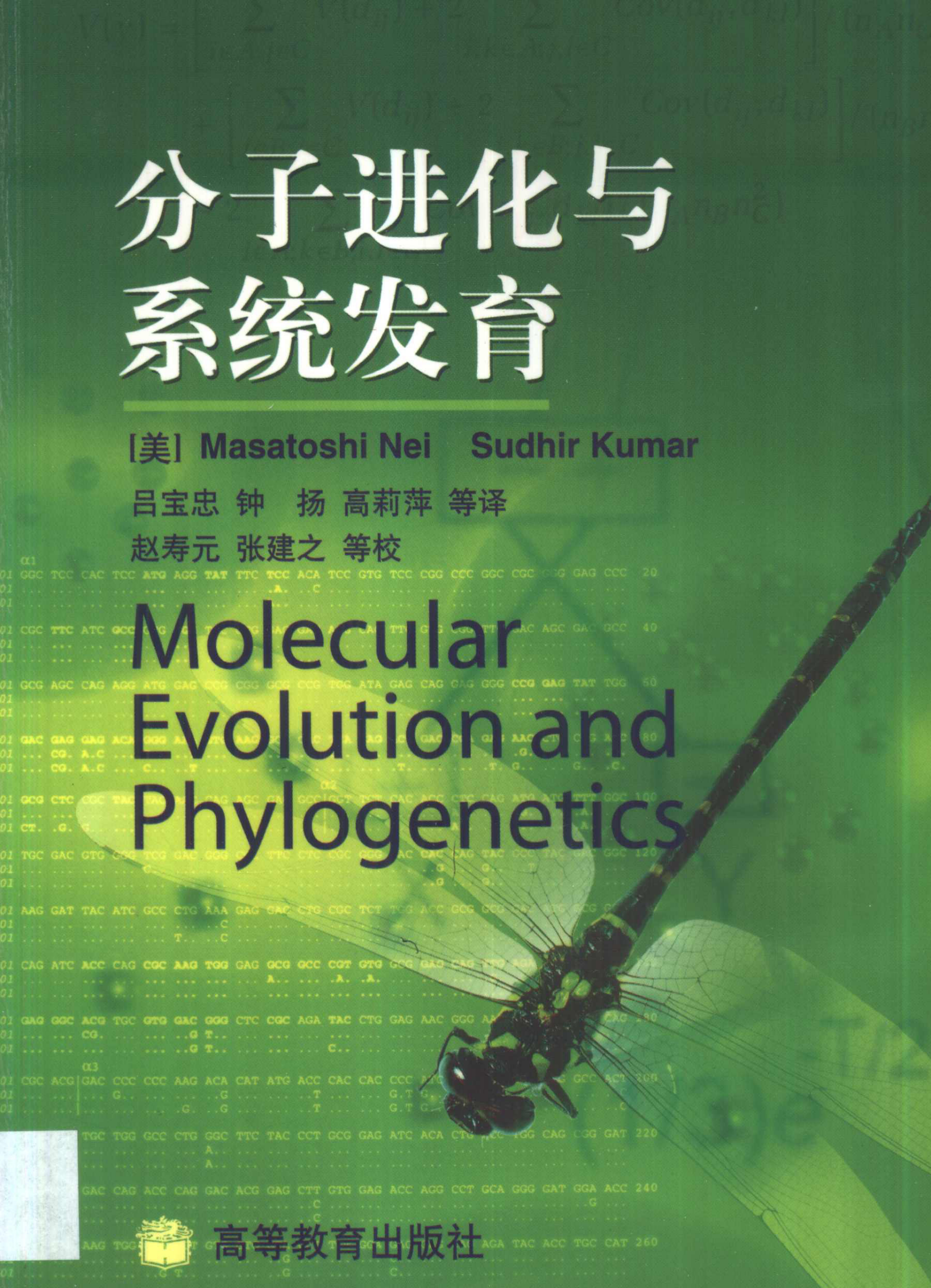
分子进化与系统发育

[美] Masatoshi Nei　Sudhir Kumar

吕宝忠 钟　扬 高莉萍 等译

赵寿元 张建之 等校

Molecular Evolution and Phylogenetics



高等教育出版社

图书在版编目(CIP)数据

分子进化与系统发育/(美)根井正利,(美)库马著;吕宝忠等译.
-北京:高等教育出版社,2002.6
ISBN 7-04-011204-3

Ⅰ.分... Ⅱ.①根...②库...③吕... Ⅲ.①分子进化
②系统发育 Ⅳ.①Q75②Q111.4

中国版本图书馆 CIP 数据核字(2002)第 037016 号

图字:01-2002-1561 号

责任编辑 王 莉 **封面设计** 王凌波
版式设计 李 杰 **责任印制** 陈伟光

出版发行 高等教育出版社
社 址 北京市东城区沙滩后街 55 号 **邮政编码** 100009
购书热线 010-64054588 **传 真** 010-64014048
免费咨询 800-810-0598 **网 址** http://www.hep.edu.cn
http://www.hep.com.cn

