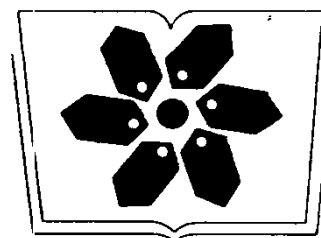


159

82-2  
J57



中国科学院科学出版基金资助出版

# 医学分子病毒学

金奇 主编



A0988341

科学出版社

2001

## 内 容 简 介

医学分子病毒学是生命科学中发展最快的领域之一,近 10 年来,由于新的分子生物学技术和方法的建立、改进和完善,极大地推动了医学分子病毒学的研究,在病毒基因组结构、致病机制和分子诊断等方面有了诸多重大突破,并由此产生了许多新的理论和新的概念。

本书着重介绍我国常见的医学病毒以及近年来国内外相关领域中的研究热点。全书共分 37 章,分别就病毒的一般特征、基因组结构、基因组编码产物及其功能、繁殖与复制、致病机制、流行特点和诊断与治疗等方面进行了系统介绍,反映了 90 年代后期分子病毒学研究的水平,使读者能够从分子水平了解本领域的最新研究动态。本书可作为从事分子病毒学和分子生物学研究领域的科研工作者、教师、研究生、大学生以及临床和检验工作者的参考书。

### 图书在版编目(CIP)数据

医学分子病毒学/金奇主编. -北京:科学出版社,2001. 2

ISBN 7-03-008132-3

I. 医… I. 金… II. 分子生物学:人体病毒学 IV. R373

中国版本图书馆 CIP 数据核字(1999)第 71317 号

**科学出版社 出版**

北京东黄城根北街 16 号  
邮政编码:100717

**新蕾印刷厂 印刷**

科学出版社发行 各地新华书店经销

\*

2001 年 2 月第 一 版 开本: 787×1092 1/16

2001 年 2 月第一次印刷 印张: 60

印数: 1—4 000 字数: 1 363 000

**定价: 95.00 元**

(如有印装质量问题,我社负责调换(新欣))

## 前 言

医学分子病毒学,是应用现代分子生物学的理论、技术和方法来研究与人类健康相关的病毒以及病毒与人类相互作用关系的一门学科,它一直是生物医学乃至整个自然科学中最为活跃的领域之一。近十年来,由于新的分子生物学技术和方法的不断建立、改进及完善,极大地推动了医学分子病毒学的发展。到目前为止,国际上已经完成了 600 多株病毒完整基因组的核苷酸序列分析,基本上阐明了所有重要医学病毒代表株的基因组结构。当今医学分子病毒学研究的重点已经从结构基因组学进入到功能基因组学,即后基因组研究时期,研究的焦点主要集中在病毒基因编码产物的功能、病毒蛋白与宿主相互作用的机制、病毒基因变异的机制以及新型生物技术疫苗等领域。在此基础上,对于病毒的生物学特性、基因结构与功能的关系、致病机制、流行特征、诊断方法和防治策略等方面的研究有了诸多重大突破,并由此产生了许多新的理论和新的概念。

病毒性传染病依然是严重危害我国人民健康的主要病种之一。由于病毒基因在自然选择和人群免疫等压力下不断发生变异以及社会人口老龄化和城市化进程加快等原因,某些病毒性传染病的流行不仅未能得到有效控制,反而更加猖獗。甲、乙、丙型肝炎,艾滋病、流行性感风和多种肠道病毒病的流行情况十分严重。此外,许多研究结果表明病毒与恶性肿瘤和多种慢性疾病的发生密切相关。医学分子病毒学最为重要和直接的意义在于从分子水平上阐明病毒基因和产物结构与功能之间的关系,揭示病毒的致病机理和本质,为最终控制病毒性疾病的流行提供重要的理论基础和科学依据。

病毒是目前所知结构最为简单的生命单位,并由此成为分子生物学的理想对象。利用分子生物学方法对病毒进行研究,其结果不仅促进了病毒学的研究,反过来对分子生物学的发展也起了巨大的推动作用。例如,RNA 肿瘤病毒反转酶的发现,就是对阐明 DNA 复制和转录、mRNA 的翻译和表达及蛋白产物的结构功能三者关系的中心法则的补充。由此可以看出:分子病毒学的发展集中体现了现代自然科学研究中不同学科间相互交叉、相互渗透、相互促进的趋势和特点。

