只有一个核苷酸的变化, 而改变了的核苷酸正好位于限制性内切酶的识别序列内, 用这种酶切割 DNA 时, 因为酶切位点的丢失或产生, 导致酶切后产生的 DNA 片段的长度和数目的变化。如果改变了的核苷酸形成了一个新的识别位点, 仍利用这种限制性内切酶酶切时, 也会使酶切后得到的 DNA 片段的数目和长度发生变化。

RFLP 标记的检测不受生物体所处环境条件和发育阶段的影响, 其在等位基因间是共显性的。由于该标记源于基因组 DNA 的自身变异, 因而其多态性丰富。基于这些优点, RFLP 特别适于构建遗传连锁图谱, 人类的第一张 RFLP 图谱就是在 1987 年由 Doniskeller 等人构建的。不仅如此, RFLP 也被用于进化遗传学的研究。

特别是近年来人们对线粒体 DNA (mitochondrial DNA, mtDNA) 的研究不断深入, 已成为研究近缘种内、种间、群体间遗传分化关系, 特别是研究人类进化的理想工具。利用限制性内切酶酶谱分析方法, 对人类 mtDNA D 环 465bp 的高变区限制性类型进行检测和比较, 探讨不同民族、人种间的亲

缘关系和演化规律，取得了举世瞩目的成果。

(2) AFLP 标记

AFLP 是扩增片段长度多态性 (amplified fragment length polymorphism) 的简称。

所有通过 PCR 反应扩增基因组 DNA 所得到的片段长度多态性均可统称为 AFLP。

1995 年由 Vos 等建立的 AFLP 技术是用于 DNA 指纹法的有效方法，是一种基于 PCR 反应的 RFLP。该方法只有那些和引物序列完全匹配的限制片段才能被扩增。如果在某一个限制酶的识别位点上发生了一个新的突变，那么该引物将不能识别此位点。于是，与这个限制位点对应的限制片段就不会出现在凝胶上。在不同个体中，电泳条带的存在与否就意味着这个位点上的多态性，这种多态性可以被用来测定种间和种内的遗传变异。

AFLP 标记是共显性或是显性的，对模板 DNA 需要的数量少而且纯度要求不高。

但是，鉴于一般 PCR 反应的要求，在进行 AFLP 分析前，必须根据基因组中被扩增的 DNA 片段两端的序列，设计相应引物，这便是进行 AFLP 分析时的最大限制因素。然而，这一局限性将会随着对已知序列的基因数目的增加而使得 AFLP 的应用范围随之扩大。

(3) RAPD 标记

RAPD 是随机扩增多态性 DNA (random amplified polymorphism DNA) 的简称。

RAPD 标记是 1990 年由 Williams 发展起来的一种新型的遗传标记。在 PCR 技术的基础上利用一系列的、随机的短引物 (通常只有 10bp) 对所研究的基因组 DNA 进行扩增获得的长度不同的多态 DNA 片段。理论上可将 RAPD 看成是一种特殊形式的 AFLP。实际上，RAPD 与 AFLP 的最大区别是用于扩增多态性 DNA 的引物不是专一的，而是随机的。这种分子标记的多态性是由于模板 DNA 中引物结合区的核苷酸变化 (如碱基插入、缺失、DNA 重排等) 而导致 DNA 片段的多态性。虽然对某一特定的引物而言，检测多态性的区域是有限的，但是利用足够多的随机引物 (一般为数十对) 就可以对整个基因组进行一次性扫描，从而检查出 DNA 的任何变化。RAPD 比 RFLP

技术操作简单易行,省时省力,可以快速而廉价地获得群体内和群体间大量的遗传标记。此外,RAPD引物的通用性也是一个突出的特点。但RAPD标记也有其不足之处,即每个标记所提供的信息量少、RAPD的检测受反应条件的影响较大、重复性较差。因此,在构建基因组的物理图谱时,需将RAPD标记克隆后转变成RFLP标记。

(4) SSCP 标记

SSCP—DNA 是单链构象多态 DNA (single strand conformation polymorphism, SSCP DNA)。当 DNA 片段的双链变性成单链之后,在非变性的条件下折叠成不同的空间构象,单链的这种构象会因为个别核苷酸的置换而改变。长度相同或相近的 DNA 单链在电泳时,由于构象不同而具有不同的迁移率,从而显示出多态性。利用此项技术可以检测基因的点突变。但由于检测技术较复杂,实际上很少用作常规的分子标记。

(5) SNP 标记

单核苷酸多态性 (single nucleotide polymorphism, SNP) 指的是不同个体的基因组 DNA 序列在某一特定位置上的单个核苷酸的差别,这种差异包括单个或多个碱基的缺失或插入,但更常见的是单个碱基的置换。具有这类单个核苷酸差异的基因座或 DNA 序列是按孟德尔方式遗传的,其出现的频率如果超过 1%,则视为单核苷酸多态性。SNP 可作为基因组作图的标记。

大多数的 SNP 属于碱基置换中的转换,即由一种嘧啶置换另一种嘧啶,或由一种嘌呤置换另一种嘌呤。在 CG 序列上出现最频繁,多半是 C 转换成 T,颠换的情况不常见,一般颠换与转换之比为 1:20。SNP 可划分为两种:一种是绝大多数分布在基因组非编码序列中的单核苷酸变异;另一种是分布在基因编码序列中,称为 cSNP (coding SNP)。cSNP 既可以编码产生同样的蛋白质,也可编码产生突变的蛋白质,导致蛋白质的多态性。cSNP 与基因的点突变之间的区别只在于其在群体中出现的频率,若出现频率低于 1% 时,则视为点突变。

在人类基因组中,SNP 分布广泛,数量巨大。如果比较任意两条同源染色体的碱基序列,就会发现平均每 1000bp 有一个碱基的差异,如果这种差异能够导致基因功能的改变,则通常称为基因的点突变,该概念与先前提及的原理相同。如果我们假定 1/1000 的碱基是多态的话,那么人类基因组中就会

拥有 300 万个 SNP 标记。人类基因组的研究已经证实,在我们的 DNA 中有 142 万个单核苷酸多态性位点 (SNP),它的密度为每 1910bp 中有一个 SNP,这就意味着超过 90% 的 20000bp 以上的序列就可能包含一个甚至更多的 SNP。由此可见,SNP 标记的数目比微卫星标记数要多出几个数量级。由于 SNP 在人类基因组中具有可观的数量,就整体而论,其多态性是极其丰富的。在 1998 年发表的第一张人类 SNP 遗传连锁图包括了 2227 个基因座,在染色体上分布的平均密度达到了 2cM。此外,由于 SNP 是二态的,易于自动化批量检测,因而越来越受到重视。

绝大多数 SNP 位于 DNA 的非编码区,由于选择压力的作用,在非转录序列中的 SNP 要多于转录区。有人估计在人类基因组内 300 万以上的 SNP 中,大约只有 20 万个存在于编码区。国际单核苷酸多态性绘图工作小组估计,他们已在基因中鉴别出了 60000 多个编码 SNP,其密度为每 1080bp 就有一个编码 SNP。对人类基因组中的这类 DNA 多态性分析有助于解释个体的表型差异,基因的连锁分析与基因定位,不同群体与个体对疾病特别是对复杂疾病的易感性,以及对各种药物的耐受性和对环境因子的反应及在法医学中的应用。SNP 与 RFLP 标记的主要区别在于,它不再以 DNA 片段的长度变化作为检测手段,而直接以 DNA 序列变异作为标记,所以 SNP 标记的遗传分析完全摒弃了经典的凝胶电泳方法,以最新的 DNA 芯片技术取而代之,因而应用前景十分广阔。

(6) DNA 重复序列标记

① VNTR (数目可变串联重复),一般表现为两种形式:一种是长序列的重复单元,通常在 6~30bp 之间,属于小卫星 DNA 一类的重复序列;另一种是重复单元的碱基数低于 6bp,一般为 2~4bp,而重复数目可变,是一类微卫星 DNA 的重复序列,也可称为 VNTR。但是近年来,将特指微卫星 DNA 重复中短序列 (2 核苷酸、3 核苷酸、4 核苷酸) 的串联重复称为短串联重复 (STR),其多态性同样是由于重复单元数目的可变性所造成。

一个特定的 VNTR 是由重复单元的碱基序列差异和重复单元的数目变化而形成丰富的多态性的。构成 VNTR 的重复单元由中间的核心区和外围的侧翼区的碱基组成。例如,在人类肌红蛋白的第一个内含子中分离出的 132bp 的重复序列中,内含 4 个重复单元,每个重复单元长 33bp,内有 13bp

(GGAGGTGGGCAGG)的核心序列。1985年发现该核心序列后,它便成为人类VNTR的第一个探针。人的基因组DNA提取后,用限制性内切酶切割成不同长度的片段,以VNTR中的核心序列为探针进行Southern杂交,可发现阳性片段的长度各不相同。不同个体的这种串联重复的重复单元的数目和位置都不相同,因而,VNTR的Southern杂交带谱就具有高度的个体特异性。

VNTR在人类基因组中分布极为广泛,约占人类基因组的10%,在人群中表现高度的个体特异性,等位基因按孟德尔共显性方式遗传,因而在个体识别、亲子鉴定、基因诊断、基因作图等研究领域具有较高的应用价值。目前,检测VNTR的方法一般都采用PCR和电泳技术,比RFLP标记更方便和快捷。

② STR短串联重复是人类基因组中另一类重要的、属于微卫星DNA的重复结构,是继RFLP及VNTR之后新发现的多态性遗传标记。STR的核心序列一般由2~4个碱基组成,核心序列重复数目的变化造成了STR基因座的遗传多态性。PCR分析时,STR等位基因片段的大小一般为100~400bp。人类基因组中的STR有(CA)_n、(GA)_n、(AA)_n以及(CAA)_n、(CGG)_n、(GACA)_n等,其中以2核苷酸重复最为常见,尤以(CA)_n、(GT)_n最多。早在1984年,Litt和Weber等人证明(CA)/(GT)的2核苷酸短串联重复序列具有高度多态性之后,Edwards等人又发现了3核苷酸和4核苷酸的短串联重复的多态性。其研究还表明,3核苷酸和4核苷酸重复序列比2核苷酸重复的PCR扩增产物更为真实。但是,4核苷酸重复序列因其特异性高、分型容易而备受关注。STR散布于整个基因组,多在基因的间隔序列、内含子及非编码区内。但3核苷酸的串联重复可能存在于某些基因的编码区以及基因3'UTR,如果其重复单元的数目异常地增多,将导致某些遗传病的发生。如已知的强直性肌营养不良遗传病是由于在其致病基因的3'端非翻译区内3核苷酸重复序列(CTG)_n的拷贝数急剧增加所致,患者的重复单元数目(n)可高达80~2000。另外,3核苷酸重复序列(CAG)_n在正常人的基因组中的重复数目为17~26,而一种罕见的脊髓及延髓性肌萎缩(X连锁遗传)遗传病患者的雄激素受体基因编码区内的3核苷酸重复序列(CAG)_n过度增加达40~52份拷贝,超出正常人的2倍。重复拷贝数改变

以后的基因的可突变性，将不同于拷贝数改变以前的基因，这类 3 核苷酸重复数的突变不同于通常所指的突变率很低、而且在长期进化过程中保持相对稳定的那些基因的突变，所以现代遗传学将 3 核苷酸扩增突变称为动态突变 (dynamic mutation)。

STR 在基因组中的分布密度为平均每 15~20kb 就存在一个 STR 基因座，在人类的 30 亿碱基对中约有 4×10^5 个左右的 STR。如此大量的 STR，无疑将提供巨大的信息量。与 RFLP 和 VNTR 相比，STR 多态性更丰富，重复单元的长度较短，更适于 PCR 快速检测。加之 STR 遵循孟德尔共显性遗传等特点，因而成为人类遗传学、基因组学、法医学、群体遗传学及分子生物学首选的多态性标记而得到广泛的应用。具有 5264 个 2 核苷酸重复的微卫星标记 (STR) 的人类遗传图已于 1996 年完成。该图谱显示，第 17 号染色体上平均每 495kb 有 1 个 STR 标记，第 9 号染色体上每 767kb 有 1 个 STR 标记，整个基因组中，只有 3 处标记间距大于 4Mb，平均分辨率已达 600kb，STR 的应用价值可见一斑。

第二节 ABO 血型系统

ABO 血型最常见的三种表型是由三种等位基因组成的基因型所决定的。三种等位基因是 A、B 与 O，它们位于 9q34.1-34.2。其所组成的基因型与表型的关系见表 6-1

