

第一篇 分子生态学

第一章 分子生态学

学习要点

1. 了解分子生态学产生的历史背景。
2. 初步掌握对分子生态学研究具有重要指导作用的理论知识。
3. 理解分子生态学与其他学科的关系及其地位。
4. 初步了解分子生态学的研究方法、手段和特点。

张德兴. 2008. 分子生态学。

见:《现代生态学》第二版(戈峰主编)

北京: 科协出版社, 8-35。

基本概念

1. 中性突变 (neutral mutation): 大多数分子水平的遗传变异 (如 DNA 序列中的碱基取代), 在选择上是中性的, 即它们并不影响生存适度 (fitness), 其命运主要是由随机遗传漂变而不是自然选择决定的。这些遗传变异被称为中性突变。这一进化理论被称为中性理论。中性理论是解释分子进化的理论, 不适于解释其他层次的进化现象, 例如, 大多数形态、生理、行为性状差异的形成应该都是通过自然选择而实现的。

2. 负选择和正选择 (negative selection and positive selection): 能降低生存适度的突变称为有害突变 (deleterious mutation), 它们在选择中处于劣势, 因而自然选择向将其从种群中淘汰的方向进行, 这种选择称为负选择。偶尔也会发生能提高生存适度的突变, 称为有利突变 (advantageous mutation)。有利突变在选择中处于优势, 因而自然选择倾向于把它们在种群中固定下来, 这种形式的选择叫正选择。

3. 固定 (fixation): 指等位基因在种群中的频率达到 1, 即种群的所有个体在该位点上都是同一等位基因的纯合体。

4. 位点 (locus): 遗传学上泛指染色体上为一个基因所占据的位置; 在分子生态学中指染色体上为一个 DNA 分子标记 (不管编码与否) 所占据的位置。

5. 基点 (site): 泛指 DNA 序列中的某个核苷酸 (碱基) 所占据的位置或蛋白质序列中某个氨基酸残基所占据的位置。例如, 某基因的第 146 个核苷酸在群体中观察到了由 A 到 G 的点突变, 可以表述为该基因在基点 146 存在多态性, 146 是多态基点。

6. 谱系 (lineage): 指具有连续共同进化历程、享有共同祖先的一个支系; 它可以是一组亚种群, 可以是一个物种, 也可以是一组物种。

7. 单倍型 (haplotype): 具有独特遗传特征的、连锁的 DNA 序列; 一个位点的每个等位基因即是一个单倍型, 不同位点的、位于同一条染色体上的等位基因也组成一个单倍型。对于线粒体 DNA, 单倍型通常被称为线粒体型 (mitotype)。

8. 基因流 (gene flow): 指基因通过个体迁移或其他途径在种群间的传播、交换。

9. 随机遗传漂变 (random genetic drift): 指种群中等位基因频率或基因型频率受随机抽样误差影响在世代间的波动。又简称遗传漂变。

10. 搭载效应 (hitchhiking effect): 指一个等位基因频率的改变不是因为它本身受选择影响, 而是因为与它连锁的另外一个位点受到选择而被“牵连”的现象。

11. 非同源相似 (homoplasy): 指性状的等同状态是通过不同进化途径形成的巧合。趋同进化、平行进化、回复突变等都可以导致非同源相似。

12. 协同进化 (concerted evolution): 指在进化中保持基因家族成员间核苷酸序列等同的分子进化机制。也就是说, 服从协同进化规律的基因家族, 其成员的序列组成在进化中会保持等同一致。为减少混乱, 作者认为生态学中描述种间相互适应、相互呼应进化关系的“co-evolution”译为“互应进化”似更合适(目前也被译为协同进化)。

13. 假基因 (pseudogene): 指一个基因的失去功能活性的拷贝。

14. 新达尔文主义 (neo-Darwinism): 20 世纪上半叶发展完善的进化学派, 认为突变向种群中引入新的基因, 因而为进化提供原料, 进化就是种群中基因频率的改变, 自然选择是改变种群基因频率的最重要机制。个别学者甚至认为自然选择是唯一的进化机制。又称为现代综合派 (modern synthesis), Fisher、Wright 和 Haldane 是这一学派最重要的奠基人。

15. 其他重要概念: 中性理论、哈德-温伯格原理、种群分化、溯祖理论等请参见正文中的解释。

近十多年以来, 许多宏观生物学问题, 特别是种群遗传学、生态学、进化生物学、保护生物学和生物地理学问题的研究, 从理论机制到实际操作等方面都取得了长足的进展, 这在很大程度上归功于分子生态学研究方法的全面引入。可以说, 分子生态学的兴起, 引发了一场宏观生物学研究的革命。由于篇幅的限制, 在此很难对分子生态学这一学科领域进行详细介绍, 只能对其核心部分作一纲要性概述。希望通过阅读本章, 读者能对分子生态学的范畴、理论、研究方法和特点有一个初步了解, 从而为更深入地学习形成一个基本思维框架。

第一节 分子生态学概述

一、分子生态学的概念

对“分子生态学”给出一个明确的定义比较困难, 因为分子生态学是多门学科交叉学科, 是多学科交叉、融合的桥梁学科, 因此, 也可以说是多个宏观生物学学科的分支学科。该学科的内涵不应“顾名思义”地理解为“分子水平的生态学”, 它也不是“分子生物学技术+生态学”的集合。因此, 我们建议读者不必急于寻求定义, 而是在阅读和理解了本章以后, 自己给它下一个定义。

分子生态学是多学科交叉的整合性研究领域(图 1-1-1), 应该理解为“宏观生物学的分子遗传水平上的进化分析”, 从研究角度概括而说, 就是综合运用分子进化和群体遗传学的理论、分子生物学的技术手段、系统发生学和数学的分析方法以及其他学科的知识(如地学、古气候学等)去研究群体中分子水平遗传变异和谱系的分布格局、形成机制和时空演化规律, 以及相关的种群、进化、生态、行为、分类、生物地理演化、生

分子生物学

古生物学

系统发生学

古地学/古气候学

分子进化

数学

1 2 3 4 5 6

图中心阴影圆所覆盖的研究领域都属于分子生态学的研究范畴。数字的含义为：1=群体遗传学；2=生态学；3=进化生物学；4=生物地理学；5=行为生物学；6=保护生物学。箭头表示主要支撑性学科的渗透、融合。“整合”指分子生态学是多学科交叉、融合的桥梁学科，是多门学科的分支学科。

国内介绍分子生态学的文章和资料多以 1992 年为“分子生态学”的诞生年，因为 *Molecular Ecology* 杂志该年在英国正式创刊。其实这是一个误解。从前一段的解释可以看出，“分子生态学”这一术语从字面上不能涵盖本学科领域的全部研究内容，因而并不是一个合适的名称。实际上，分子生态学所覆盖的不少领域的研究早在 20 世纪 50

年代后期就已开展(见后述),并用近40年的时间形成了一个综合的理论、技术、分析体系,为1992年以后的发展奠定了扎实的基础。可以说,分子生态学研究起源于早期的分子水平的群体遗传学研究和进化生物学研究,例如20世纪60年代的蛋白质多态性的研究(Lewontin and Hubby, 1966),它只是在1992年因 *Molecular Ecology* 创刊而得到了一个并不恰当的“俗名”并以此而广泛流传。

三、学科发展历史简介

分子生态学研究萌芽于20世纪50年代,其发展历史就是运用分子遗传标记研究宏观生物学问题的历史,也就是分子生物学、群体遗传学、分子进化交叉融合的历史。促成和不断加深这一融合的因素,一方面是分子生物学技术的不断突破,另一方面是分子进化理论和数据处理方法的建立和改进。下面通过列举20世纪50年代以来有关领域一些里程碑性的成就来回顾分子生态学研究发展的历史。

20世纪50年代中期,淀粉凝胶电泳技术发明(Smithies, 1955),并随之建立了蛋白质组织化学染色方法(Hunter and Markert, 1957)。二者结合使得大规模、快速、定性的蛋白质多态性分析成为可能(Lewontin and Hubby, 1966),从而掀起了自然种群遗传变异研究的第一个高潮,即60、70年代的同工酶分析时代。

20世纪60年代中后期,有两大事件,即:①分子进化的中性理论诞生(Kimura, 1968),为怎样认识遗传多态性数据提供了理论指导;②限制性内切核酸酶被发现(Linn and Arber, 1968; Meselson and Yuan, 1968),为限制性片段长度多态性(RFLP)分析提供了必要的酶学工具。

20世纪70年代有四大成就,对分子生态学的发展有重大影响。①DNA转膜杂交技术的建立(Southern, 1975);②动物线粒体DNA遗传变异性的发现(Brown and Vinograd, 1974; Avise *et al.*, 1979);③DNA序列分析技术的发明(Sanger *et al.*, 1977; Maxam and Gilbert, 1977);④DNA克隆技术的建立(体外重组)(Maniatis *et al.*, 1978)。

20世纪60、70年代的突破,催生了遗传变异分析的RFLP时代,特别是动物线粒体DNA的RFLP分析时代,并使从最基本层次分析遗传变异——即DNA序列分析进入群体遗传学研究计划中。20世纪70、80年代的遗传变异分析主要是同工酶多态性和DNA的RFLP多态性分析,并集中于对模式动物果蝇和人的研究。

真正推动分子生态学研究走向大众化,得以在众多实验室普遍开展的,是20世纪80年代的划时代发明和发现,包括①DNA聚合酶链反应(PCR)的发明;②热稳定DNA聚合酶的发现(Saiki *et al.*, 1985; 1988)。PCR的发明预示着一场DNA分析革命来临,而热稳定DNA聚合酶的发现正是促成这次革命的“东风”。从此,研究人员不必通过分子克隆就可以从微量样品出发分离、制备大量用于后续操作的目的DNA样品,并在很短时间内进行大规模样本分析。在这种时代背景下,“*Molecular Ecology*”杂志酝酿创刊,赋予了有近40年研究历史的一系列研究领域一个通俗的总称——“分子生态学”,这些研究领域也在20世纪90年代全面进入一个革命性的时代——PCR时代。我们正处在这一研究时代的高潮之中。

以上介绍的主要是贯穿分子生态学学科领域发展过程并在其中起举足轻重作用的分子生物学技术方面的突破。“实验技术”只是分子生态学赖以发展的保障之一，除此之外，尚需其他方面的保障，包括分子标记、理论和数据分析方法方面的保障。

第二节 分子生态学的理论基础

分子生态学研究的核心内容可概括如下：检测遗传变异，分析和解释遗传变异，揭示遗传变异所反映的规律，从而进一步阐明生物以及生物和环境之间是如何相互作用的，并推测生物演化到现今状态的历史过程以及未来可能的演变趋势。因此，分子生态学不仅仅是生态学的一个分支学科，也是群体遗传学、进化生物学、行为生物学、保护生物学、生物地理学、生物多样性科学等宏观生物学众多学科的分支学科。在图 1-1-1 已经形象化地描述了这种关系。