本书名为《医学分子病毒学》,主要涉及目前严重危害我国人民健康的主要病毒。全书共分 37 个篇章,较为详细地介绍了这些病毒的一般形态和结构特征、基因组结构、生物学特点、流行病学特征及检测与防治手段等内容。书中内容包括了一些新近病毒学研究领域中的热点,如病毒编码的细胞因子及其受体的模拟分子、病毒载体、病毒基因工程以及生物信息学在分子病毒学研究中的应用等。在围绕全书的基础上,各章节自成经纬,从分子水平上阐述相关病毒的最新研究进展和研究成果。

本书的作者来自国内外从事医学分子病毒学研究较为领先的机构,其中病毒基因工程国家重点实验室及其依托单位中国预防医学科学院病毒学研究所是国内开展医学病毒学研究的重要前沿基地,曾先后研制出具有我国自主知识产权的重要抗病毒药物干扰素、建立了鼻咽癌早期诊断方法、首次发现 B 组轮状病毒、完成我国痘苗病毒天坛株基因组工程、研制乙型肝炎病毒和出血热病毒疫苗等多项具有创新性成果的科研项目。本书的作

者均是在我国医学分子病毒学研究领域中脱颖而出的优秀青年科学家,他(她)们目前主持承担着相关领域中国的重大科研项目。他们丰富的科学知识和研究经验,勇敢的探索和创新精神,严谨求实的科学态度,都是本书编写质量的有力保证。

本书的成文还得到了恩师侯云德院士的指导和鼓励,在审阅和修改书稿之余,恩师还为本书欣然命序。此外,曾毅院士、洪涛院士和闻玉梅院士等多位前辈老师对本书给予的关心和帮助,亦使之大受裨益。此外,在本书编写过程中,广大同行给予了我无私和宝贵的帮助与支持,王平、董杰、朱俊萍、王瑾以及我所在课题组的同事们在承担繁重的科研任务的同时,对书稿进行了多次细致的校对,并制作了非常精美的图表。所有这些鼓励和支持使我最终完成了本书的编写。在此对他(她)们表示衷心的感谢。

21世纪是生物科学的世纪,医学分子病毒学的发展将一日千里,日新月异。我们的水平有限,虽经努力,书中错漏之处还难以避免,欢迎各位同行不吝赐教,给予批评指正。

病毒基因工程国家重点实验室



2000年仲夏于北京

# 第1章 病毒的进化

---

人类许多疾病,如肝炎、肺炎、脑炎、肾炎、脊髓灰质炎、流行性感、狂犬病、艾滋病等等,都是由病毒引起的,因而有关病毒由来、存在以及发展演变的过程、特征和规律,也就是病毒的进化,一直是公众和科学家共同关心的问题。

本章将从四个方面来阐述这个问题。首先,简单介绍一下病毒的进化研究概况;其次,简述病毒的进化研究方法,特别是病毒核酸序列上进化信息的分析方法;再者,以病毒聚合酶的核酸序列信息来阐述一些病毒的进化关系;最后,具体地谈谈流行性感、艾滋病毒、痘病毒等进化情况,并以这些病毒进化的研究为例,较为深入地分析病毒进化上的一些规律和一些研究思路。

## 一、病毒进化研究概述

### (一) 问题的复杂性和研究简史

病毒的进化是一个非常复杂而难以研究的问题,这个问题的复杂性在于:

(1) 病毒特别小,一般只有在电镜下放大几十万倍才能看到,所以常规的进化生物学研究方法难以移植到病毒进化的研究上。

(2) 病毒在漫长的演化史中基本上没有“化石”或“遗体”可言,难以获得系列的病毒材料以供进化研究。

(3) 病毒的进化具有一定的随机性,也就是说每种病毒的每一步演变都不是唯一的可能事件,而都是在一定概率下发生的结果。这样使得我们很难从病毒的现状、或短暂时间内病毒演变的情况来推论出病毒演变的过去和未来。