经 销 新华书店北京发行所
印 刷 北京民族印刷厂

开 本 787×960 1/16 **版 次** 2002 年 6 月第 1 版
印 张 19.50 **印 次** 2002 年 6 月第 1 次印刷
字 数 400 000 **定 价** 32.00 元

中 文 版 序

20 世纪以来，分子生物学在探索生命的奥秘中发挥了关键作用，它不仅揭示了基因的结构和功能，而且阐明了生命过程的基本机制——如发育、生理和繁殖。在包括从细菌到人类的所有生物体中，尽管很多过程的细节还有待明了，这些基本机制在本质上都是相同的。然而，生物学中还有许多重大奥秘有待探索，如生命的起源、脑功能以及由原始生物向高等生物的进化。这些奥秘都与进化研究有关，因而大家寄希望于进化生物学的发展，尤其是分子进化生物学对揭开这些奥秘可能具有重要的意义。

本书的目的并不在于全面论述生物学上的重大问题，而是从分子进化生物学的方法和技术着手探讨上述问题。换言之，本书系统地介绍了用于 DNA 和蛋白质进化研究的统计方法。当今，DNA 和蛋白质数据的统计分析都离不开计算机的应用。因此，本书的读者应该具备操作计算机的技能。本书初版于 2000 年 6 月，问世后又出现了一些新的相关计算机程序，而本书涉及的大量计算均可用 MEGA 2(http://www.megasoftware.net)和 PAUP* (Swafford 1998)来完成。其他程序也可从 http://mep.bio.psu.edu 获取。本书英文版的 4 次印刷均存在一些错误(见 http://lifesciences.asu.edu/mep/text)。中文版则完全更正了这些错误。

我十分高兴本书中文版的面世，尤其是基于两方面的原因：一是进化生物学在中国已快速发展，因此，中国的进化生物学研究迫切需要此类专著；另一个原因则是十余年来，许多中国学者参与了本人实验室的原创性研究，而且他们中的一些人目前已在分子进化研究领域中处于领先地位。希望本书能激励更多的中国年轻学者成为该领域未来的领导者。

最后，我衷心感谢吕宝忠教授和钟扬教授及复旦大学的学生们将本书译成中文。我还要感谢赵寿元教授和张建之博士，他们校阅了中文版的手稿并提出了宝贵意见。

根井正利(Masatoshi Nei)

2002 年 3 月 5 日于美国宾夕法尼亚州

前　言

统计学是一门用途极为广泛的学科，但有效的应用者却寥寥无几。对大多数人而言，传统的通往统计学知识之路被数学这堵令人望而生畏的高墙所阻挡。我们这里走的路就是避开这堵墙。

Efron 和 Tibshirani(1993)

本书第一作者(根井正利)于 1987 年出版了《分子进化遗传学》(Molecular Evolutionary Genetics)一书。目的是要把当时研究分子水平上进化的两门不同的学科统一起来，一门是重建生物进化历史，另一门则是研究进化的机制。最近十年来，这两个研究领域均已取得显著进展，并且不可分隔。产生这样进展的部分原因是 PCR 等生化技术的发展使得快速测定 DNA 序列成为可能。这一技术上的创新使得许多不同生物类群的大量 DNA 序列数据得以测定，大大加速了研究分子进化遗传学的研究速度。

取得上述进展的另一个重要原因是数据分析的统计方法以及计算机技术的发展。近年来，在系统发育分析和进化机制研究等方面建立和发展了很多新的方法，使分子进化研究不仅更为准确，而且更为简便。同时，高速个人计算机的不断升级，使许多研究者能对大规模数据进行细致的统计分析。

《分子进化遗传学》一书系根据数学理论和实验数据，为新形成的交叉学科提出一个框架。本书的目的则不同。本书旨在为分子进化研究提供有用的统计方法，并以实际数据为例说明如何运用这些方法。我们将介绍分析新近涌现的进化问题的各种统计方法，但不准备讨论当今分子进化遗传学上的一些生物学发现。Avise (1944)、Hartl 和 Clark (1997)、Li (1997)、Powell (1997) 以及 Graur 和 Li (1997)等已对这些新发现进行了综述，读者可以参阅他们的著作。当然，本书提供了许多实例并讨论其中有趣的生物学问题。本书还介绍了分子进化的基本知识，特别是第一章和第十章，以便读者理解本书所涉及的生物学问题。附录则提供了对分子进化研究有用的地质年代划分的资料等。