分子生态学是建立于群体遗传学理论、分子进化理论和谱系演化理论等理论支柱之上的实验性和理论性都很强的科学，其理论是交叉于这些领域的关于群体中遗传变异和谱系的分布格局、形成机制和时空演化规律的理论。因此，分子生态学将各交叉学科的理论融汇成一个崭新的理论体系，并在此基础上把它们向前推进发展。正是如此，分子生态学使得许多传统学科之间的界限变得模糊，并整合了传统上不同的研究领域，这也导致有些学者产生了分子生态学没有自己的理论的认识。下面我们从七个方面对一些基本理论知识做简要介绍。

一、选择中性是分子水平大多数变异的共同特征

分子生态学研究的核心内容是种内遗传变异，即种内遗传多态性，也就是种群的遗传多样性。遗传多样性在分子水平上表现为遗传位点的 DNA 多样性程度，即同一遗传位点（基因组的某一特定区段）在种群中所存在的不同形式的基因（称为等位基因）的数量及频率。中性理论是解释分子水平遗传变异的进化机制的理论，是“分子进化的中性理论（the neutral theory of molecular evolution）”的简称，于 20 世纪 60 年代末期提出（Kimura, 1968; 1983），经过三十多年的发展和完善，现在已成为被国际学术界广泛接受的 20 世纪最为重要的进化理论。日本学者木村（Kimura）是该理论的核心奠基人。

1. 中性理论的内容

中性理论的核心内容可概括如下：分子水平上的绝大多数突变是选择上中性的，因而它们在进化中的命运是由随机遗传漂变，而不是由自然选择所决定的。“选择上中性”应该这样理解：一个新出现的遗传变异（新等位基因），与其他已经存在的等位基因相比，在选择上是等同的，或几乎等同的，因而不同的类型所承受的选择压力几乎是一样的。这里的“中性”并不特指没有任何功能。

“在进化中的命运”指等位基因在后续的进化过程中是被淘汰，还是被固定。由于不同的等位基因之间不存在选择压力的差别，因而它们的去留问题是由取样误差随机决

定的。

应该注意的是,中性理论是解释分子进化的理论,不适于解释其他层次的进化现象,例如,大多数形态、生理、行为性状差异的形成应该都是通过自然选择而实现的。

所以,中性理论与新达尔文进化学派的冲突之处不是在于承认自然选择与否,而是在于对自然选择在进化中的地位与作用的认识上的不同。新达尔文学派(即现代综合论, modern synthesis)认为自然选择是改变种群基因频率的最重要的进化机制,中性理论则认为随机遗传漂变是改变种群中中性等位基因频率的最重要的进化机制。中性理论承认自然选择的存在,认为自然选择是适应性进化的主导力量,但它不是分子进化的主要力量。

2. 中性理论对种内遗传变异的解释

种内遗传变异是分子生态学研究的核心内容之一,因而如何从理论上认识种内变异至关重要。中性理论认为,我们所观察到的分子水平上种内遗传多态性主要是由中性突变组成的,换言之,分子水平上的绝大部分种内遗传变异是选择上中性的,遗传多态性在种群中的存亡主要是由突变速率和随机遗传漂变的程度决定的。而且,从物种进化的时间长河来看,遗传多态现象体现的是分子进化中的过渡状态,也就是说多态现象是暂时的和不稳定的(遗传多态现象最终都将是种群中消失)。

为了更好地理解这一具有极高重要性的理论,下面就以东亚飞蝗的微卫星多态性为例做进一步说明。研究表明, C67 微卫星位点(locus)在飞蝗中具有高度多态性。C67 是一个 GT 二相重复微卫星位点,表示为 $(GT)_n$; n 代表 GT 重复的次数,因而 n 的大小体现了本位点等位基因的大小。从河北白洋淀东亚飞蝗种群内已检测出 8 种 C67 等位基因, n 值分别为 12、13、14、15、19、20、21、23。依据中性理论,这 8 个 C67 等位基因是选择上中性的,因而它们在后续进化过程中的命运将主要受突变和随机遗传漂变因素的影响。因为突变速率一般较低,为简单起见,其作用在此予以忽略不计,因此这八个等位基因的命运将主要由随机遗传漂变决定。白洋淀种群繁殖三代以后,由于随机取样误差,可能 $(GT)_{15}$ 和 $(GT)_{19}$ 将被淘汰,只剩下其余 6 种等位基因;再经过 10 代,可能又有三种被淘汰,仅剩下 $(GT)_{12}$ 、 $(GT)_{13}$ 和 $(GT)_{23}$ 三种;那么再过 50 代,有可能只有其中一种,比如 $(GT)_{13}$ 存在,其余的都被淘汰了。这时, C67 位点的遗传多态性便消失了。

中性理论的一个重要特点是,它可以通过随机过程理论对等位基因在种群中的命运作出量化描述,再结合群体遗传学理论,我们便可能对种群的一些重要参数、进化过程和演化趋势作出推断。

3. 中性突变与自然选择的辩证关系

怎样看待中性突变与自然选择的关系是解决中性理论与经典进化论争论的症结所在,其实二者并不相互矛盾。首先,中性理论不否认自然选择在适应性进化中的主导作用,而只是认为它不适于解释大多数分子进化现象。

其次,中性理论强调绝大多数突变的选择中性,但并不否认不同的突变型间可能存在选择压力上的潜在差别。相反,中性理论认为,当新的环境条件改变时,某个或某些

突变型（多态型）可能会变得在选择上具有（微弱）优势。因此，中性理论认为，选择上中性的遗传多态型，尽管在当前的环境条件下没有适应性意义，但它们为将来环境变化以后，在新的生态环境下的适应性进化储备了宝贵的资源。这也正是遗传多样性保护的进化学和遗传学依据。

其三，中性突变本身不受选择压力影响，但是由于遗传连锁的存在，中性突变位点会因非中性的相邻位点而受选择的影响，因此，总体上造成中性突变位点等位基因间选择压力的失衡。这就是所谓的突变固定的“搭载效应”（hitchhiking effect），这种因搭载效应而导致中性等位基因被淘汰的现象亦称为选择扫荡（selective sweep）。

其四，中性理论使自然选择由抽象变得具体，为验证自然选择的存在提供了一个“解消假设”（无效假设，null hypothesis）。一旦实验或观察数据推翻了这个解消假设，研究者就有可能通过找出其背后的原因而发现一些新的机制和规律。在此，我们举一个易于理解的分子进化方面的例子：由于昆虫线粒体 DNA 的非编码调控区（即 A+T 富集区，A+T-rich region）的大部分区段不具备任何已知功能，根据中性理论，这一区域的大部分核苷酸在进化中应该是不保守的。但是通过比较不同蝗虫和果蝇的这一区段的核苷酸序列，研究人员发现存在一个非常保守的发卡式（hairpin）二级结构（图 1-1-2）。由此可以推断，这一保守的二级结构很可能具有某种重要功能，因而是受选择影响的。综合分析相关的研究成果，推测此二级结构可能是线粒体 DNA 的复制起点。

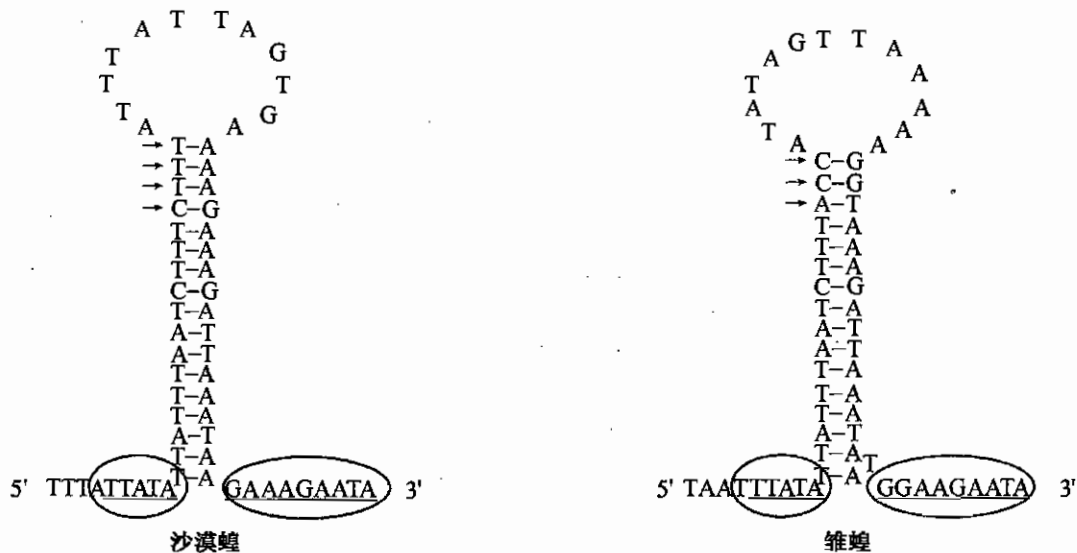


图 1-1-2 蝗虫线粒体 DNA 的非编码调控区（A+T 富集区）的一个非常保守的发卡式（hairpin）二级结构。（依 Zhang *et al.*, 1995 修改）

4. 中性理论对分子生态学研究的指导意义

种群的相互关系和演化过程是受多因子影响的复杂问题，这些因子包括选择、突变、随机遗传漂变、迁移、自然灾害、社会结构、种群统计参数（出生率、死亡率、婚配比例等）等。为了便于从遗传学上进行分析，所选择的研究体系需尽可能地简单，主要影响因子越少越好。例如，欲从遗传上研究我国东亚飞蝗各地理种群的遗传结构和相互关系，需要考虑的影响因子首先可以简化为四个，即选择、突变、随机遗传漂变和迁

移。其中，第四个因子“迁移”正是我们所要探索的问题（种群的相互关系）的一个方面，这样在余下的三个因子中，“突变”的影响可以忽略不计，因为突变速率一般都很小，在所感兴趣的时间尺度内不足以显著地改变种群的遗传结构。那么剩下来所需要考虑的便是选择和随机遗传漂变两个方面了。尽管如此，问题仍然十分复杂，因为选择的作用方式、程度等是进化生物学中最难弄清的问题。所以，必须避开“选择”的影响。那么最有效的捷径是使用选择上中性的分子标记，这样我们就不必去考虑由选择压力差别所产生的影响了。“中性分子标记”是开展大多数分子生态学研究必须满足的条件，这就是为什么说，是否有一套合适的分子标记是分子生态学研究能否取得满意成果的关键。

一旦问题简化为“随机遗传漂变”过程，在获得足够的具有代表性的分子数据以后，我们便可以利用群体遗传学理论来做深入分析，探索我们的问题的可能答案。以下几节将对一些关键群体遗传学理论作简要介绍。

二、基因频率和基因型频率是反映种群动态变化的基本参数

生物进化论在 20 世纪转变成一门可以追踪、检验、证明的科学，这得力于对其基本作用过程的阐明——进化表现为种群中基因频率的改变。因此，基因频率和基因型频率成为反映种群动态变化的最基本参数。群体遗传学的奠基性原理之一，哈德-温伯格原理（Hardy-Weinberg principle）就是关于种群基因频率和基因型频率变化规律的原理，几乎所有其他有关有性生物种群进化的遗传学理论都是在其基础上建立起来的，分子生态学的很多数据分析方法也都以哈德-温伯格平衡为前提假设条件。