即使如此,对病毒进化的研究还是在不断地向前发展,甚

至可以辩证地说,正是问题的复杂性才给研究带来了广阔的发展空间。早在 19 世纪 80 年代,Pasteur 利用兔子对尚未明了的狂犬病毒进行反复传代,最终制成有效的狂犬减毒疫苗,说明了病毒不仅可以变异,而且变异的结果具有一定的稳定性。在自然进化方面,第一个被深入研究的病毒是兔黏液瘤病毒,它本来是一种在美国引起兔子良性黏液瘤的病毒,传到欧洲后毒力变得异常强烈。1950 年,澳大利亚“引进”这种病毒超强毒株,企图用来消灭泛滥成灾的野兔,但是不久,自然变异的毒力较弱的毒株便成为流行的主流,并且野兔的特异性免疫力也大大加强了,所以野兔问题最终还是没有得到彻底解决。这个例子不仅对病毒的进化具有重要意义,而且是研究寄生物-宿主之间共进化(co-evolution)关系的范例。

分子生物学兴起,特别是 20 世纪 70 年代以来,基因操作技术的飞速发展,为病毒的进化研究提供了最直接的研究方法。在此之前,虽然通过宿主范围、病毒形态、病毒血清学特征已经将病毒分门别类,但是这种划分并没有提示各种病毒之间存在什么样的亲缘关系,而病毒核酸分子的序列分析可以表明,各种病毒之间哪些基因是同源的以及同源性的 大小,从而提示各种病毒之间或远或近的亲缘关系。

从 1982 年 Moss 和 Paoletti 等人首次利用痘苗病毒作为载体表达外源基因以来,重组病毒的基因工程发展迅速。这实际上是利用病毒(人为的)进化来制造疫苗,或生产某种具有生物活性的物质为人类服务,由此提示我们,人类和病毒之间并不是谁将消灭谁的问题,而是可以相互协调、共同发展的。当然这个进化观点不一定一下子能被人们普遍接受。

1999 年,第 11 届国际病毒学大会上病毒的进化成为一个新的主题,一方面反映了它是当前病毒学研究的热点,另一方面也反映了这一方面的研究开始从原始的资料积累阶段进入理论综合阶段。需要说明,这个方面的研究很多地方超出了医学病毒学的范畴,但不妨碍我们学术上的探讨。

## (二) 病毒存在的基本方式

病毒是以 quasispecies 形式存在于宿主体内的。quasispecies 一般被翻译为“准种”,但是这种译法不太清晰。“quasi”前缀是“带有……性质”的意思,如“quasi-official”译作“带有官方性质的”,所以 quasispecies 应该译作“带有物种特性的群体”。这个概念最早由超循环理论创立人 Eigen 在 1971 年提出,指的是大致相同,但略有不同的一群个体的组合。对于生物学而言,这种组合就叫做物种(species)。

但是现代群体遗传学基本上闲置了“物种”这个概念,而突出“居群”(population)的概念。“居群”指的是在相对独立的范围内存在的、属于同一物种的所有个体的组合,居群之间可以存在一定的基因流动(gene flow),但如果居群之间在时空上相隔太远,则基因流动的机会就很小。由于病毒比一般生物的变异要快得多,所以病毒实际上是以大致相同、但略有不同的一群病毒粒子的形式存在的,而且被宿主个体所构成的相对独立的环境分割开来,所以病毒实际上是以类似居群的形式存在的,这样根据学术的发展,quasispecies 应该翻译为“居群样”,加一个“样”是与“quasi”相对应,还可以区别于普通生物的居群,为了简洁,可直接译为“居群”。之所以说这么多,是因为病毒以居群的形式存在,而不是以均一的形式存在,更新了我们对于病毒存在的认识,对于进一步探讨病毒的遗传、变异与进化

等问题具有重要的意义。机体内某一脏器中属于同一个种的病毒,也可以看作病毒的一个居群,在局部地区流行的、属于同一个种的病毒也可以看作病毒的一个居群,即居群的大小可以随着研究的需要和实际情况而进行一定的调整。

病毒的居群内部,也就是各个毒粒之间,核酸序列并不是完全一样的,相互之间略有不同,即存在一个“差异谱”(variation spectrum);虽然存在一个“差异谱”,但是一般还是以某一种序列为主,这个序列一般叫做主序列(consensus sequence)。一般测定的病毒核酸序列可以认为是主序列,因为主序列被测定的概率最大。主序列和变异谱不是不变的,而是随着时空的转移而不断变化着,正是这种变化才引发了病毒的进化,主序列和变异谱的变化规律也是病毒进化研究的重要内容。