本书可供分子进化领域中的研究生和研究人员阅读。我们设想读者已具备了分子生物学、进化和初等统计学的基础知识。虽然本书涉及用于分子进化遗传学的统计方法，但我们不打算讨论这些统计方法的数学基础，而是想说明如何运用这些方法。当今几乎所有的数据分析都用计算机来处理，因此，我们是以计算机操作方式来解释统计方法的。

此前，我们出版了一个名为 MEGA 的计算机软件包(Kumar 等 1993)。这个

程序已经过时了，所以我们作了修订(Kumar 等 2000)。修订版(MEGA2.0 版本，MEGA2)已用来计算本书的许多数学实例。因此，我们建议读者用 MEGA2 来验算数据的分析结果，并可学习所介绍的统计方法的细节。在本书的网站 http://www.oup-usa.org/sc/0195135857 上可获得 MEGA2。本书另外一些数据分析是用 PAUP* (Swofford 1998)、PHYLIP(Felsenstein 1995)、MOLPHY(Adachi 和 Hasegawa 1996b)、PAML(Yang 1995b，1999)和其他程序运行的，这些程序可从 http://mep.bio.psu.edu 网站上获取。我们发现，用简约法和最大似然法分析 DNA 序列的系统发育关系时，PAUP* 尤其有用。在本书网页上可查到用于本书的全部原始数据和计算所需的其他资料等。

本书首先讨论了分析蛋白质和 DNA 序列数据的统计方法，然后介绍了分析等位基因频率的方法，包括我们自己实验室和其他实验室创建的多种新方法。有些方法及其统计特性是在本书撰写过程中研究并且是刚刚发表的，或者是在本书中首次披露的。收录的标准为实用性而非数学上的创新。以合乎生物学现实的假设为依据的统计方法也优先收录。由于第一作者的实验室从事该领域的研究已有 30 年，而且有些专题主要是在他的实验室进行研究的，所以本书主要收录了这一实验室的许多研究工作。本书中有关序列数据的方法多于等位基因频率数据的方法，这只是因为如今序列数据已远远多于基因频率数据。许多涉及后者的方法已在以前出版的《分子进化遗传学》作了讨论。读者如希望详细了解分析经典的等位基因频率数据的统计方法，请参考该书。

我们用了许多例子来解释数据分析。这些实例用来说明的不仅仅是计算过程而且还包括怎样从数据分析中获取生物学信息。请注意，本书的分析结果与所引用的原始论文的分析结果并不总是完全相同，这是因为我们已按照本书的目的对所有数据重新做了分析。即使引用的是我们自己以前论文中的例证也是如此。为了避免出现严重差错，我们尽量选用我们自己比较熟悉的例子。这么做对那些发表了有意义的生物学发现的作者来说恐怕有失公允。如果本书用作教材，教师不妨选用自己的例子，尤其是自己实验室所研究的。教师也可采用本书未涉及的有关理论的补充教材。

本书是两位作者近四年来的合作成果。第一作者主要负责选题并撰写正文，第二作者负责所需要的例子和数据分析。第二作者还主要负责开发计算机软件 MEGA2 并处理本书中的许多实例。

如果没有第一作者实验室的过去和现在的合作者的帮助，本书是不可能完成的。十多年来，他们齐心协力解决了在分子进化统计研究中遇到的许多挑战性难题。我们特别感谢 Andrey Rzhetsky、Tatsuya Ota、Naoko Takozaki、Koichiro Tamura、Ziheng Yang、Cludia Russo、Tanya Sitnikova、Jianzhi(George)Zhang 和 Xun

Gu。此外,访问学者如 Naoyuki Takahata、Willem Ferguson、Famio Tajima、Yashio Tateno 和 Joaquin Dopazo 也做了许多贡献。我们还要感谢宾州州立大学分子进化遗传学研究所的成员,他们提出了许多问题,促使我们进行研究。衷心感谢阅读本书各个章节初稿并提出有宝贵意义的朋友们和同事们。他们是 Tom Dowling、Alan Filipski、Rodney Heneycut、Junhyong Kim、Adrey Rzhetsky、Naruya Saitou、James Lyons－Weiler、Mike Miyamoto、Alex Reoney、Ziheng Yang、George Zhang 和 Marey Uyemoyaman。还要特别感谢 Sudhindra Gadagkar、Ingrid Jakobsen 和 Thomas Whittam,他们阅读了几乎全部手稿并提出有价值的修改意见。我们特别感谢 Joyce White 耐心地打印了好几次修改稿,并帮助我们整理参考文献。对于 Barb Backes 绘制定稿的各幅图例,谨致谢忱。