1. 哈德-温伯格原理：群体遗传变异研究的核心理论

哈德-温伯格原理可以简明地表达为：在满足下述假设条件的前提下，不管群体的初始基因型频率如何，经过一代的随机交配，等位基因频率和基因型频率将达到平衡，即在以后的世代中保持不变。换句话说就是，有性生殖不改变种群的等位基因频率和基因型频率。假设条件是：

- 生物进行有性繁殖且随机交配；
- 等位基因在雌雄两性中随机分布；
- 种群足够大（理论上可视为无穷大）；
- 世代没有重叠；
- 没有自然选择、突变和迁移。

对于具有世代重叠的生物，或两性中基因分布不等同的情况（如性染色体上基因），基因型频率和等位基因频率同样会达到哈德-温伯格平衡，但需要若干代的随机交配。

2. 哈德-温伯格原理对分子生态学研究的意义

首先，就像中性理论是检验是否存在“选择力量”的解消假设一样，哈德-温伯格模型是检验种群演化过程中是否存在其他进化力量的解消假设。哈德-温伯格原理实质上解释的是生物有性繁殖的规律，当研究发现某个种群不处于哈德-温伯格平衡时，我

们就可以推测其他进化力量可能在起作用，如可能存在种群间迁移，种群在前一代可能经过了瓶颈效应，所研究的遗传位点可能受选择影响等等。值得注意的是，如果前述的各项假设条件得到满足，有关位点将表现出哈德-温伯格平衡，但如果观察到一个位点符合哈德-温伯格平衡，我们并不能据此断定上述假设条件都得到满足。例如，很弱的自然选择力量，通常并不影响观察数据符合哈德-温伯格平衡。还需指出的是， X^2 检验方法有时难以检测出观察数据与期望值间的偏离。

其次，哈德-温伯格原理揭示的一个极为重要的方面是当没有其他进化力量作用时，群体的基因频率和基因型频率将永远保持恒定，因此孟德尔遗传本身能够保存种群的遗传多样性。这便是其深远意义之所在。正是在这样一个保证下，突变、选择、基因流等进化力量才能使生物不断地向更加丰富多彩演化。

哈德-温伯格原理隐含的另一个重要方面是，稀有等位基因主要以杂合体的形式存在于种群中。其意义为，即使该稀有基因对生物体是有害的，它也能得以在种群中保存较长时间（纯合体很容易被淘汰）。假设某个位点有几个等位基因，即 a_1 、 a_2 、……、 a_n ，令 p_i 为第 i 个等位基因在群体中的频率，假设 a_r 是稀有等位基因，亦即 p_r 在 $(p_1, p_2, \dots, p_n)$ 中最小，则杂合体 $a_r a_i (i \neq r)$ 与纯合体 $(a_r a_r)$ 在群体中的比率 $h = \frac{2(1-p_r)}{p_r}$ ，由于 $p_r \ll 1$ ，所以可以看出杂合体 $a_r a_i$ 的数量远大于 $a_r a_r$ （若令 $p_r = 0.01$ ，则 $h = 198$ ）。

许多因子会影响哈德-温伯格平衡，从而影响哈德-温伯格原理在实际问题中的适用性，建议有兴趣的读者自己进一步探索。

三、种群分化是生物进化的必要途径

为了避免混淆，我们首先对两个术语的含义进行规范化约定。种群（population）是指一个物种所有个体的集合，亚种群（subpopulation）指受特定地理环境、生态环境或生活习性等因素影响而相对独立的一个繁殖区内的个体的总和，有时也称“地理种群”（local population）。所以，可以说“种群”是所有“亚种群”的集合。

由于生态地理环境的差别、生活习性的差别或天然屏障的隔离，几乎所有生物的种群都会发生“分化”，即都会形成若干个亚种群，我们把一个物种种群分化成几个亚种群的现象称为种群分化（population subdivision）。种群分化必然导致各亚种群间遗传组成上的差别，即等位基因频率上的差别，这一过程称为种群的遗传分化（genetic differentiation）。发生遗传分化的种群往往具有层次性结构（hierarchical structure），即整个种群是由若干个有序的亚种群组成，而亚种群又由若干个更低级的有序单元组成，依次类推；“有序”在此指同一层次的同类单元间相对独立，且其所包含的更低级的单元间也相对独立。

1. 种群分化的后果

种群分化的进化后果是导致新亚种或新物种的形成。生物进化的质变在于新种的形成，种群分化是向形成新物种所迈出的第一步。所以弄清种群分化的过程、机制及其影

响因子对于从本质上认识生物演化非常重要。

种群分化的生态学后果因具体情况不同而异,它有可能提高物种的适应性,也能加速其灭亡。

种群分化在遗传学上所表现出的一个显著特征是种群的杂合度降低,因为发生种群分化以后,近交的机会增加了。读者可根据前述哈德-温伯格原理,自己推导一下这个结论。利用这一特征,我们可以检验种群是否发生了遗传分化,这涉及下一部分所要介绍的著名的赖特 F 统计方法 (Wright's F statistics)。

2. 赖特 F 统计方法

我们可以把发生遗传分化的种群区分成几个不同的研究层次,如个体、亚种群和整个种群;把相对应这三个层次的、哈德-温伯格平衡下的平均杂合度分别表示为 H_I 、 H_S 和 H_T 。

下面我们引入群体遗传学中一个非常重要的概念:Wright 的固定系数 (fixation index) F ,初学者可能感到比较难于理解 F 的含义以及为什么称之为固定系数。我们从不同的角度予以解释如下:

从种群杂合度的角度理解, F 代表在随机交配的条件下,一个层次单元(比如亚种群)相对于更高级层次单元(比如整个种群)杂合度期望值减少的程度。

从近交的角度理解, F 代表相对近交系数,即一个层次单元相对于更高级层次单元的近交程度。

从基因在种群中的分布的角度理解, F 近似地代表等位基因在种群中被固定的相对可能性,因此, F 被 Wright 称为固定系数。

现在我们定义 F_{ST} 为亚种群相对于整个种群的近交系数,则有

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

类似地, F_{IS} 为个体相对于它所属的亚种群的近交系数, F_{IT} 为个体相对于整个种群的近交系数,于是有

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

这三个层次性统计参数之间具有如下关系:

$$(1 - F_{IS}) \times (1 - F_{ST}) = (1 - F_{IT})$$

Wright 的 F 统计法是检测种群分化的最常用的方法,其中 F_{IT} 反映了种群整体趋异的程度,而 F_{ST} 是检测种群分化程度的重要参数。

值得指出的是,上述三个 F 参数的估计数值因研究体系而异,即不同物种之间 F 参数的估计值可能会有很大差异,因而没有绝对可比性;同时对同一物种而言,不同遗传位点的估计值亦会有很大差异,因而也缺乏可比性。所以,在比较不同研究的结果时,应注意研究体系间的差别。

另外, F 统计方法是基于等位基因频率数据而计算的,对于其他遗传数据,如 DNA 序列数据,或微卫星 DNA 标记数据,可以根据具体情况,对 F 统计方法做适当

修正, 比如使用 N_{ST} (Lynch and Crease, 1990) 或 R_{ST} (Slatkin, 1995) 等统计量。

3. 影响种群分化的因子

很多生物和环境因子会促进种群分化, 增加亚种群间的遗传差异, 这些因子包括自然屏障引起的隔离、距离隔离、交配制度、自然选择、突变、遗传漂变等等。从群体遗传学和进化的角度看, 各种形式的隔离(自然屏障隔离、距离隔离等)的效果都是阻断或降低了基因流, 加强了近交程度; 特殊的交配制度[如选型交配 (assortive mating)、等级制度等]和自然选择都是通过破坏了种群的随机交配而加速亚种群间的遗传分化; 突变通过直接增加遗传变异而加大亚种群间的遗传差异, 而遗传漂变通过随机减少现有的遗传变异而加大亚种群间遗传组成上的差别。由于这些因子的作用方式不一样, 所以它对一个给定种群影响也不一样, 比如, 大小不同的种群对它们的敏感性不一样, 读者可以自己作进一步分析。

四、随机遗传漂变是种群分化的重要进化动力

在没有任何其他进化动力存在的情况下, 两个起始状态完全等同的地理种群, 在完全等同的生态地理环境中经过若干世代的繁殖, 如果两地理种群间不发生任何基因交流, 则这两个地理种群也会发生遗传分化。促使其发生分化的动力是随机遗传漂变, 即由随机性取样误差所引起的等位基因频率在不同世代间的波动。种群中等位基因频率的随机波动, 最终必将导致一个被固定而其他被淘汰, 在两个起始状态完全等同的种群中, 由于相互独立(即彼此间不存在基因交流), 各个等位基因的频率会发生变化, 最终被固定下来的等位基因也会不尽相同, 因而造成两个种群遗传组成的差别, 即遗传分化。

由随机遗传漂变引起的种群进化具有如下特点:

(1) 小种群比大种群发生漂变的速度快, 因而等位基因在小种群中被固定的平均时间比大种群短。

(2) 在任何时间, 一个等位基因被固定的概率等于该等位基因此时在种群中的频率, 所以尽管稀有基因更容易被淘汰, 但它仍有一定的概率在种群中固定。

(3) 随机遗传漂变会降低种群的遗传多样性, 在没有其他具有相反效应的进化力量同时发生作用时, 随机遗传漂变最终会导致种群遗传多样性的消失。

(4) 新产生的突变在种群中被固定下来的概率等于它在种群中的频率, 因此, 新突变在小种群中被固定的可能性大于其在大种群中的可能性。

(5) 对于一个集合种群 (metapopulation) 来说, 若各局部种群 (local population) 间没有基因交流, 则局部种群越小遗传多样性丧失得越快, 局部种群间的遗传分化也越大。

(6) 随机遗传漂变的另一个重要特点是, 它同时作用于所有的遗传位点, 因此, 理论上在没有其他进化力量影响时, 不同的中性位点应该揭示出相同的演化规律, 当然这还与所研究的种群所处的进化阶段有关。

五、溯祖理论是追溯谱系演化历史的核心理论

溯祖理论 (coalescent theory) 可以说是 20 世纪后半叶理论群体遗传学的里程碑性成果, 对分析分子生态学研究中的分子数据极为重要。由于其论述比较复杂, 有兴趣的读者可以阅读 Hein 等 (2005) 和 Hudson (1990) 的综述以及那里给出的相关文献。现在我们从进化研究的角度, 特别是种群研究的角度进行简要的讨论。

1. 溯祖理论

一个种群是由不同的个体组成, 换一个角度来看, 它则是由不同的遗传谱系组成。我们知道, 对于任何一个物种来说, 现存的所有个体都是由其祖先一代代繁殖而来, 因此, 种群中的所有谱系都是由祖先谱系由远及近地传递而来。那么, 反观这个谱系传递过程, 两个谱系会在某一代回溯到它们的最近共同祖先 (most recent common ancestor, MRCA), 这一事件被称为“溯祖事件”。溯祖理论实质上是关于谱系在历史演化过程中的相互关系及其动态变化的理论, 它指出在经过一系列溯祖事件后, 最终所有的谱系都归结于一个最近共同祖先。这个过程称为溯祖过程 (coalescent process)。溯祖过程可以被看作是“分支过程” (branching process) 的逆过程。