### (三) 病毒变异的基本方式

变异是病毒进化的基础。在这个方面,病毒表现出一定的“量子”特征,即病毒是以整数倍的核苷酸、密码子或核酸节段进行变异的。在以核苷酸或密码子为“量子”进行变异时,一般表现出一定的匀速渐进的特性,相当于我们所说的“量变”;而以核酸节段为“量子”进行变异时,一般表现出一定的跃进性,相当于我们所说的“质变”,这种质变没有经过量变的积累,而是直接突现。

有时以核苷酸或密码子为“量子”进行变异时,经过遗传信息流动的放大作用也可以表现出一定的跃进性。例如,1982年在美国暴发的一次大规模禽流感,造成近2000万只鸡的死亡,原因是一株野鸟的流感病毒HA基因上一个密码子的突变而使其毒力剧增。这种点突变还使得从猫细小病毒演化出犬细小病毒,它们之间的同源性达到90%以上,但是这两种细小病毒并不发生交叉感染。

变异的数学基础是简单的组合问题,包括核苷酸的转换、颠换、插入和缺失,核酸分子内部片段的重组(recombination)和核酸节段之间的重配(reassortment)等。这些组合一般发生在病毒个体或群体的“基因库”(gene pool)内部,但很多情况下,特别是RNA病毒和反录病毒的进化还扩展到宿主的基因库上。因而,我们就不难理解许多病毒内部各个基因之间有的完全不相关,其中有些基因来源于宿主并经历过漫长的进化历史,而另外一些基因是通过病毒基因组内部各种组合快速形成的,甚至同一基因的不同部分在进化上来源也不一样,例如,痘苗病毒血凝素基因和人免疫缺陷病毒gp120蛋白的基因,在某一位置上都含有与免疫球蛋白结构域相似的序列,但是其他位置上的序列都和免疫球蛋白序列无关,提示这两种基因中不同部位来源可能不同。

病毒和宿主的核酸相互组合,可以快速地扩大病毒核酸的多样性。宿主的核酸成分在一些病毒癌基因(v-oncogene)和病毒因子(virokine)等病毒基因上表现得很明显。在某些黄病毒的基因组上还发现了宿主(ubiquitin,泛素)等核酸成分的插入。

### (四) 病毒进化的基本特征

病毒在进化中有以下几个基本特征:

(1) 新病毒不断产生,而且基本上是从另一种宿主的病毒演化而来。也就是说,新病

毒其实不完全“新”，主要是从一种生物的病毒演变为另一种生物的病毒(表 1-1)。为什么可以从一种生物病毒演化到另一种生物的病毒?原来,病毒的基因组按照上面所讲的某种方式发生了变异。另外,还有些“新”的病毒原本就已经存在,它们对当地宿主不具有致病作用,但是对外来的宿主有很强的致病作用,所以同一宿主不同群体的初次流动可以引发某些新的病毒病。

表 1-1 新出现的人类病毒的来源(Morse,1990)

病毒	症状	分布	自然宿主
人流感病毒新亚型	呼吸道症状	世界范围	禽、猪、马
汉坦病毒	出血、发热	亚洲、欧洲、北美	啮齿类
裂谷热病毒	出血、土发热	非洲	蚊子、有蹄动物
黄热病病毒	黄疸、发热	亚洲、南美	蚊子、猴
登革热病毒	发热、土出血	亚洲、非洲、加勒比	蚊子、人/猴
阿根廷出血热病毒	出血、发热	南美	啮齿类
波利维亚出血热病毒	出血、发热	南美	啮齿类
拉沙热病毒	出血、发热	西非	啮齿类
爱博拉病毒	出血、发热	非洲	灵长类
人免疫缺陷病毒	免疫缺陷	世界范围	灵长类