表 6-1 ABO 血型系统常见的基因型与表型

基因型	表型
AO、AA	A
BO、BB	B
AB	AB
OO	O

1930 年，发现 A 基因又可分 A^1 与 A^2 两个亚型， A^2 在黑人中一般占 20%~40% 以上，在白种人中大约占 20%，而在中国人中， A^2 只占 A 基因的 1%~9%。由于 A^2 等位基因起源于白种人，因而在白种人中十分常见。而

在亚洲人中该基因与正常 A¹ 等位基因的差别在于除了有一个 C467T (Pro156Leu) 突变外, 还在 1059~1061 位上三个连续的碱基中有一个缺失, 结果使读框延长了 64 个碱基, 于是 ABO 系统的基因型与表型进一步可分为如表 6-2 所示的类型。

表 6-2 A 等位基因区分为 A¹ 与 A² 后的基因型与表型

表 型	基 因 型
A ¹	A ¹ A ¹
	A ¹ A ²
	A ¹ O
A ²	A ² A ²
A ² O	
B	BB
	BO
A ¹ B	A ¹ B (或 A ¹ B/O)
A ² B	A ² B (或 A ² B/O)
O	OO

另一大量存在的 A 型等位基因在亚洲人中十分普遍, 它有一个 C467T 的突变, 引起一个 Pro156Leu 取代, 但这对其酶的活性影响很小。另外已报道, 还有两个罕见的突变 A¹ 等位基因, 一个 A¹ 等位基因除 C467T 外, 还有一个沉默突变 C564T。另一个 A¹ 等位基因与正常的 A¹ 等位基因相比多一个沉默突变 A297G。据 2001 年报道, 表现为较弱活性的 A 等位基因已由过去认为的 8 个增加至 16 个。在中国发现的 A 基因其他等位基因有 A³、AX、Am 等, 它们均表现为较弱活性。与 A¹ 相比, A³ 的碱基变异为 G871A, AX 的变异为 T646A, 而 Am 则可能是在基因的启动子区域发生了一个突变。这些亚型在中国人中均较罕见。

表现较弱的 B 等位基因已由过去的 6 个增加至 13 个。在中国人中已发现的 B 基因的亚型有 B²、B^X、B^m 等, O 基因则有类孟买型 O^m 等, 但其频率也很低。

随着研究的深入, 今后必将发现更多的新的等位基因, 但由于这些等位基因的发生率均很低, 所以对世界各人群的 ABO 血型的基因频率没有什么实

质性的影响。但罕见等位基因有时可以用于追踪某一人群的来源及分析个体间的血缘关系。

ABO 基因克隆出来之后,想知道某一个人的 ABO 血型基因型,可对 ABO 基因型的碱基序列直接进行检测,而无需再对其家系进行全面的研究。目前检测 ABO 血型基因碱基序列的方法很多,一般都在聚合酶链反应扩增后,分析限制性片段长度多态性(RFLP)或单链构象多态性,甚至直接测序等。用得最多的方法是聚丙烯酰胺凝胶电泳或琼脂糖凝胶电泳后用 EB 染色或银染。但是由于变异体很多,有的还比较常见,所得 DNA 多态性的结果也不一定十分精确。不过,如果能测到第 6 外显子中的 G261 缺失,同时能测到第 7 外显子中 A、B 及 O 的区别,则结果还是比较可靠的。

目前文献报道的 ABO 血型分布共涉及 51 个少数民族 308 个人群的 ABO 血型分布的数据,缺少俄罗斯族、塔塔尔族、门巴族、珞巴族四个民族的调查数据。所调查的 308 个人群中,藏族占了 32 个,是各民族中调查人数最多的。其次,苗族有 23 个,蒙古族有 22 个,回族有 20 个。而塔吉克族、乌孜别克族、怒族、独龙族、毛南族、仡佬族、拉祜族、布朗族、京族等族都只有一个人群被调查过,其原因主要是这些民族人口数较少,而且有的民族居住还很分散,有的民族居住地区交通很不方便等。高山族一共有 24 个人群的数据,而高山族实际上包含 9 个民族,我们仍按习惯,合在一起称之为“高山族”。

在我国西北地区塔塔尔族、塔吉克族、维吾尔族、撒拉族、东乡族和回族的部分人群(如宁夏的一些人群)中 A 基因频率较高。这和这些人群中有一定的白种人血缘有关,因为白种人的 A 型基因频率在各人种中是比较高的。但是在我国内蒙古额尔古纳右旗的俄罗斯族中, A 型等位基因频率却在 0.2 以下,这很可能是因为那里的俄罗斯族和汉族等蒙古人种的血缘混合已很严重。中国其余各少数民族的 A 型等位基因频率多数都在 0.2 以下,但彝族的一些人群的频率均比较高,在 0.2 以上,阿昌族、土家族、景颇族和佤族的 A 等位基因频率也较高。

在北方各民族(包括回族中的北方人群)中 B 基因频率较高,而南方各民族中则较低。这与汉族中北方地区 B 基因频率较高是一致的。但有的北方民族(如鄂温克族、鄂伦春族、藏族等民族)的部分人群的频率也较低,而

有的南方民族（如苗族、哈尼族）的部分人群频率也较高。

O 等位基因频率在鄂伦春族、鄂温克族、毛南族、仡佬族、傈僳族、高山族中的泰雅族、瑶族以及壮族人群中较高，在 0.7 左右，而在其余民族大部分人群中，频率在 0.5~0.7 之间。

按照郭景元、肖星甫等人的意见，根据 ABO 血型表型频率可将中国人大致分为三个组。第一组是北方组，包括黑龙江、吉林、辽宁、内蒙古、河北、山东、山西、河南、陕西、甘肃、青海、新疆、西藏，其特点是 B 基因频率高。第二组是长江流域组，包括云南、贵州、湖北、湖南、江西、安徽、江苏、浙江，其特点是 A 基因频率较高，而 B 基因频率较低。第三组是南方组，包括福建、广东、广西、海南、台湾，其特点是 O 基因频率较高。有的学者将湖南、湖北、四川单列为一个组，因为该地区的 B 基因频率特别低。过去有的学者把西藏单列为一个组，把海南归入北方组。根据目前的资料，西藏（包括藏族及汉族）明显属于北方组，而海南汉族则明显属南方组，但黎族则仍接近北方组。

应该指出，某一位点的基因频率可以在某一侧面反映出不同地区及不同民族的人群间的血缘远近，但要更准确地判断人群间的血缘远近，必须根据尽可能多的位点的基因频率，同时还需要参考考古学、历史学、语言学、人类学等各学科的有关证据，才能得出比较正确、全面的结论。

第三节 HLA 系统

HLA (human leukocyte antigen) 是人类的主要组织相容性复合体 (major histocompatibility complex, MHC)。HLA 系统是一个由一系列紧密连锁的基因座所组成的具有高度多态性的遗传复合体，也是目前已知的多态性最丰富的一个基因系统，拥有极大数量的等位基因，赋予种群巨大潜力以适应多变的内外环境。HLA 等位基因频率分布、单倍型组成及单倍型频率分布，在不同人种、不同地区均有明显差异，因此 HLA 基因的研究可作为一种遗传标记，广泛应用于民族的起源、进化、迁徙、融合等人类遗传学研究中。HLA 基因编码的分子不但在 T 细胞分化发育中是必需的，而且在天然免疫和

特异性免疫的启动和调节中也发挥着重要作用。

HLA 复合体定位于第 6 号染色体短臂的 6p21.31 区,长 3.6 Mb。HLA 系统是一个由一系列紧密连锁的基因座所组成的具有高度多态性的遗传复合体,1999 年已完成全部序列分析及基因定位。在 3.6 Mb 区域内共确认了 224 个基因座位点,到 2004 年该区域又新命名了 15 个基因座,共有 239 个基因座。其中 12 个为功能性基因,有产物表达。

中国是一个多民族国家,除了占主体的汉族外,尚有 55 个少数民族,因此 HLA 遗传多态性研究具有重要的意义。自 20 世纪 80 年代开始,我国科学工作者用国际上统一的血清学方法(补体依赖的微量淋巴细胞毒方法)及 PCR 基础上的分子生物学方法,对我国主要城市的汉族群体及至少 18 个主要的少数民族群体进行了 HLA I 类、II 类及 III 类抗原/基因频率分布研究,初步调查了中国人群中的 HLA 遗传多态性。

一、从 HLA 各座位等位基因频率分布看中国人群中的 HLA 遗传多态性

1. 中国人群 HLA 基因频率分布具有蒙古人种特点:世界上人种分为三大类,即蒙古人种(黄种人)、欧罗巴人种(白种人)和黑种人。从国际上组织的几次大规模 HLA 频率的群体调查发现,HLA 各位点基因频率的分布具有明显的种族特点。

(1) HLA-A 座位基因频率在我国各民族与群体中均是以 A2 最为常见,基因频率在 0.21~0.38 之间,几乎每 2 人中就有一个个体有 A2 基因。这一点与世界范围内绝大多数群体具有相似性。到目前为止,除巴布亚新几内亚外,黑种人、白种人各群体中也是以 A2 基因为最常见,提示 A2 基因在人类进化早期就已存在。除 A2 基因外,中国人群中常见的基因有 A11、A9、A10 和 A190。A11 基因中占主要的是 A * 1101 基因,A9 基因中主要是 A * 2402 等位基因,A * 10 基因中占优势的是 A * 2601,A * 19 基因中最常见的是 A * 3301 和 A * 3001,而 A1、A3、A23、A25 和 A28、A29、A31 均为低频率基因。这些分布与日本、韩国等蒙古人种十分相似。我国常见的这些 A 位点基因在白种人和黑种人中正好是低频基因,而他们常见的 A1、A3、A23、A28 又是蒙古人种中的低频基因,基因频率差异很大,如 A1、A3 在中国人中的基因频率均小于 1%,而在白种人中均在 10% 以上。即使是三大人种中

均为常见的 A2 基因，其等位基因频率的分布也具有明显的种族特异性，如德国白种人中 A * 0201 基因占了绝对优势，97% 的 A2 阳性个体均为 A * 0201，而在上海地区汉族人群中 A * 0201、A * 0206 和 A * 0207 均为常见等位基因，而 A * 0204 和 A * 0214 则为黑种人的典型 A2 等位基因。

(2) HLA-B 座位基因频率在中国各群体中以 B40 (尤以 B60 为主) 最高，约占 15%~23%，其次是 B46、B13 和 B15 (以 B62 为最常见)，它们的基因频率基本上都在 10% 以上，这四个基因频率的总和已超过了 50%。与白种人、黑种人相比，我国群体中这些常见的 B 基因正是他们的低频基因，而他们的高频基因如 B7、B8、B44 却又是我们蒙古人种中的低频基因。特别值得一提的是 B46 基因，在白种人中频率极低，而在黑种人中几乎不存在，因此被称为东方人种基因。黑种人中除了与白种人一样，B7、B8 为常见基因外，B42、B53 和 B70 均为较常见基因，与蒙古人种和白种人明显不同。

(3) HLA-C 座位基因频率在三大人种之间有相似性和独特性。HLA-Cw7 在三大人种均为最常见的基因，频率在 12%~28% 之间。除 Cw7 外，中国人人群中常见的 Cw 基因为 Cw1、Cw3 和 Cw6，白种人中除 Cw7 外，常见的 Cw 基因为 Cw4、Cw5 和 Cw6。黑种人中 Cw2 的基因频率则明显高于蒙古人种和白种人。

(4) HLA-DRB1 座位基因在三大人种中也存在明显的频率差异。中国人人群中常见的 DRB1 基因按频率高低一般为 DR * 15、* 09、* 07、* 12 和 * 04。白种人中常见的基因依次为 DR * 07、* 04、* 13、* 15、* 01 和 * 03。黑种人中最常见的基因是 DR * 03，其次为 DR * 13 和 * 11。值得提出的是 DRB1 * 09 基因，蒙古人种中其基因频率一般不低于 10%，甚至可高达 17%，而在白种人和黑种人中的基因频率极低，大多数在 1% 以下，因此 DRB1 * 09 基因也被称为蒙古人种基因。

2. 中华民族包含南北两大群体

吴国光等对包括 4707 人的南方汉族人群、2258 人的北方汉族人群在内的 6965 份群体用序列特异性引物聚合酶链反应 (PCR-SSP)、DNA 测序及分子克隆等以 DNA 为基础的 HLA 基因分型技术，对 HLA-A、B 及 DRB1 基因进行了多态性分析。结果发现南北汉族个体之间有 6 个基因有明显差异：除了 A * 11 基因在南方人群中频率明显高于北方人群 (31.7% 和 17.8%)