下面以一个假想的例子来对溯祖过程做进一步说明。假想图 1-1-3 是青蛙线粒体基因的谱系在 t 个单位时间 (世代) 内的演化关系图, 最上部即代表当前种群中所观察到的线粒体基因谱系 (即 $t=10$ 时,) 共有 11 个谱系; 最下部为 10 个世代前, 即 $t=0$ 时种群的遗传谱系组成情况, 即共有 7 个奠基谱系 (2 个 a_1 型、4 个 a_2 型、1 个 a_3 型)。如果我们从 $t=10$ 开始逆时倒推现有谱系的传递过程, 可以看到所有这些谱系均来源于 $t=0$ 时的同一个 a_2 谱系, 即如图中加粗的黑线所示。也就是说所有现存基因 a 的谱系均“溯祖”于 10 个世代以前的一个共同祖先 a_2 。从图 1-1-3 可以看出, 在 $t=7$ 时, 群体中尚存 5 个不同的谱系, 在 $t=5$ 时, 有 3 个谱系。在 $t=0$ 时曾存在的其他谱系, 在种群演化过程中均相继被淘汰, 对现今所观察到的线粒体 DNA 遗传谱系没有任何直接贡献。

不难看出, 溯祖过程所需要的时间与种群大小有关。我们选择一个最简单的种群演化模型对此作进一步的量化分析, 即随机遗传漂变是唯一的影响因子。假设种群在各世代都保持一个恒定的体积, 使得每一世代都包含 N 个拷贝基因 a , 可以算出种群中任意一对基因 a 在 t 个世代以前溯祖的概率为

$$P = \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^{t-1}$$

这一分布的均值即是种群中任意对基因溯祖的平均时间, 对于单倍体来说为 N 个世代, 对于二倍体则为 $2N$ 个世代。相应地, 种群中所有基因 a 拷贝溯祖到一个共同祖先的平均时间, 对于单倍体生物为 $2N$ 个世代, 对于二倍体则为 $4N$ 个世代。显然, 种群越小溯祖过程进行得越快。

2. 样本的溯祖分析

细心的读者从上面段落中不难看出, 溯祖过程与随机遗传漂变过程有某种内在联

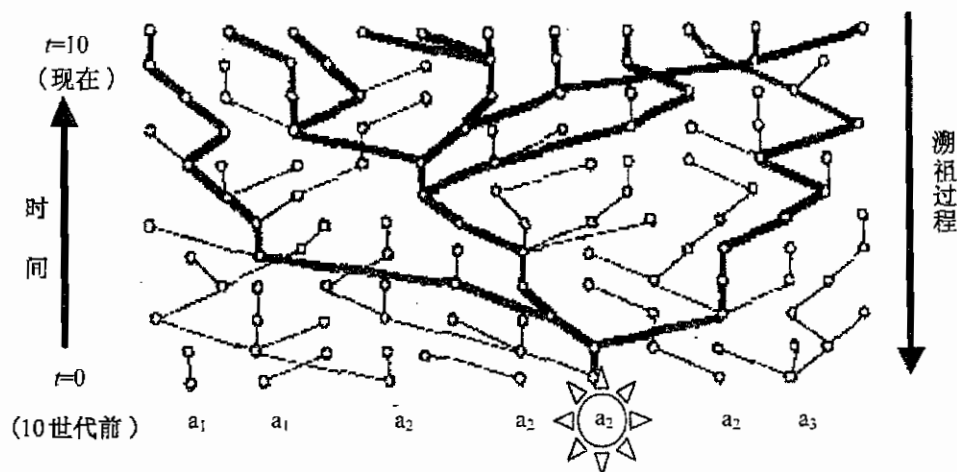


图 1-1-3 溯祖过程示意图 (自 avise, 2000 修改)

假想的青蛙线粒体基因的谱系在 t 个单位时间 (世代) 内的演化过程注意图中连线代表世代传递关系, 连线的交叉不代表任何进化过程, 而仅仅是为了节省制图空间。10 个世代前共有 7 个奠基谱系 (2 个 a_1 型、4 个 a_2 型、1 个 a_3 型), 现在 (即 $t=10$ 时) 共有 11 个谱系。如果我们从现存的 11 个谱系逆时回推, 可以看到所有 11 个谱系最终回溯于 10 个世代前的一个共同祖先谱系 a_2 , 而其他 6 个奠基谱系在进化中均先后消亡。生物谱系终会溯祖到一个共同的祖先是一个具普遍意义的规则

系。实质上, 二者是时间上互为逆向的分析过程, 是从不同的角度研究同一个问题。随机遗传漂变分析探讨种群中一个基因最终被固定下来所需要的时间 (或其概率), 而溯祖分析研究种群现有的等位基因若干世代以前享有一个共同祖先的可能性。因此, 溯祖分析实际上涉及两个互为独立的随机过程, 一个是谱系的传递过程, 一个是突变 (新谱系) 的产生过程。

溯祖分析在种群进化分析中的重要性在于在该理论的指导下, 我们可以建立一些非常有效的数学算法, 通过对种群的样本的分子遗传数据的分析而进行参数估计, 或通过模拟对种群的演化历史进行推测, 揭示影响种群进化的可能的主导因子或特征, 比如随机遗传漂变是否是最主要的影响因子, 是否有自然选择的存在, 若存在, 是正选择 (positive selection), 或负选择 (negative selection), 亦或平衡选择 (balancing selection)? 在历史演化过程中, 种群大小是恒定的还是不断增大的, 以及祖先种群的大小、自交率等等。溯祖分析要求有遗传谱系数据, 这些数据可以是种内多态性数据, 也可以是种间遗传分化数据。由于分子水平的遗传数据, 特别是 DNA 序列数据, 隐含着所研究的样本间的同祖关系 (ancestral relationship), 是最深层次上的遗传谱系数据, 所以特别适宜于溯祖分析。

溯祖分析都是基于一定的种群模型 (例如 Fisher-Wright 模型、Moran 模型) 而进行的。但现实种群很少能完全符合这些模型的前提假设条件, 好在研究表明当种群大小足够大时, 偏离这些前提假设不会带来很大影响, 因此保证了溯祖分析在种群研究中的可靠性。另外, 由于我们只能对种群的样本而不是整个种群进行研究, 所以样本的大小对溯祖分析的可靠性会有影响, 当样本过小时, 所估计的溯祖时间会有较大偏差。但是, 在样本量达到一定大小后, 继续加大样本量便不再有显著的效应。此外, 不同的分子标记的进化过程有差别, 依赖于少数几个分子标记所进行的溯祖分析亦会有很大的误差。

不同的因子对溯祖过程有不同的影响,因而研究中需根据具体情况使用不同的数学模型。这些因子包括种群增长、随机遗传漂变、自然选择、重组、近交程度等。数学模型的建立是溯祖分析重要领域之一,计算机模拟已被广泛用于检验数学模型的可靠性,一旦确立了置信度较好的理论模型,则可以根据观察数据估计一系列因子参数,如重组率、近交率、种群遗传多样性、种群增长模型、溯祖时间等。

六、分子系统发生分析是重塑生物类群进化关系的重要工具

种群差别的研究是同中求异,属于比较棘手的问题。而系统发生分析(phylogenetic analysis)则是异中求同,即在已知所研究的类群不同的情况下,揭示类群间的历史进化关系。但是,随着分子水平研究的渗入,种群差别分析与系统发生分析相互交叉融合的趋势不断加强,系统发生分析方法也被用于种群研究之中。另外,分子系统发生分析也开始被广泛用于物种分类的研究,形成所谓分子分类学(molecular taxonomy)方法。

分子系统发生分析建立于几个重要原则之上,包括:

- (1) 所使用的分子标记服从中性进化规律,即是选择上中性的。
- (2) 在所感兴趣的时间幅度内,该分子标记在不同类群中的进化速率是恒定的,即符合“分子钟”假设。
- (3) 所使用的分子标记在所有研究类群中都是直向同源的(orthology),不能是并行同源(paralogy)或外源同源(xenology)的关系。
- (4) 分子标记具有合适的遗传变异度,即变异性既不能太低也不能太高。

从方法论上看,分子系统发生分析包括以下几种:距离法(distance method)、简约法(parsimony method)、最大似然法(maximum likelihood method)和贝斯法(Bayesian method)。每种方法各有其优缺点,分别简要介绍如下。读者需仔细阅读其他有关文献资料,以便正确掌握它们的使用方法。

1. 距离法

距离法通过计算不同单元(taxon)之间的两两进化距离而对它们的相对进化关系进行比较,进化树是依据每对单元间的进化距离数值而建立的。距离法有很多种不同的算法,各种算法的表现和准确性也不同,在此不做详细介绍。一般来说,读者在使用各种距离算法时应考虑如下几点:

- (1) 各单元间的进化距离与所观察到的突变数量是否呈线性关系。
- (2) 距离算法的标准误差的大小。
- (3) 应使用不同的算法分别计算,如果算法之间在距离数值上没有显著差别,则算法越简单可能越可靠。

(4) 距离法有一些自身的缺陷,即当把DNA序列数据转化为进化距离数据时,相当一部分有用的信息便被丢失了。因此距离法仅使用了分子数据中的部分信息。一个典型的例子便是不同的数据组可以产生同一个距离矩阵,但由距离矩阵却不能推算出相应的数据组。

2. 简约法

我们介绍简约法中的最大简约法 (maximum parsimony)。该方法的理论依据是如下哲学观点: 解释一个过程的最好假说应是假设条件最少的假说。因此, 对于分子数据 (DNA 序列数据), 该方法计算在最少的碱基替换条件下能够解释整个进化过程的拓扑结构 (系统进化树)。

最大简约法也有若干不同的算法, 相对距离法而言, 最大简约法基本上不依赖任何模型, 所以受假设条件的影响较小。理论上, 在没有非同源相似 (homoplasy) 影响、所分析的核苷酸总数足够大的情况下, 最大简约法可以找到正确的进化树。但在实际运算中, 情况并不如此简单, 长枝吸引 (long-branch attraction) 现象便是最大简约法可以产生错误结果的典型例子。

尽管简约法充分运用性状分析 (character analysis) 来寻找最好的拓扑结构, 但像距离法一样, 它也并不能完全利用观察数据中所包含的所有信息。例如对于 DNA 序列数据而言, 只被观察到一次的核苷酸替换点 (称作孤频点, singleton site) 对于简约分析没有意义, 因而被忽略不计。另外, 当类群很多而性状差别很小时, 最大简约法的可靠性会大大降低。

3. 最大似然法

最大似然法所考察的是哪种进化历史模型 (即拓扑结构) 能更好地符合所观察到的数据, 其所遵循的原则是 “具有最大可能性符合所观察到的数据的模型应是最合理的模型”。在一个给定的碱基替换模型下, 最大似然法对观察数据符合每种拓扑结构的可能性进行最大化计算, 可能性最大的拓扑结构被选为终解进化树。这里可能性最大化处理所考虑的是枝长 (branch length), 而不是拓扑结构本身。与距离法和简约法相比, 最大似然法一般被认为更稳定、更可靠。原来由于其计算量很大, 当分析类群比较大时, 此法实际上难以运用, 特别是后续的置信统计检验则更难以进行; 为了克服这些弱点, 现在已发展了一些较快的算法软件, 例如 PHYL (Guindon and Gascuel, 2003), 可以较快速地进行中等大小数据组的最大似然分析。