(2) 新病毒产生后,在新的宿主以较快的速度进行变异分化。例如,对于细小病毒来说,犬细小病毒是新病毒,猫细小病毒是老病毒,两者差异不大,但是犬细小病毒进化速度较快;对于流感病毒而言,人流感病毒是新病毒,禽流感病毒是老病毒;人流感病毒进化速度较快;对于慢病毒而言,人免疫缺陷病毒是新病毒,猴免疫缺陷病毒是老病毒,人免疫缺陷病毒比猴免疫缺陷病毒变异快。

(3) 新病毒稳定后,病毒的毒力大多处在中等水平。例如,兔黏液瘤病毒在澳大利亚释放之后的十几年的跟踪研究,发现强毒株逐渐演变为以中等毒力为主的病毒(表 1-2);副黏病毒引起的禽新城疫在我国开始流行时,毒力很强,这些年逐渐产生了一些弱毒株,但是整体上,既不以强毒株为主,也不以弱毒株为主,而是以中等毒力的毒株为主;猪瘟病毒的演化也有这样的规律,使得现在的猪瘟在有些地区以非典型猪瘟、温和型猪瘟、无名高热等为主。

表 1-2 澳大利亚兔黏液瘤病毒在 1951 年到 1981 年间毒力的分布情况(%) (Fenner,1983)

毒力的等级	I	II	III	IV	V	样本量
对兔子的致死率(%)	>99	95~99	70~95	50~70	<50	
平均存活时间(d)	<13	14~16	17~28	29~50	—	
1950~1951	100					1
1952~1955	13.3	20.0	53.3	13.3	0	60
1956~1958	0.7	5.3	54.6	24.1	15.5	432
1959~1963	1.7	11.1	60.6	21.8	4.7	449
1964~1966	0.7	0.3	63.7	34.0	1.3	306
1967~1969	0	0	62.4	35.8	1.7	229
1970~1974	0.6	4.6	74.1	20.7	0	174
1975~1981	1.9	3.3	67.0	27.8	0	212



(4) 病毒的进化既有一定的随机性,又受到一定的选择压力而呈现一定的方向性和稳定性。首先,病毒以上述方式产生的各种变异都是在一定概率下随机产生的,但是这些随机产物还要经过病毒整体活性和病毒与宿主相互关系的选择。从病毒的宿主特异性来看,选择的对象是病毒的群体,而对病毒群体来说,选择的对象则是病毒个体与基因,例如,脊髓灰质炎的疫苗一般采用活的弱毒苗,它只会引起一过性感染,不会致病。但是在这个群体中,极个别的病毒会引起疫苗事故,即儿童接种疫苗后发生严重感染,经查证原来是病毒通过变异,突破了机体的防御功能(也就是选择压力),在体内大量繁殖。

(5) 病毒各个基因以及基因的各个部分在进化上具有不同的进化特征。例如,流感病毒的8种蛋白质均有各自不同的进化特性,其中病毒内部蛋白质稳定保守,病毒表面蛋白进化较快;就血凝素蛋白而言,与宿主免疫系统作用的部分变异较快,而与细胞黏附的部分在进化上相对稳定。

### (五) 病毒进化的根本目的

病毒由于自身的一些特性能够从宿主体内获取一定的物质与能量,这些物质和能量从数学上来说,转化为病毒的多样性,而且在自然过程中转化的效率向着越来越高的方向发展,引起了病毒的进化。所谓多样性,可以从它的反面来理解,它是和“少与一致”相反的。这样从病毒的进化是病毒多样性增长的结果来看,病毒进化的根本“目的”就是为了增加病毒的多样性。当然,因为病毒多样性的发展需要从宿主体内获取一定的物质和能量,所以,必然受到宿主环境的制约,这些可以用来简单地解释上面病毒进化的特征。

首先,“旧”病毒通过跨宿主转移的方式演化是较为高效地扩增病毒多样性的方式,它比完全产生一个新病毒容易,也有利于病毒在更广泛的宿主环境中进行数量上扩增和种类上发展。

其次,新的宿主环境相对病毒的多样性发展而言是一片未开垦的土地,所以一开始病毒多样性发展较快,直到病毒多样性在新宿主环境中达到饱和为止,这样新病毒出现后,在新的宿主中以较快的速度进行变异。