外, A*1、A*3、A*31 和 A*32 的基因频率在北方人群中均显著高于南方群体; 对 HLA-B 座位基因频率调查发现, 南北群体中共有 21 个基因存在明显差异, 北方人群中基因频率高于南方人群的有: B*7、B*13、B*37、B*2、B*44、B*48、B*50、B*51、B*52、B*57、B*1501、B*6701、B*1503, 其中差异最显著的有 B*13、B*44、B*50、B*52、B*57。南方人群中基因频率高于北方人群的有: B*38、B*46、B*5、B*58、B*4001、B*1502 和 B*1512, 其中差异最显著的为: B*46 及 B*4001。HLA-DRB1 座位的基因频率同样在南北人群中分布有明显差异, 其中差异最显著的是 DRB1*01 和 DRB1*07。这两个基因在北方人群中显著高于南方人群, 而 DRB1*09、DRB1*16 及 DRB1*0301 则南方人群明显高于北方人群。

1991 年对我国 7 个少数民族和 2 个南北汉族样本的 HLA 多态性分析, 进一步揭示南方少数民族(苗族和布依族)与南方汉族聚类, 北方少数民族(内蒙古的蒙古族、满族和回族)和西南地区的藏族则与北方汉族聚类。调查发现, 高加索人种起源的 HLA 抗原在维吾尔族中频率最高, 如 HLA-A3 (22.1%) 和 B8 (9.0%), 其基因频率自北向南递减, 而东南亚蒙古人种起源的 HLA 抗原在维吾尔族中频率最低或接近最低, 如 A11 (7.4%) 和 B46 (2.0%), 其基因频率自南向北递减。通过对我国南北汉族和傣族、布依族、藏族样本 HLA III 类座位 Bf、C4A 和 C4B 等位基因频率分布调查同样也发现, Bf、C4A 和 C4B 座位基因频率在我国南北群体间构成鲜明对比, Bf*S 南方汉族高于北方汉族, Bf*F 则相反。藏族 Bf 位点基因频率十分接近于北方汉族, 布依族和傣族 Bf 座位基因频率非常接近于南方汉族。C4A*3、C4B*1 和 C4B*5 藏族高于布依族, C4A*2、C4A*4 和 C4B*Q0 则布依族高于藏族。这些结果与 1986 年对 10 个汉族群体调查及 1991 年对 7 个少数民族调查结果一致, 提示中华民族包含华北和华南两大群体, 有着南北两大发源地。赵桐茂等对中国人免疫球蛋白同种异型的研究中, 也提出了中华民族起源于古代分别栖居于黄河流域和长江流域两个不同群体的假设。聂晓波等对我国 Bf 座位同种异型多态性分布调查研究的结果, 同样也支持中华民族可能有南北两个发源地的假设。杜若甫等用大量基因频率数据分析证明(38 个基因座, 30 个不同地区汉族和 37 个少数民族), 今天中国汉族与少数民族都分为南方蒙古人种和北方蒙古人种两大类型, 以长江为界, 并且从遗传学角度证明各

地汉族与当地少数民族血缘相近,说明各地汉族中已融入大量当地少数民族血缘,同样,各地少数民族也融入了部分汉族血缘。当然,蒙古人种起源和迁徙有待于更多的考古新发现,并结合语言学、人类学、历史地理学、历史学的研究成果才能得出正确结论。

第四节 线粒体 DNA 系统

线粒体 DNA 存在于胞质细胞器线粒体中,是一个闭合环状、长约 16.6 kb、遵循母系遗传的核外遗传大分子。线粒体 DNA 编码 13 种线粒体氧化磷酸化呼吸链组成酶的多肽,12S rRNA、16SrRNA 和 22 个 tRNA。基因间存在的非编码序列少,其中比较重要的是 mtDNA 复制的控制区(control region),俗称 D 环(displacement loop, D-Loop)。mtDNA 控制区序列长度约为 1.1kb,是 mtDNA 基因组中进化速率最快、多态最丰富的区域,其中的碱基变异主要集中分布在两个长度约 300—400 bp 的区域,分别称为第一高变区(hypervariable segment I, HVS-I)和第二高变区(HVS-II)。

母系遗传、缺乏重组和高进化速率是 mtDNA 重要的遗传特性,也是它作为研究现代人类起源、人群迁徙历史和人群遗传多样性很有用的一个遗传标记的内在原因。母系遗传、缺乏重组使得我们能够从人群中分布的 mtDNA 世系来追踪人群的史前历史,推测人群间的母系基因交流和区域性人群的微分化(microdifferentiation);高的进化速率(约是核基因组的 10—20 倍)一方面为法医鉴定提供了可能和依据,另一方面也使得我们能够通过群体中 mtDNA 世系上积累的变异来比较群体间的遗传多样性差异,推测人群的史前扩张、迁徙和经受建立者效应(founder effect)发生的大致时间。

早在 1987 和 1988 年,就有学者分别对我国的汉族、哈萨克族、回族和维吾尔族等少数民族人群的 mtDNA 多态进行了研究,从实验方法和数据处理等方面都做了一些可贵的探索。通过对来自上海的汉族和新疆的回族、哈萨克族和维吾尔族等民族群体的新鲜胎盘组织匀浆破碎离心提取出 mtDNA,进而用 14 种限制性内切酶进行酶切,俞民澍等在当时发现了一批新的酶切类型,并且某些类型表现出民族或种族特异性分布。而且,这四个群体中 mtD-

NA 的遗传多样性分布情况也存在差异,汉族和回族群体的内部变异较小,维吾尔族和哈萨克族群体内部变异相对较大。根据酶切数据计算的遗传距离显示,维吾尔族和哈萨克族间遗传距离最小,汉族和回族间的距离次之,而汉族/回族和维吾尔族/哈萨克族这两组群体间的遗传距离远远大于这两组群体内的距离,所以在依据遗传距离构建的群体关系树中,维吾尔族和哈萨克族最先相聚成一支,而回族和汉族聚成另一支,最后这两大支再相聚。另外,他们的结果还显示中国人群中存有 mtDNA 古老的祖先类型,认为这是支持亚洲是现代人类起源地假说的一个间接证据。

1995 年晋华黔等选择 16 种限制性内切酶分析了贵州的汉族、苗族、布依族和水族 4 个民族群体胎盘组织提取的 mtDNA 遗传变异情况。结果检测到 31 种酶切类型,其中有 3 种他们认为是新发现的类型;水族群体的内部变异度最高,苗族次之,汉族和布依族较小且接近;在构建的这 4 个民族群体遗传关系聚类树中,苗、汉最先相聚,而后再与水族相聚,布依族聚在最外。相对于国内人群 mtDNA 研究的发展来说,同时期国际上对于人群 mtDNA 的研究工作发展非常迅速,并且工作的重点也转移到对 mtDNA 控制区的两个高变区序列进行直接测定和分析。当时研究工作的另一个重要方向是对 mtDNA 全序列进行多个片段的重叠扩增,进而进行高分辨率限制性酶切分析 (high-resolution restricted fragment length polymorphism)。到 20 世纪 90 年代末期,关于我国人群 mtDNA 的研究主要见于国外一些实验室对来自我国台湾、广东和香港等地样本进行的零星分析和报道。

由于 mtDNA 母系遗传、缺乏重组的遗传特性,决定了发生在 mtDNA 上的变异的积累遵循一定的规律。除平行突变 (parallel mutation) 和热点突变 (hot spot) 外,每个线粒体世系 (lineage) 上完整地保留有祖先型突变 (ancestral mutation) 和较近时期发生的突变。这些较近时期积累的变异就成为现今在各大洲人群 mtDNA 序列中观察到的人群特异性的位点变异。含有相同的特异变异位点的一组 mtDNA 世系就构成了所谓的单倍型类群 (haplogroup), 反映在构建的系统发生关系树中,这些 mtDNA 世系就会聚成一个单系群 (monophyletic clade)。通过比较不同人群中单倍型类群的地理分布情况,并对人群中的单倍型类群的年龄进行估计,我们可以推测出群体的地理分化和迁移模式,同时可估算出这些事件大致发生的时间。

对于中国人群中 mtDNA 单倍型类群较早的界定工作，可以追溯到 1992 年 Ballinger 等的工作。在他们研究的来自东亚和东南亚地区的 7 个群体共 153 个个体中，包含有 14 个祖籍福建和广东的马来西亚个体和 20 个在台湾收集的源自湖南的样本。他们的研究表明，蒙古人种可能在距今 59 000—118 000 年前起源于中国的南部。此后，Torroni 等对我国西藏的藏族人群进行了研究，界定了单倍型类群 F、G、A、B、D 和 C 等，这批数据在 Bandelt 等的文章中被重新加以分析。虽然这些单倍型类群的划分和界定工作涉及我国的部分人群，但代表性明显不够，很多类群的系统发育关系也不是很明确，而且也没有控制区的相关信息。这种状况在近期得到较大的改善。Yao 和 Kivisild 等对我国汉族地理人群中的 mtDNA 单倍型类群进行了较为系统的分析，建立了一个框架性结构，同时也描绘了我国人群中各单倍型类群初步的地理分布格局。

在所研究的民族群体中，mtDNA 单倍型类群的频率分布表现出较大的地理分化趋势。类群 B 和 Fla 的频率总体上由南到北递减，而类群 D4 则是表现出相反的分布趋势；类群 M9a、Z、C 等在北方群体中有高的分布；类群 R9a1、R9b 和 N9a 等在南方群体中分布较多，尽管总体上的频率不是很高。在汉族人群中，南方群体间各单倍型类群分布上的差异要大于北方群体间的差异，而且，年龄比较古老的类群在南方人群中分布的频率要明显高于北方人群。汉族地理群体中的这种单倍型类群分布格局进一步证实了汉族人群的形成是多源的基因库被同一个文化即汉文化同化和统率的结果。百越部落起源的民族群体和大部分的民羌部落起源的群体在母系遗传组分上也存在着明显的差异：百越部落起源的民族群体具有更多典型的南方群体中普遍存在的类群，氏羌部落起源的民族群体中含有较多的北方群体中普遍存在的类群。一些在历史上不断迁移的氏羌族源的民族群体，如拉祜族等，可能因群体长期处于不稳定状态，遗传漂变比较严重，导致群体中的古老族源组分会不断丢失，从而使得通过基因交流获得的当地人群的遗传组分得到放大，表现出高频率的分布。此外，该民族盛行族内通婚，且近亲结婚比较多，更容易导致群体的遗传多样性降低和丢失，因而在系统树中的聚类位置比较奇怪，拥有很长的枝长，但在 mtDNA 类群的总体分布上，拉祜族表现出更为接近典型南方人群的趋势。

作为衡量遗传多样性的指标，单倍型多样性可以很好地显示出一个群体中随机抽取两个个体属于不同单倍型的概率，而平均核苷酸差异数和核苷酸多样性则可以体现群体中序列的变异情况。综合以前的研究结果显示：A. 来源于云南和贵州省的民族群体的单倍型多样性差异较大。如云南的 14 个少数民族群体中，傣族、白族和纳西族等样本的多样性较高，而怒族、拉祜族和归属于拉祜族的苦聪人等群体样本的多样性较低；来自贵州的苗族、水族、布依族和瑶族 4 个群体中瑶族的多样性明显低于其他的三个样本；来自广西的壮族和侬佬族、新疆地区的回族、蒙古族、哈萨克族、维吾尔族和内蒙古地区的蒙古族、鄂温克族、达斡尔族以及朝鲜族群体的多样性则分别表现出相近的值。B. 内蒙古和新疆两地的蒙古族遗传多样性也相近，但青海蒙古族的多样性要低一些，这可能与其样本较小（15 个个体）有关；云南的藏族和青海的藏族的多样性表现出较大的差异，云南藏族群体的多样性较低，这与该群体主要来源于迪庆地区的一个存在近亲结婚的偏僻山村，可能存在较大的抽样误差且代表性不够有关。C. 百越部落历史起源的群体如傣族、水族、壮族和布依族等具有较高的遗传多样性，而氏羌部落历史起源的群体如傈僳族、怒族等的遗传多样性普遍较低，且变化较大。D. 相对于汉族人群的遗传多样性来说，少数民族群体中遗传多样性整体上变化较大，这和目前各民族人群的人口统计大小、群体间可能的基因交流情况，以及伴随群体迁徙和定居过程中的建立者效应和遗传漂变等影响相关。