最大似然法算法本身隐含一些问题, 主要在于可能性最大化估计的对象不是拓扑结构本身, 因而最后被选为终解的拓扑结构的可能性是间接计算出来的——即通过对其枝长的估计而推断出的。

4. 贝斯法

生物进化、种群演化等都是非常复杂的问题, 我们可能永远都无法弄清楚有关影响因素的确切作用过程。近来, 贝斯法被认为对分析类似的复杂问题具有特殊价值, 并在分子系统进化分析中被广泛应用。贝斯法充分运用观察数据本身所含有的先验信息或理论所建立起来的已被其他数据证明的模型 [即先验分布 (prior distribution)], 来对观察数据进行指导性分析, 对未观察到的有关参数进行估计 [即后验分布 (posterior distribution) 估计], 有关结论则是基于后验分布而推断出来的。

与传统统计方法相比, 贝斯法是从正面直接研究有关问题, 即对某一复杂事件或某

一具有很大不确定性事件的概率进行直接估计，而不像传统统计分析那样对一个解消假设进行检验。例如，我们需要研究一个种群是否处于哈德-温伯格平衡。传统的方法则首先假设种群处于哈德-温伯格平衡，然后用观察数据检验这一解消假设是否受到支持，而贝斯法则去检验观测数据与哈德-温伯格平衡的偏离是否大到否定平衡的存在，或者现有数据是否足以用于作出有关结论 (Shoemaker *et al.*, 1998)。

最后需要指出的是，用距离法、简约法或最大似然法计算出进化树后，还需要对进化树的置信度进行估计，常用的方法有自举法 (bootstrapping) 和自减法 (jackknifing method, 又译为大折刀法)。对于种群水平的分析，则不能仅依赖于这些方法，特别是遗传分化水平较低时，通常还需要用计算机模拟来估计结果显著性。

七、多歧分化和孤频变异是种内遗传变异的显著特征

与种间遗传分歧相比，种内遗传变异有两个显著特征：①种内很多谱系间是多歧分支的关系，即三个或更多的谱系具有相同的遗传差异水平，享有同一个最近共同祖先 (图 1-1-4)。而上述系统分析方法多遵循“二歧分支”原则处理类群间的系统进化关系，因此对于种内谱系通常是不适宜的。②种内遗传变异的另一个显著特征是存在孤频变异，即很多变异只被观察到一次，在核苷酸基点上表现为许多突变只在一个序列中被观察到，在单倍型 (等位基因) 上表现为许多单倍型 (等位基因) 在种群中只出现一次。

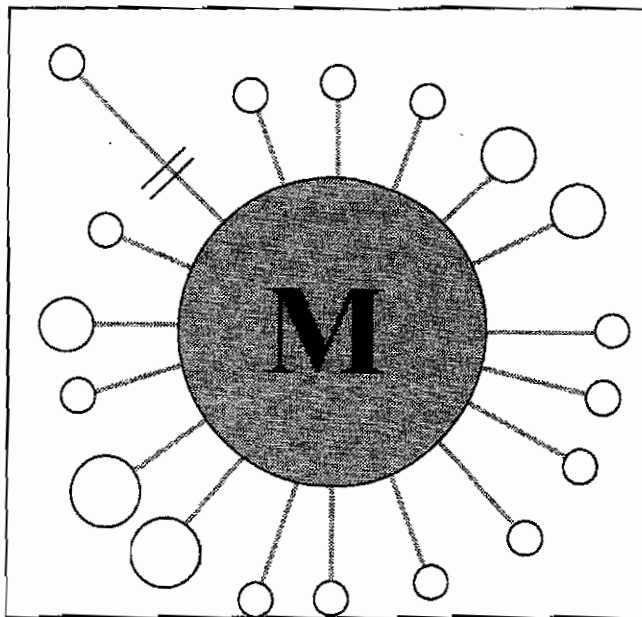


图 1-1-4 多歧分化和孤频变异是种内遗传变异的显著特征。近期经历过急剧扩张的种群的遗传谱系结构，呈现星状或辐射状网络进化关系。图中的每一个圆圈代表一个不同的遗传谱系，圆圈的面积与该谱系在种群中的频率成正比 (最小的圆圈为孤频变异)，连接圆的短线代表核苷酸突变。谱系 M 是该种群中的优势谱系 (频率最高)。注意，大多数谱系与谱系 M 仅有一个核苷酸的差异 (左上角者例外，与 M 有两个核苷酸的差异)，它们之间是多歧分化的进化关系

多歧分化主要是由种群中不同个体在进化过程中的独立变异所造成的,而孤频变异是较浅分化水平下的必然结果。通常,种群中多歧分化的水平与孤频变异的绝对频率是正相关的,即多歧分化越显著则孤频变异越多。一个 DNA 区段的种内遗传多态性水平越高,则多歧分化表现得越显著,孤频变异的绝对频率也就越高。对于快速扩张的种群,特别是经历过严重瓶颈效应而又进行了快速扩张的种群,以及受正选择影响的遗传位点,多歧分化则尤为显著,谱系间表现为星状或辐射状进化关系和网络状结构(图 1-1-4)。在这种情况下,则不同谱系间的进化关系更适于以网络状结构来展示(Bandelt *et al.*, 1995, 1999)。

根据种内谱系多样性的这些特点,我们不难看出仅仅基于频率分析的传统群体遗传学分析方法的局限性,传统的概括性统计量(summary statistics)难以全面揭示谱系数据所隐含的进化信息。例如,从谱系水平来看,各谱系间核苷酸差别的分布隐含有关于种群历史增长情况或所承受的进化压力的重要信息,需要用歧点分布(mismatch distribution)的方法进行分析;而且,遗传谱系的进化结构可反映出传统分析方法所无法揭示的许多特征,比如,一个谱系(单倍型)在网络进化关系中的相对位置和频率可反映出它出现的时间、在进化中所处的位置以及所在种群的历史演化情况。从种群的水平来看,一个种群所含有孤频变异和呈多歧分化关系的谱系的情况能反映出在进化过程中种群的增长情况或所承受的进化压力以及演化的时间尺度;谱系间的相对进化关系和谱系的地理分布格局可以帮助区分和解析历史进化因子的影响和种群的生态演化过程(Rogers and Harpending, 1992; Harpending, 1994; Templeton *et al.*, 1995; Templeton, 1998)。

第三节 分子生态学的研究方法

分子生态学是理论和实验相互结合的交叉学科领域。就实验研究本身(即撇开实验设计、数据分析和理论诠释),分子生态学研究有两个核心部分,即分子标记的建立和分子数据的获取,其中分子标记的建立具有核心基础性。人们通常所说的“分子工具”(molecular tool)实际上就是对以上两部分的概括,等于“分子标记加上遗传变异的检测方法”。

一、分子标记

分子标记包括各种蛋白质(同工酶)标记系统和 DNA 标记系统,在此仅对有关 DNA 分子标记作一简要介绍。

从基因组属性划分, DNA 分子标记有三种类型,即细胞核 DNA 标记(nuclear DNA)、线粒体 DNA 标记(mitochondrial DNA, mtDNA)和叶绿体 DNA 标记(chloroplast DNA, ctDNA);从功能和组成特点划分,分子标记包括编码区(coding region)、非编码区(non-coding region)、微卫星 DNA(microsatellite DNA)和小卫星 DNA(minisatellite DNA)等;从拷贝数量划分,有单拷贝多态性细胞核 DNA 标记(single copy nuclear polymorphic DNA, scnpDNA)、核糖体 DNA 标记(ribosomal DNA, rDNA)等。

在实际研究中,如何选择以及选择哪些 DNA 标记,则取决于具体问题的性质和研究目的,包括研究对象的类别、标本材料的形式、欲研究问题的时空跨度等等。

1. 线粒体 DNA 标记

线粒体 DNA 标记被广泛用于后生动物的各类研究中,在其众多的优点中最为重要的三点是:①母性遗传,缺乏重组;②容易设计保守性通用引物;③有效单倍性,即细胞中的众多线粒体 DNA 拷贝在遗传组成上是等同的。在哺乳动物中,线粒体 DNA 的进化速率很高,约是核 DNA 的 10 倍以上。文献中有许多关于线粒体 DNA 标记的综述,1996 年以前的大多数综述在 Zhang 和 Hewitt (1997a) 中均有引用,有兴趣的读者可以查阅。有关通用引物的文献报道,主要有 Simon 等 (1994)、Kocher 等 (1989)、Zhang 和 Hewitt (1997b) 等。

与动物中的情况相反,植物线粒体 DNA 作为分子标记的使用价值相对局限,因为其进化速率很低且基因重组频繁。

近来发现,在不少物种的核基因组中存在线粒体假基因,即在生物进化过程中由线粒体转移到细胞核基因组中的线粒体 DNA,它们与线粒体 DNA 是外源同源 (xenology) 的关系。这些假基因会对线粒体 DNA 标记的使用造成很大干扰,以至于在有些生物类群中,如很多蝗虫中,线粒体 DNA 无法被用作分子标记。研究人员务必注意检查在其研究对象中是否有线粒体假基因的干扰。建议读者进一步参考这方面的两篇综述文章:Zhang 和 Hewitt (1996)、Bensasson 等 (2001)。

2. 叶绿体 DNA 标记

对于植物而言,叶绿体 DNA 作为分子标记的价值犹如线粒体 DNA 之于后生动物。叶绿体在很多植物中都是母性遗传(也有父性遗传或双亲遗传的例子)。但由于叶绿体 DNA 的进化速率较小,作为分子标记,它更适合于种以上水平的系统发生研究。它在种内遗传变异研究方面的作用和有效性,则因植物类群而异。关于叶绿体 DNA 保守性引物,可参阅 Taberlet 等 (1991) 和 Demesure 等 (1995) 的文章。

3. 细胞核 DNA 标记——单拷贝多态性细胞核 DNA 标记

单拷贝多态性细胞核 DNA (scnpDNA) 标记指细胞核基因组中拷贝数为 1 的 DNA 序列。每一个多态性 scnpDNA 标记实际上包含一到多个相互连锁的单核苷酸多态性 (single nucleotide polymorphism, SNP) 基点。scnpDNA 标记对于运用溯祖理论进行谱系进化分析非常重要,无论在其可用性、变异性、实验和理论检测技术和方法,还是在数据分析方面都已具备成为本学科领域新的增长点和突破点的条件,可望推进分子生态学和种群进化研究迈向一个新的时代。Zhang 和 Hewitt (2003) 对此有非常详尽的阐述。过去影响 scnpDNA 标记使用的四个主要实验技术和数据分析障碍现在都可以克服或逾越,包括:

(1) 可用的 scnpDNA 标记:模式生物获益于基因组测序工程而有大量可以直接使用的 scnpDNA 标记;对于非模式生物也存在一些成熟的建立 scnpDNA 标记的方法,其中一个捷径就是使用单拷贝微卫星位点的侧翼区域。

(2) 单倍型相的分离: 由于二倍体或多倍体生物中杂合性的存在, 杂合体中包含两个不同的等位基因 (单倍型), 现在已经建立了一些适合于大规模分析的、行之有效的、单倍型相的统计和实验分离方法。