本书得到了美国国立卫生研究院(NIH)和国家科学基金会(NSF)给第一作者的研究资助。本书有一部分是第一作者在日本国立遗传研究所(NIG)作学术休假时撰写的,得到了日本科学促进会(JSPS)的资助(邀请人:Takashi Gojobori)。对上述慷慨资助的机构谨致谢意。

根井正利(Masatoshi Nei)

苏德海尔·库马(Sudhir Kumar)

目　录

第一章

进化的分子基础

1.1 生命的进化树

自达尔文时代起，许多生物学家都有一个梦想，那便是重建地球上所有生命的进化历史并以系统树的形式描述这部历史(Haeckel 1866)。理想的途径应该是利用化石证据，但是化石是如此的零散且不完整，致使大多数研究者转向比较形态学和比较生理学的方法。通过后两条途径，经典进化学家已得出有机体进化历史的主要框架。然而，形态和生理性状的进化如此复杂，以致不可能产生一幅进化历史的清晰图像。不同学者重建的系统树在细节上几乎总是可争议的。

分子生物学的进展大大地改变了这种局面。由于所有生物的蓝图都用 DNA(在某些病毒中则用 RNA)来书写，因而人们可以通过比较 DNA 来研究它们的进化关系。分子途径较经典的形态学和生理学途径有如下优点。首先，DNA 仅由 4 种碱基组成，即：腺嘌呤(A)、胸腺嘧啶(T)、胞嘧啶(C)和鸟嘌呤(G)。所有生物，不论是细菌、植物和动物中的 DNA 均由这 4 种碱基组成。因而，可用它们比较所有有机体的进化关系。这在经典进化研究方法中是不

可能做到的。

其次，DNA 的进化演变或多或少是有规律的，因而能用数学模型来描述其变化并可比较亲缘关系较远的生物间的 DNA。形态性状的进化演变，即使在一段较短的进化时间，也是极其复杂的。因而，形态的系统发育研究必然会有各种各样的假设，但这些假设往往难以令人信服。第三，所有生物的基因组都是由长长的核酸序列组成，比形态性状包含的系统发育信息要多得多。鉴于上述原因，分子系统学有望澄清生命系统树中多处对于经典途径来说极为棘手的问题。

系统学或分类学是生命科学中争议最多的领域之一。种、属、科以及更高的分类单元的定义常常带有主观性。对同一类群（如果蝇）进行研究的两位专家，在将这一类群归属于亚种还是种或属等分类单元时，判断会很不一致。较之分类学，系统发育学内的矛盾要少一些，因为它首要考虑的是有机体间的进化关系，而将某一类群归属到一个确定的分类单元等级，则是次要的工作。然而，系统发育学与分类学的关系相当紧密，因为有机体的分类应反映它们的进化历史（Darwin 1859；Mayr 1968）。由此可见，系统发育学对发展系统学的科学基础具有重要作用，尽管它还不可能解决分类学的所有难题。分子系统学的新近进展，已经为生物分类问题的许多方面提供了崭新见解，这将在以后的章节中叙述。

1.2 进化机制

进化的第一原因是基因突变。由核苷酸替代、插入/缺失、重组和基因转换等引发的突变基因或 DNA 序列，通过群体水平的遗传漂变和/或自然选择进行扩散（参见 Nei 1987；Hartl 和 Clark 1997），并最终在物种中得以固定。倘若此突变基因产生新的形态或功能性状，除非基因再次突变，此性状将会传递给其所有后裔。因此，当对某一类群构建了一棵有效的系统树，我们就可以找到具有此突变性状的谱系。

上述信息可用以研究特定性状的进化机制。将具有该特定性状的谱系所处的环境条件与无此性状的谱系所处的环境条件进行比较，就可能会搞清该性状是由自然选择还是随机演化所决定的。如果我们能鉴别出所涉及的基因并研究其进化演变，将会明了何种类型的突变产生了特定的形态或生理性状。

这类研究已经进行，如与反刍动物和叶猴的双肠道消化系统进化有关的溶菌酶和核糖核酸酶的分析（Stewart 等 1987；Jermann 等 1995）。这些动物的前肠道寄生着能发酵草料和树叶的细菌。这些细菌在后肠道被溶菌酶消化，释放出的 RNA 被核糖核酸酶降解。发酵混合物，包含被消化的细菌，为宿主提供养料（Bernard 1969）。通过统计方法，现在已有可能推导出祖先物种蛋白质的氨基酸序列（Fitch 1971；Maddison 和 Maddison 1992；Yang 等 1995b），然后通过定点诱变（site-

directed mutagenesis)重建祖先蛋白质。据此,可以研究古蛋白质的催化活性(Jermann 等 1995)。这样,就有可能研究基因功能的进化演变。