第五节 药物代谢酶系统

药物通常是和特定的靶点结合而产生药物效应的，因此药物效应与靶部位的药物浓度有密切关系，而靶部位药物浓度又与血药浓度间存在着动态平衡。血药浓度不仅取决于药物剂量，而且还与各种药物代谢动力学参数密切相关。由于药物代谢动力学参数常存在不同程度的个体差异，因此仅通过药物剂量往往不能有效地把血药浓度控制在一定范围之内，也就是说无法真正实现个体化给药。因此如能了解个体的药物代谢动力学参数，就能有效地预测靶部位的药物浓度，使个体化给药成为可能。在各种药物代谢动力学参数

中,药物的代谢速率是影响血药浓度的最主要因素之一。药物的代谢速率和代谢方式主要是由特定的药物代谢酶决定的,这是因为药物代谢绝大部分是在特定的药物代谢酶催化下完成的。现已知大多数药物代谢酶的催化活性存在明显的个体差异,因此对药物代谢酶的研究引起了广泛的兴趣。尤其是随着人类基因组计划的完成和相关学科的飞速发展,对药物代谢酶的研究和了解也达到了前所未有的深度,也使得药物遗传学(pharmacogenetics)、药物基因组学(pharmacogenomics)、药物蛋白组学(pharmacoproteomics)等新学科应运而生。现有研究已经证明药物代谢酶活性的个体差异多数与编码酶蛋白的基因多态性有关,基因多态性可从量上甚至质上影响这些基因所编码蛋白的功能,进而产生药物效应的差异。

根据药物代谢酶催化底物能力的不同,在表型上可分为快代谢型(extensive metabolizers, EM)、中间代谢型(in termedia te meta bolizers, IM)和慢代谢型(poor metabolizers, PM)。一般情况下,快代谢型表现为药物在体内的代谢速率加快、曲线下面积(area under the curve, AUC)缩小、半衰期缩短、血药浓度降低,这使得常规剂量给药时靶位点不易达到有效浓度而导致治疗失败;相对于快代谢型而言,慢代谢者表现为清除率降低、AUC增加、半衰期延长、易出现药物蓄积引起的各种 ADR。Phillips 等(2001 年)研究发现,59%易出现 ADR 的药物的代谢酶活性存在明显个体差异,而较少出现 ADR 的药物中只有 7%的药物的代谢酶活性存在类似的个体差异,临床上常见的 A 型 ADR 多与药物代谢酶的多态性有关。

药物在机体内的代谢主要有两种类型的生化反应,I 相反应(Phase I reaction)包括氧化、还原、水解反应等;II 相反应(Phase II reaction)主要是结合反应,通过和体内物质结合生成极性产物,以利排出体外。相应的将这些酶称为催化 I 相反应的药物代谢酶和催化 II 相反应的药物代谢酶。

催化 I 相反应的药物代谢酶主要有细胞色素 P450 酶系(cyto chromes P450 enzymes, CYP450)、还原酶(reductase)、氧化酶(oxidase)、酯酶(esterase)等,其中最重要的是 CYP450。已证明存在遗传多态性的 I 相药物代谢酶主要有属于 CYP450 的 CYP1A1、CYP1A2、CYP2A6、CYP2B6、CYP2C8、CYP2C9、CYP2C18、CYP2C19、CYP2D6、CYP2E1、CYP3A4、CYP3A5 等;属于氧化酶的有乙醇脱氢酶(alcohol dehydrogenase)、乙醛脱

氢酶 (aldehyde dehydrogenase)、单胺氧化酶 (monoamine oxidase)、过氧化氢酶 (catalase)、超氧化物歧化酶 (superoxide dismutase)、三甲胺-N-氧化酶 (trimethylamine N-oxidase)、二氢嘧啶脱氢酶 (dihydropyrimidine dehydrogenase); 属于酯酶的有丁酰胆碱酯酶 (butyrylcholinesterase)、芳香基酯酶 (paraoxonas); 属于还原酶的有葡萄糖-6-磷酸脱氢酶 (glucose-6-phosphate dehydrogenase)、NAD (P) H: 醌氧化还原酶 [NAD (P) H: quinone oxidoreductase] 等。

催化 II 相反应的代谢酶主要有 N-乙酰转移酶、甲基转移酶、尿苷二磷酸、葡萄糖醛酸转移酶、磺基转移酶和谷胱甘肽-S 转移酶等。

1. CYP2A6

人类 CYP2A 亚家族中已发现的基因有: CYP2A6、CYP2A7、CYP2A13 等, 其中研究最多的是 CYP2A6。CYP2A6 主要在肝脏表达, 约占肝脏中 CYP450 总量的 4%, 可被苯巴比妥、地塞米松 (dexamethasone)、利福平 (rifampicin) 等诱导, 它的抑制剂有毛果芸香碱 (pilocarpine)、8-甲氧补骨脂素 (8-methoxypsoralen) 等。它的底物有香豆素、法罗唑啉 (fadololol)、洛加酮 (loxigamone) 等药物及前致癌物等, 特别是 70%—80% 的尼古丁由 CYP2A6 代谢灭活。目前已发现 20 多种 CYP2A6 的等位基因, 中国人中常见的非野生型等位基因频率见表 6-3, 和其他种族相比有明显的差别。中国人中频率较高的基因型依次是 CYP2A6 * 1/CYP2A6 * 1 (77.2%)、CYP2A6 (1/CYP2A6) * 4 (11.4%)、CYP2A6 * 1/CYP2A6 * 8 (7.0%)、CYP2A6 * 7/ CYP2A6 * 7 (1.8%)。

表 6-3 中国人中常见的 CYP2A6 等位基因及其频率

等位基因	碱基变化	氨基酸变化	功能变化	分布频率
CYP2A6 * 1B/C	-395G>A, 51A>G			0.4
CYP2A6 * 2	T479>A	160Leu→His	无	0.0
CYP2A6 * 3	CYP2A6/ CYP2A7 杂合		?	0.0
CYP2A6 * 4	基因缺失		无	6.6
CYP2A6 * 7	T1412>C	471Ile→Thr	降低	2.2
CYP2A6 * 8	G1454>T	485Arg→Leu	正常	3.5
CYP2A6 * 10	T1412>C 1454G>T	Ile471→Thr 485Arg→Leu	降低	0.4

和野生型相比, CYP2A6 * 7、CYP2A6 * 10 的酶活性显著降低, 这可能与 471 位氨基酸从 Ile 变成强极性的 Thr, 影响了肽链的折叠有关。CYP2A6 * 8 酶活性与野生型没有明显差别, CYP2A6 * 2 和 CYP2A6 * 4 则没有明显酶活性。目前关于 CYP2A6 各种等位基因对药物代谢影响的研究较少。

2. CYP2C9

基因 CYP2C 定位在 10q24.2, 是哺乳动物中最大的亚家族, 至少已发现近 30 个成员, 在人类至少已鉴定出 6 个 CYP2C 的 cDNA: CYP2C8、CYP2C9、CYP2C10、CYP2C17、CYP2C18、CYP2C19。其中最重要的是 CYP2C9。

CYP2C9 是 CYP2C 亚家族的一员, CYP2C9 和 CYP2C10 仅有 2 个氨基酸不同, 且二者的催化活性极为相似, 故一般常把 CYP2C9 和 CYP2C10 统称为 CYP2C9/10 或 CYP2C9。CYP2C9 含有 9 个外显子和 8 个内含子, 全长约 5.5 kb。在 CYP2C9 的 5' 上游 2.2 kb 内含有几个与糖皮质激素效应元件 (glucocorticoid response elements, GRE) 一致的序列, 另外在起始密码子上游 57 bp 和 110 bp 处分别有一个 TATA 和 CAAT 盒。在 5' 调控区至少存在 7 个 SNP, 它们可能影响 CYP2C9 的转录, 其中位于 HNF-1 (hepatic nuclear factor-1) 结合位点的 T-1912C 参与了基因转录的激活。另外, 第一外显子内的 6 bp 重复也参与了基因表达的调控。若缺失从 -155 到 0 这段序列, 基因转录水平可降低至 1/8~1/7, 提示可能有重要的未知调控元件位于该区间。CYP2C9 编码的酶蛋白分子量为 53KD, 含有 490 个氨基酸。CYP2C9 在人类肝微粒体内含量丰富, 约占 CYP450 总量的 20%, 仅次于 CYP3A。临床上约有 10% 的常用药物由 CYP2C9 催化代谢, 常见的底物有磺胺类、氨基比林、双香豆素、氯霉素、西咪替丁、咪唑类抗真菌药、甲苯磺丁脲等。CYP2C9 常用的工具药为华法林。

CYP2C9 主要的多态性位点有: C430T (Arg144C)、A1075C (Ile359Leu)、C1080G (Asp360Glu), 它们的频率存在明显的种族差异。144Arg、144Cys 在中国人、日本人、黑人、白人中的频率分别为 100%、0%, 100%、0%, 96.8%、3.2%, 88.7%、11.3%。359Ile、359Leu 在中国人、日本人、黑人、白人的频率分别为 96.7%、3.3%, 97.8%、2.2%,

98.3%、1.7%，92.6%、7.4%。CYP2C9 已发现 12 种等位基因，重要的有 3 种：野生型 CYP2C9 * 1、突变体 CYP2C9 * 2 (430C→T, 144Arg→Cys) 和 CYP2C9 * 3 (1075A→C, 359Ile→Leu)。中国人群主要的基因型是 CYP2C9 * 1/* 1 (96.5%)、CYP2C9 * 1/* 3 (3.5%)。144 位氨基酸残基虽不在公认的底物识别位点区域内，但却紧靠第一个识别位点 SRS-1，在包括 144 位氨基酸残基在内的区域内有一 α 螺旋结构，当 144 位氨基酸发生由 Arg 到 Cys 的突变时，这种 α 螺旋结构遭到破坏，导致 CYP2C9 * 2 蛋白结构和功能发生改变，从而改变了酶的催化活性。359 位氨基酸位于底物识别位点区域内，该区域有一 β 折叠结构，当 Ile 突变成 Leu 时破坏了这个结构的稳定性，造成酶对底物的亲和力降低，使酶活性下降 90%。临床上表现为药物的有效剂量与 CYP2C9 的活性密切相关，具体到抗凝药华法林而言，CYP2C9 * 3 纯合子患者每天的治疗量仅需 0.5 mg 消旋华法林，约为野生型病人的十分之一。

3. CYP2D6

CYP2D 是第一个被发现存在遗传多态性的 CYP450，目前哺乳动物体内至少发现了 21 种 CYP2D。而人类的 CYP2D 位于 22q13.1，由 CYP2D6、CYP2D7P 和 CYP2D8P 共同构成，其中 CYP2D7P 和 CYP2D8P 为假基因，仅 CYP2D6 有功能蛋白表达。CYP2D6 主要在肝脏中表达，CYP2D6 共有 497 个氨基酸组成。CYP2D6 不仅氧化代谢某些内源性的类固醇激素，而且还是 20%~25% 常用药物的主要代谢酶。CYP2D6 的底物包括 β 受体阻断剂普萘洛尔、抗心率失常药奎尼丁、普罗帕酮、降压药异喹胍、镇痛药曲马多、抗精神病药奋乃静、氟哌丁醇 (haloperidol) 等，止咳平喘药可待因 (codeine)、右美沙芬等。

CYP2D6 基因型分布表现出明显的种族差异，包括人群中特殊的基因突变和种族特异基因突变，如白种人的 CYP2D6 * 4、亚洲人的 CYP2D6 * 10 和非洲人的 CYP2D6 * 17，CYP2D6 * 14 突变基因首先在中国人中发现，之后的研究发现该基因型仅在亚洲人中存在，而 CYP2D6 * 3、* 6、* 8 在中国人群中未见报道。常见的 CYP2D6 突变等位基因及其在不同种族之间的分布频率见表 6-4。

表 6-4 常见的 CYP2D6 等位基因在白种人、非洲黑人及亚洲人中的分布频率

基因型	CYP2D6 等位基因	白种人 (%)	非洲黑人 (%)	亚洲人 (%)
功能型	* 1	33-40	28-50	23-42
	* 2	22-34	11-78	9-20
功能下调型	* 9	0-2.9	0	3.3
	* 10	1.9-8	3.1-8.6	38-70
	* 17	0.1-0.3	9-34	0.5
非功能型	* 41	8	-	-
	* 3	1-3.9	0-0.5	0.8-1
	* 4	12-23	1.2-7	0-2.8
	* 5	1.6-7.3	0.6-6.1	4.5-6.1
	* 6	0.7-1	0	-
复制型	* 1×2	0.2-0.5	3.3	0.5
	* 2×2	0.7-1.6	1.6-2.5	0-1
	* 4×2	0.1-0.2	0.9	-