(3) 细胞核 DNA 多态性数据的分析: 由于多歧分化和孤频变异是种内遗传变异的显著特征, scnpDNA 的变异水平通常较低, 以及核基因组中基因重组的广泛存在, 增加了 scnpDNA 数据分析的难度, 对此目前从理论上已有较系统的认识, 在分析技术上也有显著进展。

(4) scnpDNA 标记对种内变异分析的可用性: 由于总体上细胞核 DNA 的突变速率较线粒体 DNA 为低, 例如在哺乳动物中前者只是后者的十分之一或更低, 所以此前学术界对于 scnpDNA 标记是否能揭示出足够的种内变异有一定疑虑。实践证明, 种群中含有足够多的 scnpDNA 变异, 可以用于分子生态学研究。例如, 我们实验室对飞蝗一段 251bp 细胞核 DNA 分析发现, 种群中存在 71 个多态基点, 由 63 个单倍型构成了 141 个不同基因型, 种群的核苷酸多样性是 0.0243——这对于细胞核 DNA 标记来说是相当高的 (Zhang and Hewitt, 2003)。

4. 细胞核 DNA 标记——核糖体 DNA

核糖体 DNA (rDNA) 是重复序列, 在基因组的拷贝数可达数千, 但却是目前最为广泛使用的细胞核 DNA 分子标记, 这主要是因为它服从协同进化 (concerted evolution) 规律, 因而表现出有效单倍性。由于 rDNA 簇 (单元) 是由一系列编码区和非编码区嵌合而成, 所以可以设计出进化上非常保守的通用引物, 用以在不同物种中扩增变异性较高的非编码区 [比如被转录的内间隔区 (internal transcribed spacer, ITS)]。下列文献给出了许多 rDNA 保守引物: Hillis 和 Dixon (1991)、Schlotterer (1996) 和 Ji 等 (2003a)。

使用 rDNA 标记时, 有两点值得注意: ①就目前的情况来看, rDNA 标记对于种内变异研究的价值因物种而异, 甚至因种群而异, 其变异性相对线粒体 DNA 则低很多; ②并不是在每个物种中 rDNA 都遵循协同进化规律, 研究人员必须首先对此进行检验以确定其适用性。

5. 细胞核 DNA 标记——微卫星 DNA 和小卫星 DNA 标记

微卫星 DNA 和小卫星 DNA 有时统称为简单重复 DNA (simple sequence repeat, SSR)。前者通常指重复单位的核苷酸长度为 1~6 个的 DNA 序列, 如 GTGTGTGTGTGTGTGTGTGTGT, 表示为 (GT)₁₀; 后者重复单位的长度通常是十几到几十个核苷酸。二者的共同特征是变异性很高。微卫星 DNA 主要表现为重复单位拷贝数的变化, 主要突变机制是复制滑动错位 (replication slippage)。目前的发展趋势是, 在分子生态学研究中, 小卫星 DNA 标记由于其长度较大, 基因型自动分析不方便, 使用的频率 (相对比例) 在降低。微卫星 DNA 标记则是越来越广泛地被应用。这里主要介绍微卫星 DNA 标记的有关情况。图 1-1-5 显示的是东亚飞蝗的一个 (GT)_n 微卫星位点。

微卫星标记的使用大体上有三种不同方法。最原始的方法是从多位点小卫星指纹分析衍生出来的多位点微卫星指纹分析法, 即用微卫星 DNA 为探针与经过内切酶消化的

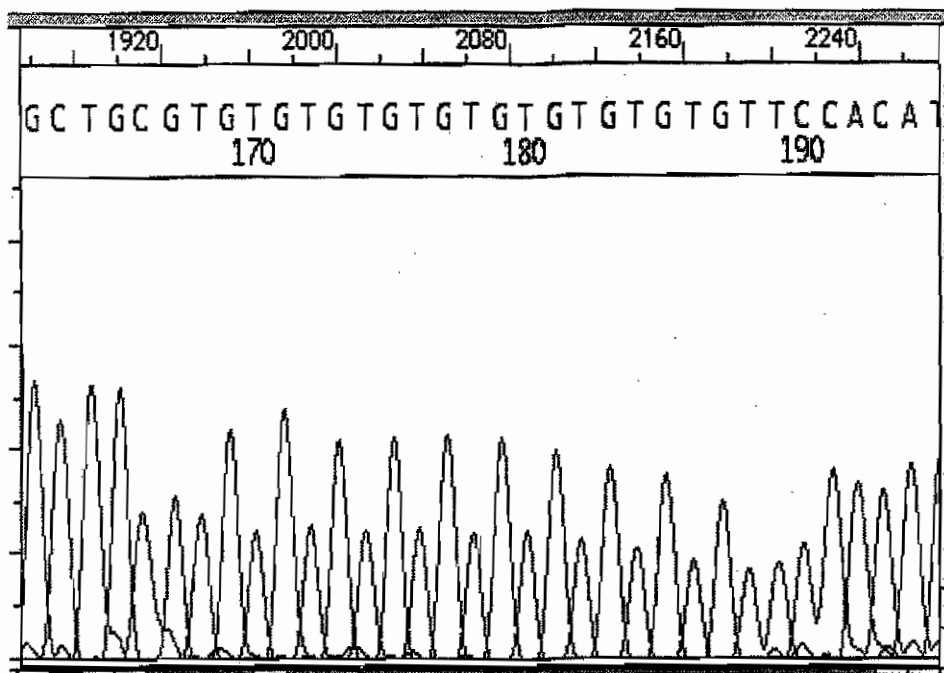


图 1-1-5 东亚飞蝗的一个 (GT)₁₁ 微卫星位点序列示意图

总量 DNA 进行杂交, 从而检测出遗传变异。这一方法虽然灵敏度较高, 但存在与小卫星指纹分析相同的问题, 即难以对数据进行深度进化遗传学分析, 对 DNA 质量和数量要求很高, 难以对每个条带进行量化处理等。PCR 技术普及以后, 演化出另一种多位点微卫星分析法, 即以微卫星序列为 PCR 引物, 对总量 DNA 进行扩增, 根据所扩增出的条带的差异来检测遗传多态性。该方法虽然简单易行, 容易操作, 能找出很多带型差异, 但却隐含许多不确定性, 主要表现在 PCR 的重复性、带型有效性、条带同源性等方面, 例如同一个模板 DNA 样品, 在不同的 PCR 条件下可以扩增出不同的带型。一般说来, 该方法不宜用于种群的遗传进化研究。第三种方法, 位点特异性微卫星分析, 是经过反复验证、目前被公认的方法, 又称为单位点分析法。它是通过对每一个微卫星位点分别进行特异性变异检测而实现的。原理如下: 许多微卫星 DNA 两侧的区域 (称为侧翼区域, flanking region) 是非重复的、单拷贝 DNA, 因而可以在侧翼区域设计特异性引物, 通过 PCR 而扩增出该特定位点上的微卫星 DNA, 然后对其进行基因型鉴定。微卫星单位点分析法特异性强, 结果稳定, 能对坐落于该位点上的所有等位基因进行精确基因型鉴定。通过荧光标记法和错落搭配法, 研究人员可以同时若干个不同的微卫星位点进行特异性分析 (图 1-1-6)。

微卫星 DNA 标记也存在许多问题, 首先是其分子突变机制尚不完全清楚, 其突变速率受等位基因长度的影响, 并且可能有方向性; 其次, 由重复突变 (recurrent mutation) 或平行突变 (parallel mutation) 引起的非同源相似 (homoplasy) 在种间和种内都存在; 再者, 等位基因的祖先状态 (ancestral state) 难以确定, 因此会影响对有关数据进行谱系分析; 再者, PCR 引物在不同物种间的保守性较差, 研究人员通常需要为自己的研究系统开发一套特异微卫星标记, 这无论在技术难度上还是在经费上要求都很高。此外, 在有些物种或类群中, 例如许多鳞翅目昆虫, 合适的微卫星 DNA 标记较

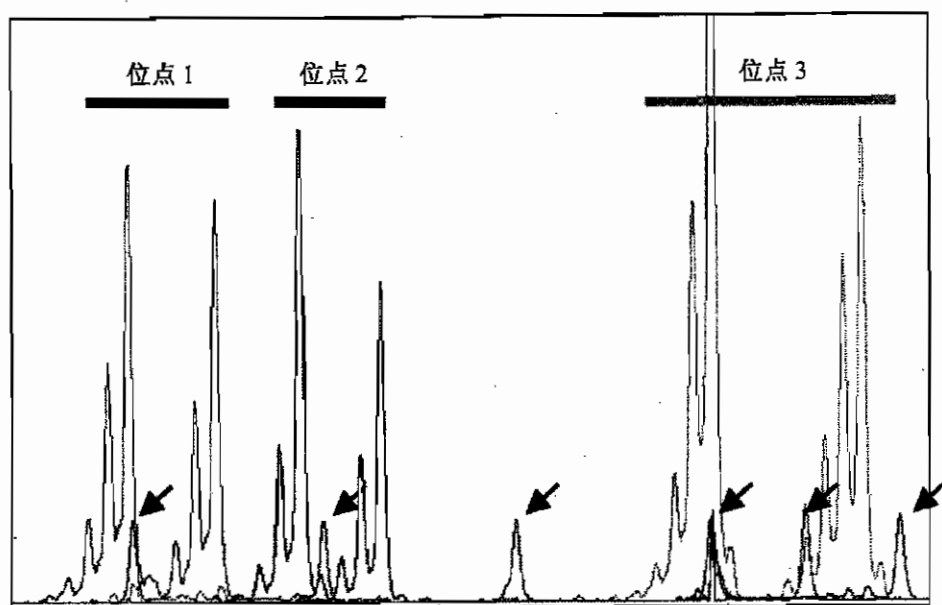


图 1-1-6 用 GeneScan 的方法同时对东亚飞蝗三个不同微卫星位点进行基因型检测
用于测试的个体在这三个微卫星位点上都是杂合体。基部的矮峰（箭头所指者）为分子质量内
标。这三个微卫星位点采用的是不同的荧光标记

难建立 (Ji *et al.*, 2003b; 吉亚杰和张德兴, 2004; Zhang, 2004)。

6. 细胞核 DNA 标记——AFLP 标记

扩增片段长度多态性 (amplified fragment length polymorphism, AFLP) 是 1993 年由 Zabeau Marc 和 Vos Pieter 创立的一种技术通用型多态性标记。其原理是把 RFLP 与 PCR 相结合, 对多态性片段进行选择检测。在这种方法中, 首先用限制性内切核酸酶把基因组 DNA 切割成大小不等的片段, 再把双链人工接头接到基因组 DNA 的酶切片段两端。然后以此为模板, 用标记的具有不同选择强度的特异引物 (引物选择强度由附接于 3' 端的与接头不互补 1~3 个选择碱基控制) 进行 PCR 扩增。遗传多态性通过扩增出来的限制性片段的差别体现出来。AFLP 反应所需 DNA 量少, 得到的条带丰富, 分辨率高, 快速高效。它与下面将要介绍的 RAPD 标记相比, 具有多态性丰富、重复性好、条带稳定等优点。但该技术对样品 DNA 质量要求较高, 多态片段中有的显性表达, 有的则共显性表达, 并且有些多态片段是不相互独立的。