研究突变、自然选择、遗传漂变和重组等的相对重要性是群体遗传学的一个重要课题。为此,群体遗传学家现在正在对一个基因座上的不同等位基因进行测序,来了解它们的进化历史。在这里,问题不是物种的系统树,而是同一物种内不同等位基因的系统树。这一类研究获得的一个有趣的结果是,哺乳类主要组织相容性复合体(major histocompatibility complex,MHC)基因座上某些等位基因谱系在群体中已持续了好几百万年(Figueroa 等 1988;Lowlor 等 1988;McConnell 等 1988;Hughes 和 Yeager 1998)。该结果与 MHC 分子的抗原识别部位受超显性选择的观点相吻合(Hughes 和 Nei 1988)。多态等位基因的系统发育分析也指出,在基因内重组出现的频率相当高(Robertson 等 1995;Fitch 1997)。

多态等位基因的系统发育分析也可以为研究两群体间的基因交流的程度提供重要信息。30 年前,Prakash 等(1969)对北美洲和南美洲(哥伦比亚的波哥大)的果蝇(*Drosophila pseudoobscura*)群体的某些酶进行电泳分析,发现许多等位基因为这两个群体所共有。根据这一观察,他们认为,波哥达群体是最近才形成的,可能在 1950 年左右由北美迁入。然而,Coyne 和 Felton(1977)在对上述实验数据进行了仔细分析后,对他们的结论提出质疑。其后,Schaeffer 和 Miller(1991)根据南、北美洲两群体的乙醇脱氢酶多态等位基因的 DNA 序列数据构建了系统树,发现波哥大群体早在 100 000 年前就可能已形成了。

上述诸例清楚表明,分子系统学已经成为研究进化机制的一个重要工具。

1.3 基因的结构与功能

尽管基因的分子生物学不是本书主题,但要读懂本书,必须对基因的基本结构和基本功能有所了解。从功能来看,基因分为两个大类:蛋白质编码基因(protein-coding gene)和 RNA 编码基因(RNA-coding gene)。蛋白质编码基因首先将遗传信息转录至信使 RNA(mRNA),再由 mRNA 将遗传信息翻译成蛋白质的氨基酸序列,RNA 编码基因则产生转移 RNA(tRNA)、核糖体 RNA(rRNA)、核内小 RNA(snRNA)等等。这些非信使 RNA 是 RNA 编码基因的最终产物。核糖体 RNA 是作为蛋白质合成机器的核糖体的核心组分,而 tRNA 在将 mRNA 的遗传信息转移至蛋白质的氨基酸序列中起到至关重要的作用。snRNA 仅仅存在于核内,其中某些 snRNA 与内含子剪切或其他 RNA 加工反应有关。

真核蛋白质编码基因的基本结构见图 1.1 所示。基因是一条由 A、T、G 和 C 线性排列成的长链,由 DNA 的转录区以及 5'和 3'非转录侧翼区 3 个部分组成。

侧翼区对控制转录和前信使 RNA(pre-mRNA)的加工是不可缺少的。一个前信使 RNA 由编码区和非编码区组成。编码区含有该基因所编码的氨基酸的信息,而非编码区则含有某些调节多肽链产生的必要信息。非编码区某些片断在成熟信使 RNA 产生过程中被剪切。剪切片断称为内含子(intron),而保留部分称为外显子(exon)(图 1.1)。基因的外显子数目在不同基因中是不同的。原核基因没有内含子,而某些真核基因(如肌营养不良基因)的内含子多达 78 个(Roberts 等 1992)。内含子的功能尚未完全了解。通常,一个内含子以双核苷酸 GT 起始,并以 AG 终止。这两个双核苷酸提供了正确的内含子剪切信号。