CYP2D6 基因突变的发生主要是因为碱基的缺失、插入导致阅读框架移位或碱基替换引起氨基酸的改变，甚至是大片段基因的丢失以及基因的复制或扩增。

碱基的插入及替换通常引起读码框移码，从而导致基因突变的发生。CYP2D6 * 4 是 70% 白种人酶活性降低的主要原因，其第 3 号内含子 3' 末端 1934 位的 G 替换为 A；CYP2D6J 基因在外显子 1、9 分别有 100C>T 和 4268G>C 突变；CYP2D6 * 40 是在第 4 外显子内插入了 9bp 碱基，同时有 1023T、1661C 和 2850T 三个点突变；CYP2D6 * 69 为外显子上出现 4 个位点的突变，分别是第 1 个外显子的 100C、第 2 个外显子的 1062G、第 6 个外显子的 2850T 以及其末端 2988A，其中包括与 * 10 和 * 41 重合的重要突变点，其标志性突变为 1062A>G，* 69 据此标志性突变点可以和 * 10、* 41 区分。

常见出现氨基酸取代的等位基因有 CYP2D6B、CYP2D6 * 10 和 CYP2D6 * 17、* 45、* 46 等。CYP2D6B 是第 4 内含子的接口处 1934 位 G>A 突变，使 RNA 剪切改变，引起氨基酸的替换；CYP2D6 * 10 是亚洲人常见的单核苷酸多态性等位基因，它的突变点分别是第 1 外显子的 100C>T 突变，造成 Pro→Ser 氨基酸取代及第六外显子的 4268G>C 突变，造成 Ser486→Thr 氨

氨酸取代,二者均降低 CYP2D6 的酶活性; CYP2D6 * 17 是另一重要的 CYP2D6 的突变体,它携带了三个无意义的编码区域,出现 Thr107→Ile、Arg296→Cys、Ser486→Thr 氨基酸取代,该基因型在非洲人中的分布频率约是 21%。突变体 * 45、* 46 均有外显子 3 上 Glu155→Lys 氨基酸取代及外显子 5 上的 2575C>A 无义突变,* 46 因携带有另外一个氨基酸取代 (R26H) 而与 * 45 区分,尽管 * 43 也有该取代,但与 * 46 相比,* 43 缺少其他单核苷酸多态性位点。

CYP2D6 * 3 的突变是由于外显子 5 上的 2637A 丢失导致阅读框移码,使得翻译提前终止而造成酶蛋白的合成障碍; CYP2D6 * 5 是由于染色体上基因大片段丢失,造成 CYP2D6 酶不完全表达; CYP2D6 * 45B 虽然有 A * 692 (TGTG) 的丢失,却对基因的表达没有显著的影响。

Cai 等在非洲裔美国人中发现多重复制型等位基因: * 4Dxn、* 10Bxn、* 17xn、* 41xn 和 * 2Lxn,其中 * 17xn 和 * 2Lxn 基因型属于首次发现。* 41xn 和 * 2Lxn 均在启动子区域有一 1584C/G 突变,但在 * 41xn 的内含子 1 区发现了一个基因转换单元,据此可将二者区分开来。* 10Bxn 突变位点分别位于外显子 9 (4180C) 和外显子 1 (100T),虽然这两个位点与 * 4 的突变位点相同,但前者在内含子 3 与外显子 4 的边缘没有 1846A 点突变。

其他突变体如 CYP2D6 * 62 仅是碱基序列与 CYP2D6 * 1A 的序列不同 (R441C),这一罕见突变体在白种人中的突变频率不到 0.1%,且研究发现该基因型为非功能性的等位基因。而 * 64、* 65 均为 CYP2D6 等位基因的杂交体,如 * 64 是 * 10 和 * 17 的杂交体,* 65 是 * 10 和 * 2 的杂交体。二者对 CYP2D6 酶功能的影响并没有确定性的结论,据推测 * 64 对 CYP2D6 酶功能影响类似于 * 10 和 * 17。在此,我们将 CYP2D6 基因常见等位基因型及其突变机制列表 6-5 中。

表 6-5 常见功能性的等位基因型的突变机制和突变结果

等位基因	碱基突变	氨基酸的改变	突变位点 (外显子)	突变结果 (酶活性)
* 1	—	—	—	野生型
* 2	2850C>T	R296C	外显子 6	酶活性不改变
	4180G>C	S486T	外显子 9	—

续表

等位基因	碱基突变	氨基酸的改变	突变位点 (外显子)	突变结果 (酶活性)
* 5	—	染色体基因丢失	—	酶活性丧失
* 10	100C>T	P34S	外显子 1	酶活性降低
	4180G>C	S486T	外显子 9	酶活性降低
* 14	G124 (212) A	P34S	外显子 1	酶活性降低
* 18	4125 — 4133 (4213 — 4221) insGTGCCCACT	468—470ins VPT	外显子 9	酶活性降低, 甚至丧失
* 21	2573insC	273ter	外显子 5	酶活性降低, 甚至丧失
	2850C>T	267Frameshift	外显子 6	酶活性丧失
	4180G>C	267Frameshift	外显子 9	酶活性丧失
* 27	3853G>A	E410K	外显子 8	—
* 36	2D6 与 2D7 换位	7 个氨基酸改变	外显子 9	—
* 39	4180G>C	S486T	外显子 9	—
* 44	2950G>C	—	外显子 6	酶活性丧失
* 47	73C>T	R25W, P34S	外显子 1	—
	4180G>C	S486T	外显子 1	—
* 48	972C>T	A90V	外显子 2	—
* 49	100C>T	P34S	外显子 1	—
	1611T>A	F120I	外显子 3	—
	4180G>C	S486T	外显子 9	—
* 50	1720A>C	E156A	外显子 3	—
* 51	2850C>T	R296C	外显子 6	—
	3172A>C	E334A	外显子 7	—
	4180G>C	S486T	外显子 9	—
* 53	1617G>T	F120I	外显子 3	—
	1611T>A	A122S	外显子 3	—
* 54	100C>T	P34S	外显子 1	—
	2556C>T	T261I	外显子 5	—
	4180G>C	S486T	外显子 9	—
* 55	2850C>T	R296C	外显子 6	—
	3835A>C	K404Q	外显子 8	—
	4180G>C	S486T	外显子 9	—
* 57	100C>T	P34S	外显子 1	—
	887C>T	R62W	外显子 2	—
	2D6 与 2D7 换位	7 个氨基酸改变	外显子 9	—

注：表中列出的是亚洲常见功能性的等位基因型及分布频率。沉默突变、启动子突变和内含子突变未列出。

CYP2D6 代谢表型多态性是由其基因多态性决定的。根据基因突变对 CYP2D6 酶活性的影响,一般可将 CYP2D6 的等位基因及相应的代谢表型划分为以下几类:

泛代谢型 (Extensive Metabolizers, EMs): EMs 发生通常是因为携带有 CYP2D6 的野生型等位基因,是纯合子等位基因产生的正常酶的表达,如:

* 1, * 2, * 35。

弱代谢表型 (Poor Metabolizers, PMs): 包括 CYP2D6 * 7、* 8、* 11、* 12、* 13、* 14、* 15、* 16、* 18、* 19、* 20、* 21、* 38、* 39、* 40、* 42、* 44 和 * 36x2。PMs 发生通常是因为 2 个非功能性等位基因的表达引起了酶活性的丧失。而非功能性等位基因又称为无效等位基因,如 CYP2D6 * 3、* 4、* 5 和 * 6,它们分别占无效等位基因的 71%、16%、6% 和 4%。

中间代谢型 (Intermedia Metabolizers, IMs): IMs 发生是因为同时携带了一个 CYP2D6 无效等位基因和一个损坏的等位基因,致使酶活性降低,包括 * 9、* 10、* 17、* 29、* 36、* 40、* 41 (?), * 43、* 45 和 * 46。

超快代谢型 (Ultrapid Metabolizers, UMs): UMs 发生通常是由于 CYP2D6 的多基因拷贝,使酶蛋白高表达,导致酶活性的显著增高。包括 * 1xn、* 2Axn 和 * 35xn。这种表型在瑞典人中的携带频率约为 1%,西班牙人为 7%,黑人为 29%。需要特别注意的是,有的 CYP2D6 等位基因发生多重复制但并不导致 UMs 型的产生,如 * 36x36。

CYP2D6 代谢表型多态性的发生也存在着明显的种族差异。在白种人中 PMs 的发生频率是 5%~10%,阿拉伯人(约旦人)为 7.1%,非洲黑人 0~2%、东亚人 1~2% (包括中国人、日本人、韩国人),在印度尼西亚人,僧伽罗族人和 Cuna A merin—dinas 人中至今未检测到 CYP2D6 的弱代谢型。在中国人群中,不同民族间 PMs 的发生频率也有一定差异,如汉族为 0~1%、藏族 1.52%、维吾尔族 0.63%、蒙古族 0.81%、侗族 0.8%、苗族 0。UMs 主要分布在东部非洲及西班牙人群中,在白种人、亚洲人及非洲西部少见分布。通常 UMs 在瑞典人中的携带频率约为 1%,西班牙人为 7%,黑人为 29%。IMs 主要分布在亚洲人群中,如中国人群中常见的基因突变型 CYP2D6 * 10,其发生频率为 64%~70%。

第七章 中华民族的源与流

第一节 东亚人的起源

关于现代人的起源问题，一直是学术界争论的焦点。近几年来，随着分子遗传学研究手段的深入及考古学新证据的不断被发现，现代人起源问题的研究正逐步走向系统化。目前的学术界存在两大学派，“非洲起源”学派和“多地区起源”学派。在近20年内，两大学派为证明各自的主张展开了激烈的学术争论，这一争论目前还在继续。随着争论的不断升级，问题中存在的许多不明因素也变得越来越明朗，现代人的起源和史前迁徙路线被清晰地勾画出来。在人群起源和史前迁徙研究方向上，分子遗传学、考古学和古DNA等研究手段正逐步成熟，能够将三者很好地结合起来，形成了一个很有效的综合研究手段。正因为如此，仅有少数考古学家还坚守的“多地区起源”学说的呼声变得越来越微弱，现代人“非洲起源”学说逐步被广大的学者所接受。

各方面的研究证据都支持直立人起源于非洲，大约在100万年前直立人走出非洲扩散到欧亚大陆，目前这一结论在学术界已经达成共识。但是，关于直立人在哪里进化成为现代人这一问题上却存在很大分歧，学术界基本上分化为两大阵营，即“走出非洲”学派和“多地区起源”学派。“走出非洲”假说认为直立人在非洲进化成为现代人，地球上所有的现代人类都来源于非洲，现代人的祖先大约在10万年前走出非洲之后，完全取代了世界各地的直立人并繁衍下来。该假说自20世纪80年代末被提出之后被越来越多的遗传学和考古学证据所支持，并且备受广大学者认同。“多地区起源”假说主张直立人进化为现代人不仅发生于非洲，走出非洲之后扩散到东亚地区的直立人后裔也

独立进化成为东亚现代人。因为在中国以及东亚的其他地方发现了很丰富的人类化石。这些人类化石，不管从形态特征的过渡上还是从空间与时间的分布格局上，东亚地区的人类进化看似一个完整的从直立人到现代人的连续进化过程，持这一观点的主要是一些考古学家。目前，两大学派都在寻找各种证据，为自己支持的理论提供支持。

几乎所有的遗传学证据都支持现代人起源的“走出非洲”学说。大量的线粒体 DNA (mtDNA) 的研究证据虽然支持“走出非洲”学说，但在该遗传证据的解释上存在矛盾，甚至冲突。此外，在过去大部分的遗传学研究中，中国人的样本主要来自于广东。虽然广东存在许多少数民族群体，但是样本的代表性还不足以对中华现代人存在独立起源的学说形成全面的批驳。目前，对欧洲、美洲和大洋洲等地区的现代人的起源及史前迁徙路线的研究已经很清楚，但是对于中国现代人的来源及早期迁徙路线研究还是一个谜。中华现代人的起源和史前迁徙问题的解决，对解决目前两大学派之争起决定性作用。

对于中国现存人群遗传结构的研究，最初开展的是采用传统的遗传学标记来进行，如血型、免疫球蛋白和等位酶等。中国各个人群之间尽管几千年以来存在共同历史及大规模的内部迁徙，但是南北人群之间还是存在显著的遗传差异。这就引出了一个问题，中国现存的南北人群究竟是同一人群的后代还是不同来源人群的后代。