7. 细胞核 DNA 标记——RAPD 标记

随机扩增 DNA 多态性 (random amplified polymorphic DNA, RAPD) 是于 20 世纪 90 年代初在两个实验室分别独立建立的。其原理和实验程序均很简单, 因为基因组中随机分布着很多寡核苷酸序列, 所以用 10bp 左右的寡核苷酸随机单引物对基因组 DNA 进行 PCR 扩增, 便可扩增出有个体特异性的产物, 然后通过电泳分离可检测扩增产物的多态性。RAPD 的优越性在于: ①技术通用性, 引物无种属特异性, 一套引物可用于不同物种的研究; ②技术“简单、高效、快速”; ③所需样品量少。然而, RAPD 技术却同时存在一些重要的缺陷, 如: 重复性较差, 对实验条件极其敏感; 有

假阳性或假阴性结果；为显性标记，不能提供完整的遗传信息等等。所以 RAPD 分子标记，看似“简单、高效、快速”，实际并非如此，在进行实验时，应采用标准的反应条件，建立系统性实验对照，慎重对待实验结果。

总体看来，以上各种分子标记在分子生态学研究中的应用呈现出如下变化趋势：目前占优势的标记是微卫星 DNA 标记和细胞器（线粒体和叶绿体）DNA 标记，这种趋势短期内不会改变；SNP 和 scnpDNA 标记正在被越来越广泛地应用，必将成为分子生态学研究的核心分子标记；AFLP 在一定领域内有较好应用，但仍占据次要的地位；除了一些特定研究领域外，RAPD 标记正逐渐退出分子生态学的核心研究领域。

二、遗传变异的检测方法

电泳技术是最基本的实验技术，大多数遗传变异检测方法都应用到它，而上述 AFLP 标记和 RAPD 标记更是将它融为一体。从所揭示的遗传变异信息的类型和层次上看，遗传变异的检测方法不外乎两类：序列分析（sequence analysis）和片段分析（fragment analysis）。序列分析指直接对 DNA 标记的核苷酸组成进行分析；片段分析指通过检测多态片段的长度或构象来区分不同的基因型，所以，片段分析有时也被称作基因型分析。

1. 序列分析

遗传信息的分子表现形式为 DNA 序列（或 RNA 序列）。因此，通过 DNA 序列分析，可以获得生物遗传变异最确切、最深层次的信息 [图 1-1-5 是对一个 (GT)_n 微卫星位点的序列分析]。随着生物技术的长足进步，DNA 序列分析已经全面自动化，并成为分子水平遗传研究的常规手段。DNA 序列数据与其他遗传变异数据相比还有如下优点：所包含的进化信息最全面，含有确切的共祖信息，序列差异程度的衡量标准比较统一，能够揭示从个体间到物种间不同层次的遗传差别等。目前，DNA 序列分析的实验成本仍相对较高，因而研究位点的大小和数量往往很受限制。但对于研究许多进化问题，序列分析往往是不可替代的手段。

2. 片段分析

目前，已有许多成熟的片段分析方法，它们所依据的原理或者是长度差异本身，或者是分子构象的差异。下面主要介绍四种不同的方法，其中，长度分析，如 GeneScan，属于对等位基因长度的直接分析，SSCP、DGGE 和 TGGE 属于根据构象差异而区分不同等位基因，DHPLC 则可从上述两个方面进行分析（因具体情况而异）。

（1）长度分析：大多数学者把这种分析方法直接称为“片段分析”，以微卫星 DNA 的 GeneScan 为例。GeneScan[®] 本是 Applied Biosystems 公司发展的一套用于微卫星 DNA 标记片段分析的软件，由于它被广泛使用，GeneScan 已代表一种分析方法。它利用荧光素对微卫星引物进行标记，然后用荧光标记的引物对目的片段 PCR 扩增，随后在 DNA 序列自动分析仪上测出目的片段的长度。由于它可以同时测定出杂合体中两个等位基因的大小，所以在分析的同时也可检测出个体的基因型。目前，DNA 序列自动

分析仪基本上都可以进行基因型分析,所以这类仪器的正式名称已逐渐变为“自动化遗传分析系统”。图 1-1-6 是通过 GeneScan 同时对东亚飞蝗三个不同微卫星位点进行基因型检测的例子。

(2) SSCP 分析:单链构象多态性 (single-strand conformation polymorphism, SSCP) 检测方法是 Orita 等于 1989 年建立的。其原理在于,在非变性条件下, DNA 单链分子可以形成具有序列特异性的构象;在非变性电泳条件下,不同构象的单链分子表现出不同的迁移率,因而根据迁移率的差别即可检测出遗传变异。SSCP 的基本特点是,目的片段上所发生的突变,甚至单个碱基的替换、缺失或插入等微小变化,也将导致其单链构象改变,并在电泳迁移率上表现出来。SSCP 检测有一定的适用范围,200~400bp 最理想,产物太大时,检测灵敏度会变得很低。SSCP 分析需要对实验条件作一定优化,因为实验条件因 DNA 片段而异,且受电泳温度、缓冲液和电泳凝胶的配方等因素影响较大。在最优条件下,SSCP 可以达到很高的分辨率。

(3) DGGE 和 TGGE:变性梯度和温度梯度凝胶电泳 (denaturing gradient gel electrophoresis, DGGE; temperature gradient gel electrophoresis, TGGE) 方法由于其具有较高的分辨率而越来越受重视,并开始用于点突变检测和等位基因的分离。这一方法的理论原理是,一个 DNA 分子的不同区段往往具有不同的动力学特性,表现为解链温度 (T_m) 的不同,因而其解链温度曲线具有序列专一性。如果 DNA 分子发生了突变,突变点所在区段的 T_m 就会被改变,从而改变解链温度曲线,因此可把突变型分子与野生型分子区别开来。

这一方法的技术原理如下: DGGE 所用的聚丙烯酰胺凝胶的不同部位的变性强度不同,一般从左到右或从上到下变性强度连续性递增。当双链 DNA 分子在凝胶中沿着由低到高的变性梯度迁移时, T_m 最低的区段将在适当的位置优先解为单链,形成一个局部变性的枝状或叉状 DNA 分子,从而降低 DNA 分子在电泳凝胶中的迁移速率。如果在这一变性区段发生了突变,其 T_m 也会改变,在变性梯度电泳凝胶上表现出不同的迁移特征,因而可以被检测出来。

DGGE 所用的电泳凝胶的变性梯度是通过化学变性剂 (如尿素、甲酰胺) 的浓度的梯度性变化来实现的,而 TGGE 的电泳凝胶具有均一的变性强度,其变性梯度是通过空间上或时间上的连续性温度梯度来实现的。目前应用较广泛的是 DGGE。它可以检测出单个或多个碱基的差别,在优化的条件下,分辨率可达到 95% 以上。

(4) DHPLC:变性高效液相色谱 (denaturing high performance liquid chromatography, DHPLC) 是 2000 年以来才逐渐发展成熟的核酸分析技术。其原理是异源双链分析 (heteroduplex analysis)。异源双链 (heteroduplex) 指由不完全互补的两条 DNA 单链形成的双链 DNA,同源双链 (homoduplex) 指由完全互补的两条 DNA 单链形成的双链 DNA。简略说来,异源双链与同源双链对介质有不同的亲和力,因此,它们从介质中被洗脱下来的条件有差别。利用这一特性,可以把二者区分开来。所以,把一个未知样品 (例如一未知 mtDNA 片段) 同一个已知样品 (即已被定性分析的 mtDNA 片段) 混合后,让它们形成足够比例的异源双链,通过分析是否有异源双链吸收峰便可以知道这两个样品是否等同。对于核 DNA,纯合体只有同源双链吸收峰,而杂合体则除有同源双链吸收峰外,还有异源双链吸收峰。

DHPLC 可以快速高效地对核酸片段进行半自动化分析，并可以对样品进行自动化回收。该技术允许在完全变性、部分变性或非变性条件下分析核苷酸片段，已被广泛用于核酸片段分离、突变检测、基因型分析、PCR 引物纯度分析和纯化、SNP 分析等研究，单核苷酸突变检测准确率稳定在 96% 以上。在最佳条件下，该技术甚至可以对有些杂合体的两个等位基因进行分离。其缺点是，要求有特定分析仪器，对每一类样品，要通过反复实验摸索出最优条件，对于较大 DNA 片段的分析不够灵敏。图 1-1-7 显示的是 DHPLC 对棉铃虫的一个微卫星 DNA 位点的突变检测情况。

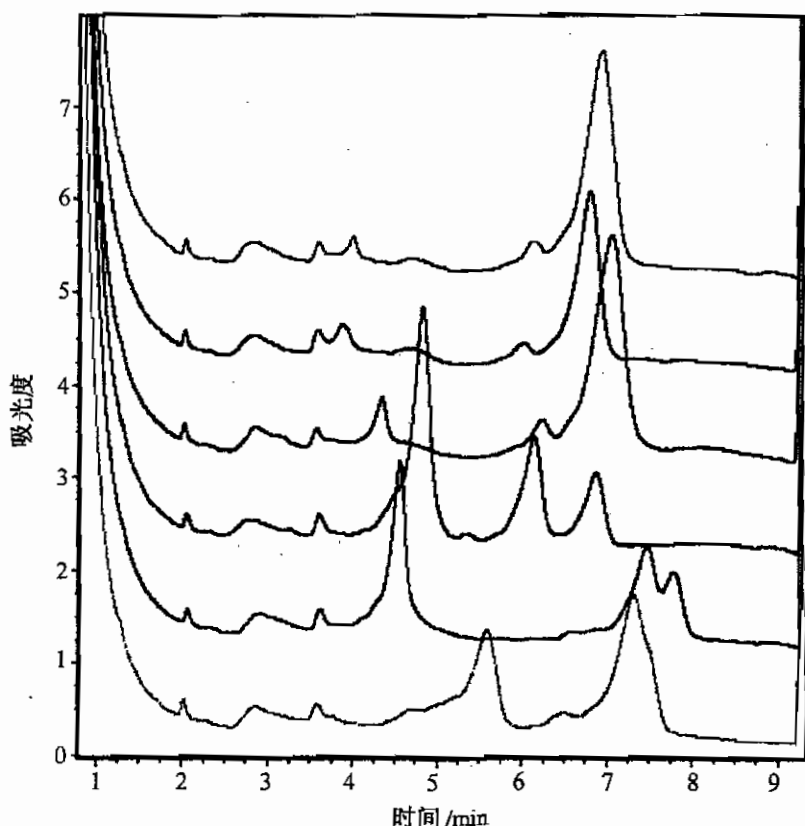


图 1-1-7 一个棉铃虫微卫星 DNA 位点的 DHPLC 突变检测
上面三个个体为纯合体，下面三个分别为等位基因长度相差 5、2、1 个重复单位的杂合体

三、分子生态学研究的特点

综上所述可以看出，分子生态学研究是理论跨度大、实验技术性很强的交叉性复合学科，其研究特点可概括为如下几条：

(1) 起点高，可以获得最深层次（DNA 序列组成）上的遗传变异信息，因而可以达到很高的分辨率。

(2) 揭示能力强：分子生态学研究方法在解决各种进化问题、历史演化问题方面有其他方法难以比拟的揭示能力。

(3) 技术集成度：分子生态学研究的有效开展，要求有较全面的分子生物学及分子进化等方面的技术方法，属于技术密集型研究。

(4) 学科整合性强: 为了发现相应的主要影响因子, 往往需要从多学科、不同角度去研究同一问题, 因此要求研究人员具有多学科的知识积累。

(5) 持续型研究: 由于本学科研究的问题是非常复杂的宏观问题, 正确解决任何一个问题都需要用多套相互独立的分子标记系统进行持续性研究, 因此研究的周期一般较长, 对前期工作积累依赖性大, 研究者决不可草率地依据少量分子数据对问题下定性结论。