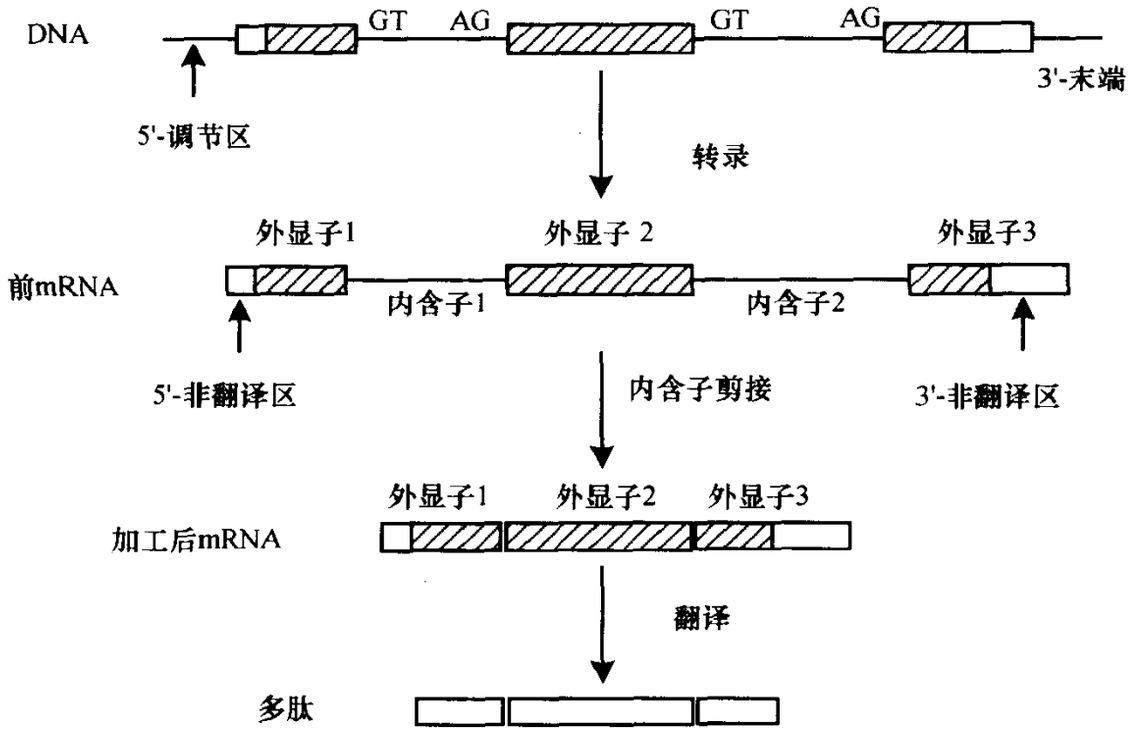


图 1.1　与转录和翻译有关的真核蛋白编码基因的基本结构

一个基因的核苷酸所携带的遗传信息首先被转移到 mRNA,通过两者核苷酸间简单的 1 对 1 的转录过程实现。转移至 mRNA 的遗传信息决定了蛋白质的氨基酸序列。mRNA 上的核苷酸,以 3 个为一体(即三联体),连续进行翻译。根据遗传密码,每一个三联体或密码子(codon)被翻译为生成的多肽链上的一个特定氨基酸。

除少数例外,原核类和真核类中核基因的遗传密码看来是通用的。这个遗传密码(通用或标准遗传密码)对叶绿体基因同样适用,但线粒体基因使用了略有不同的遗传密码。表 1.1 给出了标准遗传密码。

在此表中,氨基酸以三字母符号表示(表 1.2)。对 4 个不同核苷酸 U(相当于 DNA 中的 T)、C、A 和 G 而言,共有 $4^3=64$ 种可能的密码子。

表 1.1 标准或"通用"遗传密码

密码子		密码子		密码子		密码子	
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Ter	UGA	Ter
UUG	Leu	UCG	Ser	UAG	Ter	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

表 1.2 单字母和三字母氨基酸密码

名称	密码		pH 7 时侧链的特征
	单字母	三字母	
丙氨酸	A	Ala	非极性(疏水性)
半胱氨酸	C	Cys	极性
天冬氨酸	D	Asp	极性(亲水性,酸性)
谷氨酸	E	Glu	极性(亲水性,酸性)
苯丙氨酸	F	Phe	非极性(疏水性)
甘氨酸	G	Gly	非极性
组氨酸	H	His	极性(亲水性,碱性)
异亮氨酸	I	Ile	非极性(疏水性)
赖氨酸	K	Lys	极性(亲水性,碱性)
亮氨酸	L	Leu	非极性(疏水性)
甲硫氨酸	M	Met	非极性(疏水性)
天冬酰胺	N	Asn	极性(亲水性,中性)
脯氨酸	P	Pro	非极性
谷氨酰胺	Q	Gln	极性(亲水性,中性)
精氨酸	R	Arg	极性(亲水性,碱性)
丝氨酸	S	Ser	极性
苏氨酸	T	Thr	极性
颉氨酸	V	Val	非极性(亲水性)
色氨酸	W	Trp	非极性
酪氨酸	Y	Tyr	极性