Y 染色体遗传标记一直没有被应用于重建人类进化历史，主要的障碍是因为这些遗传标记的多态性很低，难于发现。然而，一种新的高效检测突变位点技术——变性高效液相色谱的出现扫除了这种障碍，许多 Y 染色体双等位遗传标记在近几年被陆续发现。

1999 年，宿兵等采用 Y 染色体遗传标记系统研究东亚人群中父系遗传结构，该研究所采用的一套共 19 个双等位 Y 染色体遗传标记是由斯坦福大学发现的，主要适用于研究东亚人群。该研究收集大量东亚群体，包括 21 个中国少数民族群体、22 个省区的汉族群体、3 个东北亚和 5 个东南亚群体，以及 12 个亚洲之外群体（来自非洲、美洲、欧洲和大洋洲）。对大量东亚代表性群体通过高分辨率的 Y 遗传标记分析，得出结论为：东亚南方人群的遗传多样性比北方人群高，揭示出东亚大陆南部为现代人进入东亚大陆后最早的居住地。同时，根据 Y 染色体微卫星重复次数变异的方差估计出现代人最早到达

东亚南部的时间为 1.8 万~6 万年以前, 研究认为东亚现代人在到达东亚大陆南部以后, 从南向北扩散到中国北部甚至西伯利亚, 在时间上与最后一次冰期的衰退相符合。

“走出非洲”学说在两个方面受到来自“多地区起源”学派支持者的批评。其中一个就是多地区起源学派认为“多地区起源”学说目前被误解, 主张在直立人在 100 万年前走出非洲之后, 不同地区之间的群体仍然进行频繁的基因交流, 导致现代人共享多地区进化的结果。他们认为目前支持“走出非洲”学说的大量遗传学证据也能在“多地区起源”学说框架下解释清楚。但是, 这一理论无法对 Y 染色体单倍型数据进行解释。不同 Y 单倍型在人群中的分布具有地区特异性, 只支持现代人“走出非洲”学说, 没有发现非洲祖先之外的本地祖先成分, 并且所有的非洲之外发生的突变都晚于距今 5 万年。假设在过去 100 万年中不同地区之间存在广泛的基因交流, 那么在现存非洲人群中古老的 Y 单倍型也能在东亚人群中发现, 但事实并不是这样。

“多地区起源”学派支持者的另一个批评来自于东亚地区的考古学证据, 在东亚地区尤其是中国, 发现了一系列从直立人到现代人的连续的进化过程的化石证据, 被认为是支持东亚现代人独立本地起源的有力证据。经过仔细分析发现, 在中国发现的人类化石在时间上并不具备所谓的连续性, 所有的直立人化石都早于距今 10 万年, 而所有的现代人化石都晚于距今 4 万年(大部分处于距今 3 万~1 万年之间), 东亚地区的人类化石在距今 10 万~4 万年之间处于一个空白, 这是一个不规则的现象。而且, 在中国发现的存在距今 10 万~4 万年空缺的人类化石之间有一个很大的跨度, 早于距今 10 万年的化石与晚于距今 4 万年的化石之间在形态上存在显著的差异, 这个鸿沟不能被粗放地连续起来。这种现象合理的解释就是, 东亚当地的古人灭绝了, 后来来自非洲的现代人到达东亚, 并完全取代了东亚古人的位置。巧合的是, 这 6 万年的空缺时间正好是最后一次大冰期, 在最后一次冰期中, 非洲起源的现代人已经来到了东亚大陆的南部。

“走出非洲”假说的一个重要补充就是要清楚地勾画出现代人走出非洲之后沿什么路线迁徙到达东亚。关于东亚现代人史前迁徙的研究中发现一个有趣的现象, 东亚现存的现代人在南北人群之间存在遗传上和形态上的显著差异。对这一现象目前存在三种解释。第一种解释认为现代人是从北部进入东

亚，早期的迁徙是由北向南，最后与澳洲土著会合。第二种解释认为现代人是从南部进入东亚大陆，然后由南向北迁徙。第三种解释是前两种解释的折中，认为东亚现代人是沿南北两条线独立进入东亚，最后在东亚汇合。对于这三种解释，第一种认为东亚人与澳洲土著有混合，这种事件几乎不可能发生，因为 Y 单倍型的研究发现，东亚人群中特异的 Y 单倍型在澳洲土著群体中没有发现，H17Y 单倍型是澳洲人群特异的单倍型，但在东亚人群中没有被发现。大量东亚人群的 Y 染色体研究发现，北部人群具有的单倍型只是南部人群单倍型中的分支，南部人群拥有更复杂的单倍型组，认为这一现象为早期人群由南向北迁徙过程中产生瓶颈效应的结果。第三种解释看似合理，但在最近的研究中发现这种解释也是不太可能的。东亚现代人最早到达的是东亚南部，主要的迁徙路线是由南向北，这一历史时间得到了来自 Y 染色体和 mtDNA 遗传证据的支持。

东亚现代人群中一个主要的东亚特异 Y 单倍型组是 M175（约占东亚 Y 的 80%），在这一大单倍型组中 M122 组最大（约占东亚 Y 的 45%），通过对东亚 2333 份男性 Y 的 M122 单倍型组的研究发现，东亚南方人群的多态性比北方人群高，用微卫星数据构建的网络图可以看出北方人群为南方人群的分支，支持东亚人最早到达东亚南部大陆，在 4 万~2 万年前由南向北扩散。在 M175 分支下还存在另外两大分支：M119 和 M95，M95 仅见于东亚南方人群中，北方人群几乎很少发现，M119 在东亚南方群体中普遍存在，北方人群的频率很低，可能也是由南向北扩散过程造成的结果。

Yap+ 单倍型在世界人群中广泛存在，该单倍型分为两大支，所有东亚的 Yap+ 都属于 M174 这一支（H2 和 H3），为东亚人群的特有单倍型，而其他地区的 Yap+ 为 M40/M96 这一大支，其下面又分为各个地区特有单倍型。东亚的 Yap+ 在 M174 突变下又分为三支，一支带有 M55 等突变，为日本人特异单倍型，另外一大支带有 M15 突变，为中国大陆人群的特异单倍型，主要分布在藏族群体和中国南方的少数民族群体（苗瑶支系及壮侗支系）中，东亚大陆的其他人群中也有少量分布。在东亚大陆上分布的 Yap+ 还有一支为只有 M174 突变，没有 M15 和 M55 突变，这一单倍型主要分布在藏缅语族的各个民族以及汉族人群中，苗瑶和壮侗语族的各个群体也有部分分布。在东亚现存人群中，Yap+ 的频率呈现出不连续的分布，分布频率除了藏族（40%

~65%) 和日本人 (30%~50%) 较高, 在苗瑶人群中也很高 (15%~50%), 占东亚人口主体的汉族人群的 Y_{ap+} 频率很低, 蒙古人群和东南亚人群仅有少量分布, 西伯利亚人群没有发现该单倍型组。根据目前的数据我们猜测, Y_{ap+} 也是由南部进入东亚大陆, 之后的人群由南向北迁徙把这一单倍型带到北方。汉族人群是以 M122 突变为主的群体, 近 3000 年来, 汉族人群快速膨胀, 形成该地区最大的民族, 分布范围不断扩大并占据中国大陆的绝大部分。目前 Y_{ap+} 在东亚大陆上的分布格局可能就是由于汉族群体的扩大造成的残留分布, 大量的 M122 单倍型组的人群稀释了 Y_{ap+} 的频率。

来自非洲的现代人到达东亚大陆南部之后, 除了主要向北扩散外, 另外一部分向南开拓新的居住地, 这一史前迁徙历史得到遗传学证据的支持。这一部分南迁的人群占据东南亚之后, 他们的后裔在 3000 年前从印度尼西亚群岛出发扩散到太平洋诸岛 (包括美拉尼西亚、大美拉尼西亚、玻利尼西亚以及夏威夷群岛)。

第二节 中华民族的源流

根据很多对我国各民族与人群以各种不同指标所进行的遗传距离、主成分分析及遗传拓扑学分析, 对中华民族的源与流、各民族的遗传结构及彼此间血缘关系, 目前有如下结论。

从人种关系上讲, 中国各民族中, 俄罗斯族原属高加索人种, 但近一二百年来, 中国的俄罗斯族已与汉族等蒙古人种民族发生了血缘混杂。在中国俄罗斯人数虽不多, 可是居住分散, 且各地俄罗斯人群与其他民族的血缘混杂程度不一。因此, 不可能对俄罗斯民族中高加索人种血缘还占多少成分作出估计。在中国各民族中, 塔吉克族、塔塔尔族、柯尔克孜族、乌孜别克族等民族, 具有较多的高加索人种的血缘, 在迁入中国后, 与其他民族血缘混杂很少, 但其主要血缘成分属蒙古人种还是高加索人种尚有待进一步的研究, 以积累更多的研究资料, 才能作出判断。而维吾尔族虽也有一定的高加索人种血缘, 可是绝大部分是蒙古人种血缘, 因此可以肯定属蒙古人种。哈萨克族与回族等民族, 也有一定的高加索人种血缘, 但其成分比维吾尔族所含有

的还少。在全国范围内，高加索人种的血缘自西向东、自北向南越来越少，而最少的是在南方。可见，高加索人种基因流入中国，主要是从中国西部，从中国东部几乎没有，仅从东北由俄罗斯人带入少量高加索人种血缘。而虽然印度人也属高加索人种，离中国南方也较近，但却没有带入什么高加索人种血缘到中国南方人中。

在全国各地人群的遗传结构差异中，就少数民族而论，高加索人种与蒙古人种间基因流动所造成的差异占主要地位，把少数民族与各地汉族人群放在一起分析时也如此。仅仅就汉族而言，则在各地汉族人群间的遗传差异中，北方蒙古人种与南方蒙古人种间基因流动所造成的差异占主要地位。

除对新疆 3~4 个少数民族尚需进行深入一些的研究外，可以认为，中国人由北方蒙古人种与南方蒙古人种两个亚人种组成，即无论汉族与少数民族，都可分为南、北两大群，分别属于南方蒙古人种与北方蒙古人种这两个亚人种。南、北蒙古人种的分布，基本上是以长江为界。过去，有人提出以北纬 300 为界，虽然北纬 300 与长江中、下段相当接近，但是纬度是一条人为的直线，而长江却是隔离人群的天然屏障，所以以主张长江为界的提法更为合理。自重庆以西，则四川茂县羌族和阿坝州、甘孜州直至青海、西藏等地的藏族应属北方蒙古人种，成都市等地汉族仍是南方蒙古人种。而四川与陕西交界地区，南、北蒙古人种的分界线在何处，尚待深入研究。

过去，在文献中曾提出过蒙古人种东亚类型和南亚类型及北亚类型并列。现在看来，东亚类型并不单独存在，他们或者应归入南亚类型（即南方蒙古人种），或者应归入北亚类型（即北方蒙古人种），有的则是基本上属南方蒙古人种，但融入有北方蒙古人种一定的血缘；或原来起源于北方，但迁移到南方后已融入有南方蒙古人种的一定的血缘，已基本上属南方蒙古人种。还有人提出，中国人除南北两大类型外还有第三种类型，他们称之为“藏彝走廊类型”，包括从甘肃南部到云南西部南部属藏缅语族的各少数民族（保安族、东乡族、纳西族、藏族、羌族、白马藏族等）。他们认为有藏彝走廊类型这第三类型所依据的只是用人体测量数据进行的聚类分析，用其他遗传标记作依据进行聚类分析时，这一类型并不出现。

在各地汉族人群中，北方人群之间的遗传距离比南方人群之间的要近。这可能是因为在北方有不少广袤的平原，黄河冬天封冻而且经常改道，几千年

来多次因战乱与灾荒造成人口大迁移与民族大融合也发生在北方，因此北方的人口流动比南方更为频繁，也较容易。

长江以北的汉族人群属北方蒙古人种，长江以南的汉族人群属南方蒙古人种。这在以红细胞血型、白细胞血型、皮纹学指标、免疫球蛋白同种异型Gm因子、人体测量学，以及多指标为依据所进行的遗传距离与聚类分析的结果中，都明显地可以看到。