(6) 风险型研究: 投入与产出未必成比例, 研究结果经常与预期的不相符, 研究进度经常比预期的慢, 因此, 不可急于求成。

(7) 两面性研究: 好的数据和客观的结论能有效推动有关研究领域的健康发展, 坏数据和片面的结论则会产生很大误导。后一点尤其应引起我们足够的重视。

第四节 分子生态学研究的发展趋势

一、我国分子生态学研究现状

自 *Molecular Ecology* 杂志 1992 年创刊以来, 分子生态学得到了长足的发展, 已经成为宏观生物学领域具有举足轻重地位的交叉学科。这可以从该期刊的影响因子得到体现: *Molecular Ecology* 2006 年的 SCI 影响因子为 4.825, 在生态学领域的 114 种期刊中排名第 6 位, 在进化生物学领域的 35 种期刊中排名第 5 位。近年来, 尽管分子生态学研究在我国也得到了较大的发展, 但我国的总体研究水平与发达国家相比还有很大差距, 主要表现为: 在研究单元的地域分布上存在很大的不平衡性, 尚未形成一个具有较强科研竞争能力的研究网络; 在科研课题的设置上, 总体上仍呈现为个案研究和机动性研究缺乏系统性的、具有区域特色的研究体系; 在理论、技术或分析方法的原创性贡献上, 我国的产出十分匮乏, 目前的研究主要集中于区域性的应用性研究, 在对该学科领域发展趋势的影响上尚缺乏发言权。

二、分子生态学的未来发展趋势

分子生态学研究方法将继续向所有与生态、种群和进化有关的学科领域全面渗透, 并不断取得开拓性成果, 这已成为学科发展的时代大趋势。分子生态学在未来的发展仍将依赖于理论和分析方法上的突破和分子技术上的进步, 在若干重要领域的进展会继续显示出多学科综合交叉的优势。

从学科领域的全局看, 分子生态学研究的热点将体现在复杂宏观问题的系统研究方面, 如生物的生态适应, 全基因组水平的遗传变异分析, 较大时空范围内生物生态地理演化的规律和影响因子的研究, 遗传多样性与群落及生态系统稳定性, 家系结构、行为特征对种群分化和物种形成的影响, 物种间的相互作用, 生物保护以及种群控制等, 这些问题一般都需在较好的研究系统下, 经过较长时间的研究积累, 才能取得突破。但分子生态学涵盖庞大的研究空间, 任何其他方向, 只要能取得实质性进展, 均可成为新的热点领域。我国勿须盲目追踪国际潮流, 而应结合自己的优势, 开辟适合我国实际情

况、满足国家发展需求的研究领域,例如有害生物的种群结构和动态方面,生物生态地理演化方面等。

从技术手段方面看,近期的热点仍将是发展高效、快速的遗传变异检测、筛选方法,以便能达到从全基因组的水平进行遗传变异分析。这方面将主要在现有的模式生物中进行,特别是人、果蝇、拟南芥、水稻等,因为这些生物的基因组测序计划已基本完成,有大量可供选择的各种 DNA 标记,如微卫星 DNA 标记、SNP 标记等。

从分子标记方面看,微卫星 DNA 和线粒体 DNA 仍将是首选的分子标记,并且其分子进化机制方面的研究会继续加深;scnpDNA 标记将成为新的热点标记并被广泛应用,它有望使分子生态学研究迈上一个新的台阶;规模化 SNP 标记和全基因组分析将成为重大研究计划认真考虑的标记系统。此外,转录和表达水平的多态性标记将逐步成为进行生态适应机制研究的分子标记,从而推动分子生态学向更高层次拓展。

从进化分析角度看,遗传谱系分析将变成主流,这将主要依赖于 DNA 序列数据和其他能较容易推测出祖先状态的单倍型数据。以溯祖理论为核心的、针对具体科学问题的各种计算机模拟和参数估计方法可望会得到空前发展,而且溯祖理论指导下的计算机模拟和参数估计将成为分子生态学研究中的常规的数据分析模式。

(张德兴)

思考题

1. 谈谈随机遗传漂变对于生物进化的意义。
2. 请推导一下溯祖概率方程

$$P = \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^{t-1}$$

3. 在存在自然选择或基因流的情况下,溯祖时间会有怎样的变化?
4. 在各种进化力量中(自然选择、随机遗传漂变、基因流、近交、突变等),哪些会增加种群的遗传多样性,哪些会降低种群的遗传多样性?为什么?
5. 选择 DNA 分子标记时,应考虑哪些方面?
6. 试谈分子水平的研究与宏观研究的关系,并举例加以说明。

推荐读物

1. Avise JC. 2004. Molecular Markers, Natural History and Evolution. 2nd edition. Sunderland: Sinauer Associates
2. Avise JC. 2000. Phylogeography: the History and Formation of Species. Cambridge: Harvard University Press
3. Hartl DL, Clark AG. 1997. Principles of Population Genetics. 3rd edition. Sunderland, MA: Sinauer Associates
4. Hein J, Schierup MH, Wiuf C. 2005. Gene genealogies, variation and evolution: A primer in coalescent theory. Oxford: Oxford University press
5. Kimura M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press
6. Li WH. 1997. Molecular Evolution. Sunderland, MA: Sinauer Associates
7. Zhang DX, Hewitt GM. 1996. Nuclear integrations: Challenges for mitochondrial DNA markers. Trends in Ecology and Evolution, 11: 247~251
8. Zhang DX, Hewitt GM. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and

参考文献

- 吉亚杰, 张德兴. 2004. 鳞翅目昆虫基因组中微卫星 DNA 的特征及其对分离的影响. *动物学报*, 50 (3): 608~614
- Avice JC. 2000. *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge
- Avice JC, Gyllen-Davidson C, Laerm J, et al. 1979. Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences (USA)*, 76: 6694~6698
- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37~48
- Bandelt HJ, Forster P, Sykes BC, et al. 1995. Mitochondrial portraits of human populations using median networks. *Genetics*, 141: 743~753
- Bensasson D, Zhang DX, Hartl DL, et al. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution*, 16: 314~321
- Brown WM, Vinograd J. 1974. Restriction Endonuclease Cleavage Maps of Animal Mitochondrial DNAs. *Proceedings of the National Academy of Sciences (USA)*, 71: 4617~4621
- Demesure B, Sodzi N, Petit RJ. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, 4: 129~131
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52: 696~704
- Harpending RC. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, 66: 591~600
- Hartl DL, Clark AG. 1997. *Principles of Population Genetics*. 3rd edition. Sinauer Associates, Sunderland, MA
- Hein J, Schierup MH, Wiuf C. 2005. *Gene genealogies, variation and evolution: A primer in coalescent theory*. Oxford University press, Oxford
- Hillis DM, Dixon MT. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Reviews of Biology*, 66: 411~453
- Hillis DM, Moritz C, Mable BK. 1996. *Molecular Systematics*. 2nd edition. Sinauer Associates, Sunderland
- Hudson RR. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, 7: 1~44
- Hunter RL, Markert CL. 1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science*, 125: 1294~1295
- Ji YJ, Zhang DX, He LJ. 2003a. Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer (ITS) regions in insects and other invertebrates. *Molecular Ecology Notes*, 3: 581~585
- Ji YJ, Zhang DX, Hewitt GM, et al. 2003b. Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. *Molecular Ecology Notes*, 3: 102~104
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature*, 217: 624~626
- Kimura M. 1983. *The Neutral Theory of Molecular Evolution*, Cambridge University Press.
- Kocher T, Thomas W, Meyer A, et al. 1989. Dynamics of mitochondrial DNA sequence evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences (USA)*, 86: 6196~6200
- Lewontin RC, Hubby JL. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54: 595~609
- Li WH. 1997. *Molecular Evolution*. Sinauer Associates, Sunderland, MA
- Linn S, Arber W. 1968. Host specificity of DNA produced by *Escherichia coli*, λ *in vitro* restriction of phage *fd* replicative form. *PNAS*, 59: 1300~1306

- Lynch M, Crease TJ. 1990. The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution*, 7: 377~394
- Maniatis T, Hardison RC, Lacy E, et al. 1978. The isolation of structural genes from libraries of eucaryotic DNA. *Cell*, 15: 687~701
- Maxam AM, Gilbert W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences (USA)*, 74: 560~564
- Meselson M, Yuan R. 1968. DNA Restriction Enzyme from *E. coli*. *Nature*, 217: 1110~1114
- Rogers AR, Harpending RC. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9: 552~569
- Saiki RK, Gelfand DH, Stoffel S, et al. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239: 487~491
- Saiki RK, Scharf S, Faloona F, et al. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230: 1350~1354
- Sanger FS, Nicklen AR, Coulson. 1977. DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences*, 74: 5463~5467
- Schlotterer C. 1996. Ribosomal DNA probes and primers. In "Molecular Tools for Screening Biodiversity: Plants and Animals" (eds. Karp A, Isaac PG & Ingram DS), Chapman & Hall, 267~276
- Shoemaker J, Painter I, Weir BS. 1998. A Bayesian characterization of Hardy-Weinberg disequilibrium. *Genetics*, 149: 2079~2088
- Simon C, Frati F, Beckenbach A, et al. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain-reaction primers. *Annals of the Entomological Society of America*, 87: 651~701
- Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139: 457~462
- Smithies O. 1955. Zone electrophoresis in starch gels: group variations in the serum proteins of normal individuals. *Biochemical Journal*, 61: 629~641
- Southern EM. 1975. Transfer of DNA fragments to Millipore filters after restriction endonuclease cleavage followed by agarose gel electrophoresis. *Journal of Molecular Biology*, 98: 503~517
- Taberlet P, Gielly L, Pautou G, et al. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, 17: 1105~1109
- Templeton AR. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, 7: 381~397
- Templeton AR, Routman E, Phillips CA. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, 140: 767~782
- Zhang DX, Hewitt GM. 1996. Nuclear integrations: Challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution*, 11: 247~251
- Zhang DX, Hewitt GM. 1997a. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*, 25: 99~120
- Zhang DX, Hewitt GM. 1997b. Assessment of the universality and utility of a set of conserved mitochondrial COI primers in insects. *Insect Molecular Biology*, 6: 143~150
- Zhang DX & Hewitt GM. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, 12: 563~584
- Zhang DX. 2004. Lepidopteran microsatellite DNA: redundant but promising. *Trends in Ecology and Evolution*, 19: 507~509
- Zhang DX, Szymura JM, Hewitt GM. 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution*, 40: 382~391