新病毒稳定后,病毒的毒力大多处在中等水平,这是因为毒力太强,最后造成宿主和病毒共同灭亡,双方都不能传代,不利于宿主和病毒双方多样性的发展,将在循环中被逐步淘汰。中等毒力的毒株既能保持一定的病毒数量,又能不断传代下去,因而呈现一定的优势。而弱毒株虽不能对宿主造成损伤,但也不能产生大量的病毒,这样也不利于病毒多样性的增长。也有例外,如水禽流感病毒既能产生大量后代,又不对水禽致病,所以水禽流感病毒以弱毒株为主。

病毒按照随机组合的方式产生很多种类的病毒,所以病毒多样性的增长具有一定的随机性;但是病毒多样性增长的基础是要从宿主体内获得一定的物质和能量,这就要求病毒必须维持自身的某些特性和遵循宿主环境的某些特殊要求,并且不能导致宿主群体的彻底毁灭,这样使得病毒的进化承受一定的选择压力,也使得宿主和病毒在长期相互作用中维持着相对的稳定性。由于病毒各个基因和基因的各个部分承受的选择压力有所不同,所以病毒各个基因以及基因的不同部分具有不同的进化特性。

## (六) 病毒在现代进化生物学中的地位

过去多以植物、动物的实体或化石进行结构与功能的考察来研究进化问题。但是这些生物进化往往跨越特别长的时段,而且影响因素往往特别多,因而常常得出的只有一些推测,不是无可置疑的科学结论。而病毒一方面可以看作大分子的复杂装置;另一方面,又可以看作一些生物个体或群体。从这一点来讲,病毒是沟通分子进化理论和传统进化理论的桥梁;而且病毒基因组特别小,进化特别快,受干扰的因素特别少,因而在分子生物学飞速发展的今天,它是进化研究很好的材料,从中能够得出一些较为肯定的、清晰的结论。总之,病毒在现代进化生物学中应当占有重要的地位。新近,陈继明等从数学上推导出的新的进化理论,不仅能够更好地解释以前进化理论难以解释的一些复杂的进化现象,而且在病毒的进化研究中还找到一些更为直接的实验证据。本章第五部分将指出,进化生物学和病毒的进化联系起来进行研究,也是当前人免疫缺陷病毒进化研究的主要特色。

## 二、病毒核酸序列上进化信息的研究方法

病毒的进化是病毒学中综合性较高的研究领域,其研究方法涉及流行病学方法、临床病毒学方法、分子病毒学方法等,但是其特色的研究方法却是对病毒核酸序列进行进化信息的处理方法。

### (一) 病毒核酸序列进化的特征

病毒核酸序列相对于核酸酶切图谱而言,更完整地说明了各个病毒核酸之间的相似程度,从而更明确地提示它们之间同源性的的大小,因而具有重要的进化意义。由于PCR技术和DNA自动测序技术的推广,像病毒这样较小的基因测序工作已经不太复杂了。但是序列的测定只是分析进化关系中的第一步,还要经过一些复杂的信息处理,才能较为准确地揭示病毒之间的进化关系。

分析核酸序列上的进化信息实际上是要揭开这种核酸序列进化的特征。反过来,病毒核酸序列进化的特征又是确立信息处理方法的基础,也是信息处理效果的检验标准。

在大量实验资料积累的基础上,现在可以总结出一些病毒核酸序列进化的特征。首先,同一个病毒“种”内的不同分离株之间一般只有少数点突变造成的差异,而且经常出现在密码子的第三位,属于同义突变,核酸的进化也因此比蛋白质的进化要快一些。核酸序列上可能存在一些缺失和插入,但是一般不会有移码性突变,否则会急剧地破坏病毒活性。在信息学上插入和缺失价值相当,所以处理过程中一般将它们统称为插缺(indels)。其次,同一病毒“属”内不同病毒“种”的基因组之间可能有较多的点突变的差异;不同的基因、基因上不同的位置在进化上的保守性也不一样,有的进化较快,有的进化较慢。第三,同一病毒“科”内不同的病毒“属”之间一般核酸序列上有明显的差异;相似的只是存在于病毒某些酶与结构蛋白的基因序列上;有时这种相似性明显地表现在某一较短的特征序列上,即基序(motif),而基序之外相似性很小,提示它们从某一祖先演化过程中发生了很