各地汉族在主成分分布图人群和二维同步排序图中，都可清楚地看出，在历史上黄河中游的华夏族曾向四周扩散。各地汉族人群与地理相近的汉族人群遗传距离最近，和当地的少数民族的遗传距离居其次，而与地理距离远的汉族人群、特别是南方汉族人群与北方汉族人群间的遗传距离更远。这充分说明汉族与本地的少数民族间已有了许多基因流动，他们的遗传结构已十分接近了。从历史记载看，这主要是少数民族融入汉族的结果，也有部分是汉族融入少数民族的结果。而且汉族今天之所以能成为世界上人口最多的民族，主要是不断地大量融入其他民族的结果。

最终的结论如下：

1. 蒙古族、朝鲜族、满族、锡伯族、赫哲族、达斡尔族、鄂伦春族、鄂温克族等是典型的北方蒙古人种民族。其中鄂温克与鄂伦春比较特殊，因为他们原来在更北方的在西伯利亚，后来才迁到黑龙江以南。

2. 我国回族没有统一的起源。各地回族融入有一定的高加索人种血缘，而且其成分多少不一。一般说来，西北地区宁夏等地的回族融入的高加索人种血缘较多。可是回族中融入的高加索人种血缘比维吾尔族、哈萨克族等民族少，比塔吉克族、塔塔尔族、柯尔克孜族、乌孜别克族、俄罗斯族等民族则更少。北方的回族肯定属北方蒙古人种，但福建、云南的回族有可能属南方蒙古人种，特别是海南的回族，肯定属南方蒙古人种。

3. 藏族肯定属蒙古人种，而且在平行排序图上可以明显地看出，藏族起源于北方。

藏族自北方向青藏高原迁移发生在史前时代，但究竟什么时候，还有待进一步研究。据考古发现，西藏也发现了新石器时代的人类遗址。那么，自北方迁至西藏的藏族先民到达西藏之前，西藏有没有人类？如果有的话，这些人的血缘在目前藏族中占多大比例？很可能，在西藏发现的新石器时代人

类遗址，也是从北方迁移过去的人类留下的。由于藏族一直住在高寒地区，其生活条件与中国北方相似，而且他们与外界特别是南方蒙古人种的各民族接触较少，因此他们至今与一些典型的北方蒙古人群的遗传距离仍十分近。

4. 羌族也起源于北方，但现都居住在四川北川、茂县一带，他们明显地仍属于北方蒙古人种。

5. 维吾尔族虽然也有部分高加索人种血缘，却比较少，大部分是蒙古人种的血缘。维吾尔族与哈萨克族相聚也不是偶然的，这是因为他们的高加索人种血缘比柯尔克孜族、乌孜别克族、塔塔尔族、塔吉克族等都少，而北方蒙古人种血缘都较多。在以 38 个基因座的基因频率进行聚类时，维吾尔族与哈萨克族仍首先相聚，再与塔吉克族、柯尔克孜族相聚，然后才与我国其他北方民族相聚，说明这些民族由于都具有一定的高加索人种血缘，所以遗传结构是相似的。

在用 HLA 基因频率进行聚类分析时，维吾尔族没有和中国的其他民族相聚，却最后勉强地与白种人人群相聚了。这一方面说明维吾尔族实际上同时含有蒙古人种与高加索人种血缘的人群，另一方面也说明，虽然维吾尔族最后与高加索人群相聚，但很勉强。其聚类可能与 HLA 这几个基因座的特殊性有关。如果用其他基因座或更多基因座的基因频率来分析，维吾尔族就不一定与高加索人群相聚。事实也正如此。在用微卫星 DNA 和 HLA 的基因频率进行聚类时维吾尔族也首先与蒙古人种人群相聚，而不是首先与高加索人群等其他人群相聚。在用 38 个基因座的基因频率进行聚类时，维吾尔族很自然地与中国北方民族先相聚，然后与南方民族相聚。

6. 壮族、傣族、黎族、瑶族、布依族、侗族、京族等都是典型的南方蒙古人种民族。

7. 彝族、景颇族、纳西族、阿昌族、白族、基诺族、拉祜族、哈尼族、保安族、傈族等，有不少是起源于北方的，但今天已融有许多南方蒙古人种的血缘，因此已基本上属南方蒙古人种。不过当所用分析的指标较少、或基因频率数据的准确性不是很高时，这些民族中有的还会显示出与北方蒙古人种人群相近。

8. 对仡佬族、佤族、水族、毛南族、畲族等人口较少的民族研究得较少，但可以完全肯定，他们都属南方蒙古人种。

9. 苗族和土家族遗传结构比较相近，他们既不属于北方蒙古人种，也不同于典型的南方蒙古人种，但基本上属南方蒙古人种。土家族是久居当地的土著，苗族则可能一度曾居住在长江以北地区。苗族人口众多，居住地区也很分散，有许多分支，值得进一步深入研究。

10. 德昂族、布朗族、佯族三个民族，其语言同属孟—高棉语系，在遗传结构上也相似。他们是云南最早的土著，在血缘上也与东南亚、南亚的蒙古人种最为接近。

11. 高山族实际上包括九个民族，即泰雅族、赛夏族、布农族、邹族、鲁凯族、排湾族、卑南族、阿美族和雅美族。原来还有一个平埔族是住在平地的，已完全融入汉族了。他们大部分来自菲律宾、马来西亚、文莱等地，小部分来自中国大陆南方。他们属南方蒙古人种，而且与东南亚人群在遗传结构上十分相近。

参 考 文 献

- 1 中国大百科全书·民族卷. 北京: 中国大百科全书出版社, 1986
- 2 国务院人口普查办公室, 国家统计局人口和社会科技统计司. 中国 2000 年人口普查资料. 北京: 中国统计出版社, 2002
- 3 郭大烈, 董建中. 中华民族知识通览. 昆明: 云南教育出版社, 2000
- 4 杜若甫. 中国人群体遗传学. 北京: 科学出版社, 2004
- 5 郭净, 段玉明, 杨福泉. 云南少数民族概览. 昆明: 云南人民出版社, 1999
- 6 赵锦元, 戴佩丽. 世界民族通览. 北京: 中央民族大学出版社, 2000
- 7 郝文明. 中国民族. 北京: 中央民族大学出版社, 2001
- 8 王鍾翰. 中国民族史概要. 太原: 山西教育出版社, 2004
- 9 杜若甫, 叶傅. 中国的民族. 北京: 科学出版社, 1994
- 10 王鍾翰. 中国民族史(增订本). 北京: 中国社会科学出版社, 1994
- 11 高文德. 中国少数民族史大辞典. 长春: 吉林教育出版社, 1995
- 12 刘海琦, 李自然. 少数民族史话. 北京: 中央民族学院出版社, 1996
- 13 杨圣敏. 中国民族志. 北京: 中央民族大学出版社, 2003
- 14 孙开来. 人类发育与遗传学. 北京: 科学出版社, 2004
- 15 Benjamin Lewin. Genes VIII. Pearson Prentice Hall, 2004
- 16 王琳芳, 杨克恭. 医学分子生物学原理. 北京: 高等教育出版社, 2003
- 17 夏家辉. 医学遗传学. 北京: 人民卫生出版社, 2004
- 18 陈竺. 医学遗传学. 北京: 人民卫生出版社, 2001
- 19 周柔丽. 医学细胞生物学. 北京: 北京大学医学出版社, 2006
- 20 王修海, 单长民, 杨康鹏. 医学遗传学实验指导. 北京: 科学出版社, 2008
- 21 张涛, 马爱民. 医学遗传学. 北京: 北京大学医学出版社, 2008
- 22 金力, 褚嘉佑. 中华民族遗传多样性研究. 上海: 上海科学技术出版社, 2006
- 23 戴玉景, 丁建生, 邹占彪. 甘肃裕固族体质特征初步研究. 人类学学报, 1987, 6(3): 227-235

中国少数民族遗传学概论

- 24 郁瑞生,戴玉景,薄岭. 青海撒拉族体质特征研究. 人类学学报, 1995, 14(1): 32-39
- 25 杨东亚,戴玉景. 甘肃保安族体质特征研究. 人类学学报, 1990, 9(1): 55-63
- 26 戴玉景. 青海土族体质人类学研究. 人类学学报, 1997, 16(4): 274-284
- 27 戴玉景,杨亚东. 甘肃东乡族体质特征研究. 人类学学报, 1991, 10(2): 127-134
- 28 艾琼华,赵建新,肖辉等. 新疆蒙古族体质人类学研究. 人类学学报, 1994, 13(1): 46-55
- 29 王静兰,邵兴周,崔静等. 新疆蒙古族土尔扈特部体质特征调查. 人类学学报, 1993, 12(2): 137-146
- 30 朱钦,刘文忠,李志军等. 蒙古族的体格、体型和半个多世纪以来的变化. 人类学学报, 1993, 12(4): 347-356
- 31 吕泉,袁华生,代素娥等. 内蒙古赤峰地区蒙古族成人体质特征的研究. 人类学学报, 1998, 17(1): 32-44
- 32 冯麟. 呼和浩特市蒙古族学生身体发育调查报告. 中华卫生杂志, 1965, 10(1): 41-44
- 33 韩向君,何欣,段秀吉等. 吉林省满族体质特征调查. 人类学学报, 1993, 12(1): 55-63
- 34 邵兴周,王笃伦,崔静等. 新疆察布查尔锡伯族体质特征调查. 人类学学报, 1984, 3(4): 349-362
- 35 施全德,胡俊清,赵贵新. 赫哲族体质特征. 人类学学报, 1987, 6(4): 336-342
- 36 陈良忠,杜若甫. 鄂温克族与鄂伦春族的群体遗传学研究. 人类学学报, 1983, 2(3): 282-292
- 37 朱钦,富杰,刘文忠等. 达斡尔族成人的体格、体型和半个多世纪以来的变化. 人类学学报, 1996, 15(2): 119-126
- 38 施全德,胡俊清,杨宏有. 黑龙江省达斡尔族体质特征调查. 人类学学报, 1983, 2(1): 60-71
- 39 胡兴宇,顾国雄,汪澜等. 对甘肃省玛曲县境内安多藏族青壮年体质特征的调查研究. 泸州医学院学报, 1991, 14(2): 102-108
- 40 泸州医学院,阿坝州卫生学校,红原(马尔康,茂汶,南坪)县卫生局等. 四川藏族、羌族青壮年体质特征调查. 解剖学通报, 1984, 7: 13
- 41 张振标. 藏族的体质特征. 人类学学报, 1985, 4(3): 250-258
- 42 夏元敏,吴融西. 藏族体质形态的初步研究,见: 中国人类学会编,中国八个民族体

- 质调查报告. 昆明: 云南人民出版社, 1982, 97-114
- 43 庞祖荫, 李培春, 梁明康等. 广西德峨苗族、彝族体质调查. 人类学学报, 1987, 6(4): 324-335
- 44 刘冠豪, 李明, 余发昌等. 保安族的体质特征研究. 人类学学报, 1990, 9(2): 122-129
- 45 刘冠豪, 余发昌, 李明等. 云南纳西族的体质特征研究. 人类学学报, 1992, 11(1): 13-19
- 46 李明, 李跃敏, 余发昌等. 云南普米族的体质特征研究. 人类学学报, 1995, 14(3): 227-232
- 47 任家武, 李严斌, 唐茂林等. 湖南苗族体质特征. 人类学学报, 1996, 15(3): 260-262
- 48 余发昌, 李明, 刘冠豪等. 云南苗族的体质特征研究. 人类学学报, 1994, 13(4): 321-326
- 49 罗远才, 韩承柱, 肖冠宇等. 湖南土家族的体质特征. 人类学学报, 1985, 4(2): 160-172
- 50 梁明康, 李培春, 吴荣敏等. 贵州仡佬族体质特征. 人类学学报, 1994, 13(1): 64-71
- 51 梁明康, 李培春, 庞祖荫等. 广西隆林县仡佬族体质特征. 右江民族医学院学报, 1989, 11(4): 1-9
- 52 顾牛范, 江三多, 王克勋等. 畲族隔离群体的人类遗传学调查. 遗传与疾病, 1985, 2(1): 14-17
- 53 曾宪智, 戴福珍, 史可舜等. 福建省福安市畲族成人人体质调查报告. 福建医药杂志, 1996, 18(5): 211-216
- 54 李培春, 梁明康, 吴荣敏等. 水族的体质特征研究. 人类学学报, 1994, 13(1): 56-63
- 55 陈翁良, 吴融西等. 白族体质形态的初步研究. 见: 中国人类学会编, 中国八个民族体质调查报告. 昆明: 云南人民出版社, 1982, 88-96
- 56 李明, 李跃敏, 余发昌等. 云南拉祜族的体质特征, 人类学学报, 2001, 20(1): 39-44
- 57 王桂伦, 戴星翼, 陈翁良等. 哈尼族体质形态的初步研究. 见: 中国人类学会编, 中国八个民族体质调查报告. 昆明: 云南人民出版社, 1982. 49-68
- 58 陈翁良, 郭常富, 王桂伦等. 基诺族体质形态的初步研究. 见: 中国人类学会编, 中