3 个密码子 UAA、UAG 和 UGA 为终止密码子(termination 或 stop codon),它们不编码任何氨基酸,其余 61 个密码子(有义密码子,sense codon)能编码氨基酸。然而,用来构成蛋白质的氨基酸只有 20 种(表 1.2),因此,必然有几个不同密码子编码同一种氨基酸。编码同一种氨基酸的不同密码子称为同义密码子(synonymous codon)。在遗传密码表中,AUG 编码甲硫氨酸,而它也是起始密码子(initiation codon)。被起始密码子编码的甲硫氨酸处于修饰状态,当形成多肽链后即被剪切。近来的研究已表明,在某些核基因中,CUG 和 UUG 也可作为起始密码子(Elzanowski 和 Ostell 1996)。研究 DNA 序列进化时,起思密码子必须排除,因为多数情况下它们处于不变化状态。同样,终止密码子也应排除。

表 1.3 示出脊椎动物线粒体基因的遗传密码,它与标准遗传密码略有差别。线粒体中,密码子 UGA 并非终止密码子,它编码色氨酸。AGA 和 AGC 不编码精氨酸,而变为终止密码子。在核基因中编码异亮氨酸的 AUA,却在线粒体上用来编码甲硫氨酸。脊椎动物线粒体的遗传密码不一定适用于其他生物。事实上,海鞘、棘皮动物、果蝇、酵母、植物和原生动物均有略为不同的遗传密码(详见表 1.4)。纤毛原生动物如四膜虫和草履虫的核基因的遗传密码也与标准遗传密码稍有差别。这里 UAA 和 UAG 在核基因也非终止密码子,而成了谷氨酰胺的密码子。还有,在原核类的一种支原体(*Mycoplasma capricolum*)上,通用密码中终止密码子 UGA 用来编码色氨酸(Osawa 1995)。

表 1.3 脊椎动物线粒体遗传密码(与标准遗传密码不同的用黑体表示)

密码子		密码子		密码子		密码子	
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Ter	UGA	**Trp**
UUG	Leu	UCG	Ser	UAG	Ter	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	**Met**	ACA	Thr	AAA	Lys	AGA	**Ter**
AUG	Met	ACG	Thr	AAG	Lys	AGG	**Ter**
GUU	Val	GCU	Ama	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

表 1.4 和标准遗传密码不一致的其他遗传密码

细胞器/有机体	密码子						
	UGA	AUA	AAA	AGR	CUN	CGG	UAR
标准遗传密码	Ter	Ile	Lys	Arg	Leu	Arg	Ter
线粒体密码							
脊椎动物	Trp	Met	·	Ter	·	·	·
海鞘	Trp	Met	·	Gly	·	·	·
棘皮动物	Trp	·	Asn	Ser	·	·	·
果蝇	Trp	Met	·	Ser	·	·	·
酵母	Trp	Met	·	·	Thr	·	·
原生动物	Trp	·	·	·	·	·	·
霉菌	Trp	·	·	·	·	·	·
腔肠动物	Trp	·	·	·	·	·	·
核密码							
四膜虫	·	·	·	·	·	·	Gln
支原体	Trp	·	·	·	·	·	·
euplotid	Cys	·	·	·	·	·	·

注:·表示与标准密码一致。R 代表 A 或 G,N 代表 T、C、A 或 G。

植物线粒体基因的密码子 CGG 并不直接翻译成色氨酸,而是该密码子中的 C,在 mRNA 形成后,转化为 U,此转变成的 UGG 按标准遗传密码编码色氨酸。该过程称为 RNA 编辑(RNA edition)(Covello 和 Gray 1993)。然而,在不同植物的氨基酸序列比较中,人们可将 CGG 作为色氨酸密码子。事实上,RNA 编辑也在其他真核类的某些线粒体基因中出现。我们在进行这些基因的蛋白质翻译中应当谨慎。

1.4 DNA 序列的突变

既然所有形态与生理性状最终都是 DNA 携带的遗传信息所控制的,那么这些性状的突变就是 DNA 分子中的某些变化的结果。DNA 变化有 4 种基本类型:一个核苷酸被另一不同的核苷酸替代(substitution,图 1.2A)、核苷酸缺失(deletion)(图 1.2B)、核苷酸插入(insertion)(图 1.2C)和核苷酸倒位(invertion)(图 1.2D)。插入、缺失和倒位的出现是以一个碱基或多个碱基为一个单元。如果插入或缺失出现在蛋白质编码区,它们有可能改变核苷酸序列的阅读框。这些插入和缺失称为移码突变(frameshift mutation)。

核苷酸替代分为两类:转换(transition)和颠换(transversion)。转换指的是一个嘌呤(purine,腺嘌呤或鸟嘌呤)被另一个不同的嘌呤所替代,或一个嘧啶