中国少数民族遗传学概论

- 国八个民族体质调查报告. 昆明: 云南人民出版社, 1982. 11-29
- 59 李明, 李跃敏, 程宏忠等. 云南阿昌族的体质特征. 人类学学报, 1992, 11(1): 20-26
- 60 Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans; a Y chromosome perspective. *Science*. 2000. 290(5494): 1155-1159.
- 61 Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill p, Cavalli-Sforza L, Chakraborty R, Jin L. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet*. 2000. 107(6): 582-590.
- 62 Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Deka R, Su B, Chakraborty R, Jin L. Genetic evidence supports demic diffusion of Han culture. *Nature*. 2004. 431(7006): 302-305.
- 63 Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet*. 2002. 71(3): 466-482.
- 64 Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HL, Villems R. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol*. 2002. 19(10): 1737-1751.
- 65 Seielstad MT, Minch E, Cavalli-Sforza LL. Genetic evidence for a higher female migration rate in humans. *Nat Genet*. 1998. 20(3): 278-280.
- 66 Wolpoff MH, Hawks J, Caspari R. Multiregional, not multiple origins. *Am J Phys Anthropol*. 2000. 112(1): 129-136.
- 67 Zhao Z, Jin L, Fu YX, Ramsay M, Jenkins T, Leskinen E, Pamiilo P, Trexler M, Patthy L, Jorde LB, Ramos-Onsins S, Yu N, Li WH. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. *Proc Natl Acad Sci USA*, 2000, 97(21): 11354-11358.
- 68 王广结, 王钢, 尚锦青等. 对内蒙古地区少数民族红细胞血型的研究. 内蒙古医学杂志, 1993. 13(4): 1-5
- 69 王广结, 王钢, 尚锦青等. 内蒙古地区汉、蒙、回、满四个民族 12 个红细胞血型系统的研究. 内蒙古中心血站输血专辑, 1992, (1): 7-17

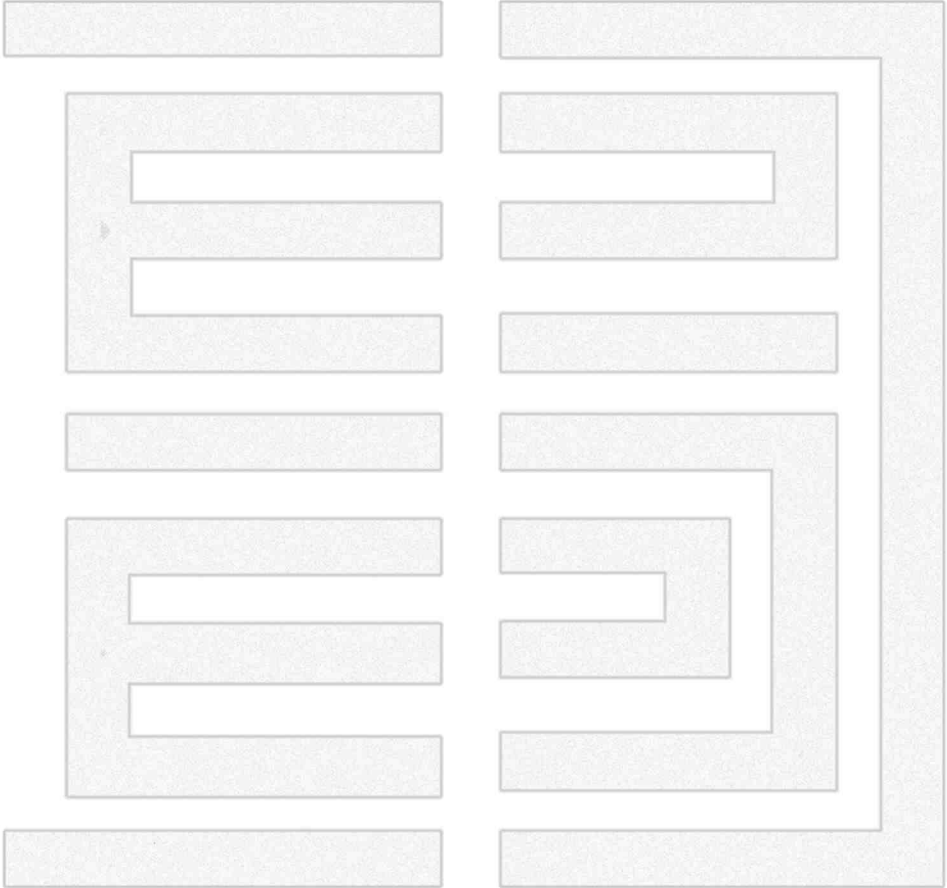
参 考 文 献

- 70 尚书颂, 郭寅同. 中国各省区及民族 ABC 血型的分布. 天津医药杂志(输血及血液学附刊), 1963. (1): 57-61
- 71 陈稚勇, 赵桐茂, 张工梁. 中国人 ABO 血型分布. 遗传, 1982, 4(2): 4-7
- 72 李振翩. 中国民族的血属. 中华医学杂志, 1930, 16(1): 5
- 73 穆继鸿. 牡丹江地区 ABO 血型分布. 中华医学检验杂志, 1987, 10(5): 302
- 74 汪家志, 孙志芳. 伊春地区 5143 例 ABO 血型分布. 中华医学检验杂志, 1989, 12(1): 27
- 75 陈稚勇, 郭寅同. ABO 血型及其亚型的调查. 中华内科杂志, 1962, 10(7): 40S-409
- 76 陈仁彪, 赵桐茂, 叶根耀等. 中国大陆 HLA 多态性联合报告. 上海免疫学杂志, 1987, 7(6): 321-324
- 77 陈仁彪, 叶根耀, 庚镇诚等. 我国大陆主要少数民族 HLA 多态性聚类分析和频率分布对中华民族起源的启示. 遗传学报, 1993, 20(5): 389-398
- 78 吴国光, 邓志辉, 高素青等. 6965 名汉族骨髓供者 HLA 多态性分析. 中华血液学杂志, 2004, 25(8): 473-477
- 79 Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Rood AI, Zegura SL, Hammer MF. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet.* 2001. 69(3): 615-628.
- 80 Lell IT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet.* 2002. 70(1): 192-206.
- 81 Fay JC, Wu CI A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol Biol Evol.* 1999, 16(7): 1003-1005.
- 82 Li T, Etlar DA New Middle Pleistocene hominid crania from Yunxian in China. *Nature.* 1992, 357(6377): 404-407.
- 83 Brooks AS, Wood R Palaeoanthropology. The Chinese side of the story. *Nature.* 1990, 344(6264): 288-289.
- 84 Zhang ZR An analysis of the physical characteristics of modern Chinese. *Acta Anthropol Sin.* 1988, 7: 314-323.
- 85 Zhang H Distribution of dermatoglyphic parameters in fifty-two Chinese populations. *Acta Anthropol Sin.* 1988, 7: 39-45.

中国少数民族遗传学概论

- 86 Ding YC, Wooding S, Harpending HC, Chi HC, Li HP, Fu YX, Pang JF, Yao YG, Yu JG, Moyzis R, Zhang Y. Population structure and history in East Asia. *Proc Natl Acad Sci USA*, 2000, 97(25): 14003-14006.
- 87 Kayser M, Brauer S, Weiss G, Schiefenhover W, Underhill PA, Stoneking M Independent histories of human Y chromosomes from Melanesia and Australia. *Am J Hum Genet*, 2001, 68(1): 173-190.
- 88 Cordaux R, Weiss G, Saha N, Stoneking M The northeast Indian passageway: a barrier or corridor for human migrations? *Mol Biol Evol*, 2004, 21(8): 1525-1533.
- 89 Qian Y, Qian B, Su B, Yu J, Ke Y, Chu Z, Shi L, Lu D, Chu J, Jin L. Multiple origins of Tibetan Y chromosomes. *Hum Genet*, 2000, 106(4): 453-454.
- 90 Jin HJ, Kwak KD, Hammer MF, Nakahori Y, Shinka T, Lee JW, Jin F, Jia X, Tyler-Smith C, Kim W. Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. *Hum Genet*, 2003, 114(1): 27-35.
- 91 Wells RS, Yuldashewa N, Ruzibakiev R, Underhill PA, Evsceva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA*. 2001, 98(18): 10244-10249.
- 92 杜若甫, 肖春杰. 从遗传学探讨中华民族的源与流. *中国社会科学*, 1997, (4): 139-149.
- 93 陈德珍. 中国新石器时代居民体质类型及其继承关系. *人类学学报*, 1986, 5(2): 114-125.
- 94 刘武. 华北新石器时代人类牙齿形态特征及其在现代中国人起源与演化. *人类学学报*, 1995, 14(4): 360-378.
- 95 刘武. 蒙古人种及现代中国人的起源与演化上的意义. *人类学学报*, 1997, 16(1): 55-73.
- 96 翁自力, 袁义达, 杜若甫. 中国人群遗传结构分析. *人类学学报*, 1989, 8(3): 261-268.
- 97 谭茜, 杜若甫. 中国 21 个人群的遗传拓扑学分析. *人类学学报*, 1993, 12(1): 80-87.
- 98 肖春杰, 杜若甫. 中国人群基因频率的主成分分析. *中国科学*, 2000, 30(4): 434-442.

- 99 腾百军, 陈峰, 傅松滨等. 我国东北地区 3 个群体 DYS390 多态位点的遗传学研究. 人类学学报, 2000, 19(4): 318-323
- 100 杜若甫. 用人类群体遗传学数据分析中国各民族的起源. 中国遗传学研究(1987 - 1990), 中国遗传学会编, 北京: 中国科学技术出版社, 1991

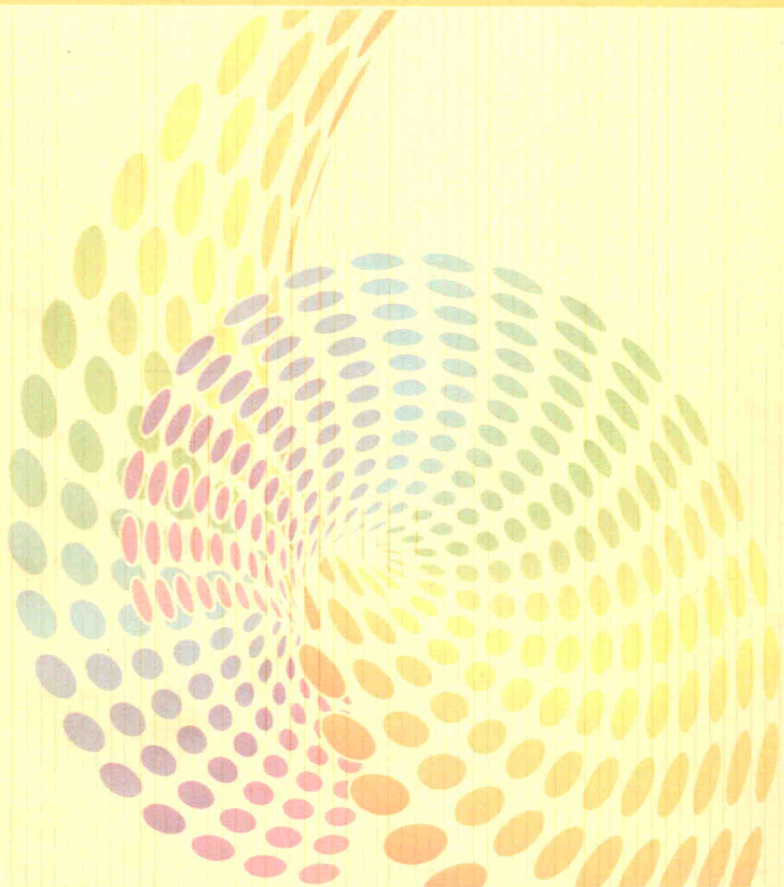




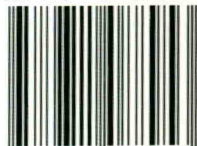
责任编辑 / 常远岐

封面设计 / 布拉格

Genetics For the Ethnic Groups of China



ISBN 978-7-81108-895-3



9 787811 088953 >

定价：53.00元