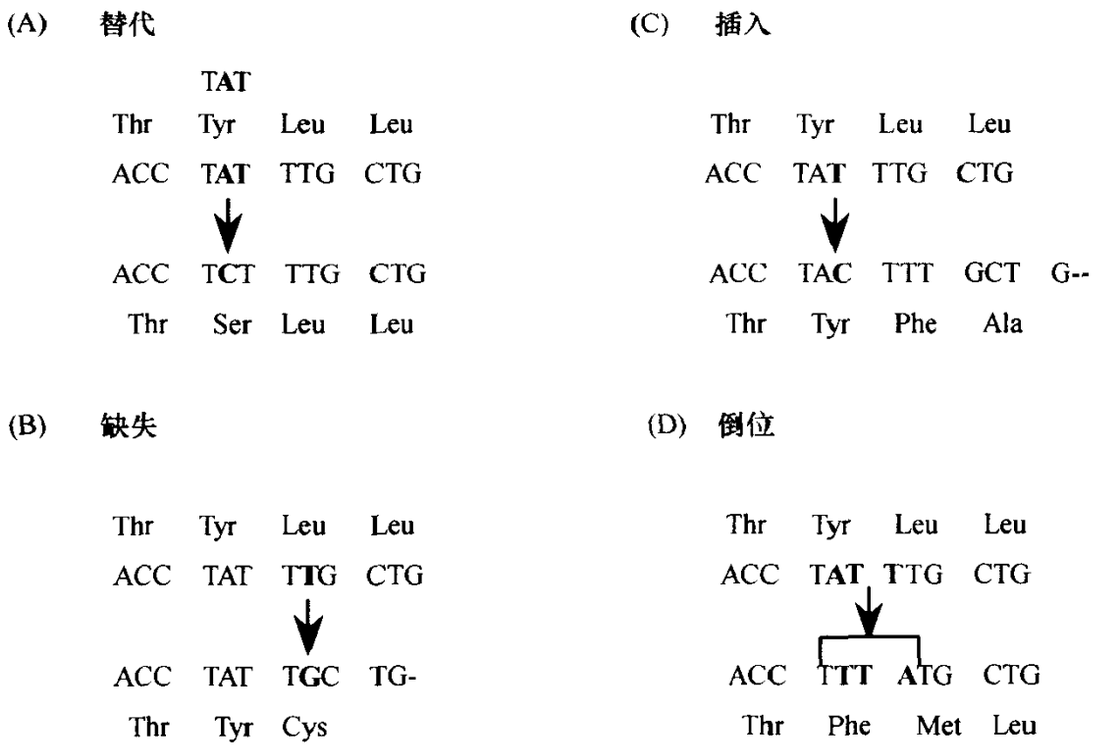


图 1.2 核苷酸水平突变的 4 种基本类型(由突变导致的核苷酸变化以黑体显示)

(pyrimidine,胸腺嘧啶或胞嘧啶)被另一不同嘧啶所替代(图 1.3)。其他的核苷酸替代皆为颠换。在大多数 DNA 片段中,转换出现的频率比颠换要高(Fitch 1967; Gojobori 等 1982;Kocher 和 Wilson 1991)。在蛋白质编码基因中,仍为同义密码子的核苷酸替代称为同义或沉默替代(synonymous 或 silent substitution),而导致非

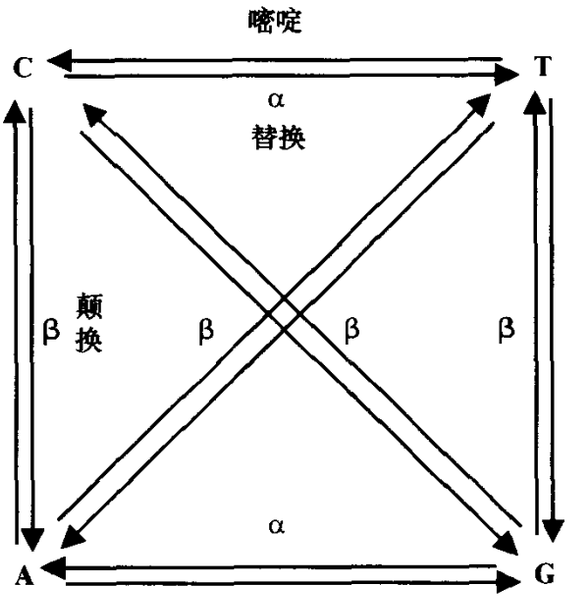


图 1.3 核苷酸的转换(A↔G 和 T↔C)和颠换(其他)(α 和 β 分别为转换和颠换率)