多次插缺;有时不同的病毒“属”之间一般核酸序列上没有多少相似性,但是核酸和蛋白质的二级和三级结构很相似,提示一级结构比二级结构和三级结构进化较快。例如,几乎所有20面体对称的RNA病毒的结构蛋白在空间上都有一定的相似性。

不同病毒属共有的一些基因有时来自完全不同的病毒,甚至来自不同宿主的基因,而且在组合和排列上可能也有所不同,对它们的处理尚未达成一致意见,因而给这些基因的处理带来一定的难度。另外,还有一些基因是有“属”特异性的,它们可能和各个属共有的基因在进化上完全不相关,使得病毒属之间的进化关系难以从这些基因来孤立地判断。

病毒核酸序列上的相似虽然一定程度上反映了参比序列间的亲缘关系,但是亲缘关系较远的毒株在相同的环境下承受相似的选择作用而表现出一定的趋同作用,导致相似性有时难以准确反映亲缘性的真实值。

上一节提到病毒变异的数学基础是各种层次上“量子”的组合问题,通过不同层次的“量子”的组合,形成了更高层次的多样性,如4种核苷酸组合形成了一些相对独立的特征序列,包括基序、模块(module)和结构域(domain),这些特征序列与一些非特征序列一起通过组合构成更为多样的基因,而这些基因通过组合形成数以亿计的物种。层次太低的“量子”,如单个的核苷酸包含的进化信息几乎为零;基序、模块等“量子”包含的进化信息虽然很明显,但是信息总量不大;结构域和基因等层次上的“量子”既包含了较为清晰的进化信息,又具有较大的信息量,是较好的分子进化研究对象;而基因组或个体层次上的“量子”虽然包含了非常多的进化信息,但是正因为信息量太大而显得错综复杂。病毒的核酸序列上也有这些进化特征。

## (二) 病毒核酸序列上进化信息的分析方法

核酸序列的信息处理是非常复杂的过程,但是由于近年来分子生物学技术的飞速更新和环球网(world wide web)的发展,以及各种软件的开发,开展这项工作已容易了许多。虽然如此,我们还需要了解一下这个信息处理的基本过程和基本原理,以便知其所以然。

### 1. 病毒核酸序列测定

一般病毒核酸提取后用PCR方法进行扩增,扩增的产物纯化后就可以直接进行测序,不必像以前那样克隆再测序。对于多节段的病毒而言,各节段上有时共有某一序列,如果利用这些共有序列设计的一些通用引物来进行PCR扩增,得到的产物可能含有不同的节段,纯化较为困难,所以仍然需要克隆到测序质粒里面,这样虽然增加了克隆这个步骤,但却减少了引物设计的困难,也可以节省引物合成费用。

### 2. 查找相似的序列

当我们拿到病毒某一段核酸序列后,只有从已有的数据库中找到和它相似的一些序列后才有足够的信息处理材料。对于一个未知病毒的序列,通过相似序列的检索还有鉴定的作用。这个工作如果手工检索,则工作量是巨大的,但是现在可以从www上利用美国生物技术信息中心(NCBI)提供的BLAST服务器或欧洲生物信息研究所(EBI)提供的FASTA服务器进行快速检索,它们的网址分别是<http://www.ncbi.nlm.nih.gov>和

<http://www.ebi.ac.uk/htbin>。在 NCBI 和 EBI 的主页上选择相应的图标,如 Program 项选择 blastn 表示输入和检索核酸序列,blastp 表示输入和检索蛋白质序列,Database 项一般选择 nr,表示在 GenBank + EMBL + DDBJ + PDB 几个大型数据库中无重复地检索。然后按照 FASTA 格式输入(或从磁盘读入)测定的序列(一次只能读入一个序列),最后选择“提交查询内容”,屏幕上将显示检索结果。用户输入自己的 E-mail 地址后,检索结果也将同时寄到用户的电子信箱中。对于 BLAST 而言,不太熟悉的用户可以先使用 Basic BLAST 进行检索,只需要输入序列就行,然后根据检索结果尝试性设置一些参数使用 Advanced BLAST 进行检索。近年来,又有更新的服务器 Power BLAST 问世,在敏感度、特异性和检索速度上都有所提高。

### 3. 序列的同源性大小计算以及进化树的制作

从核酸序列的相似性可以计算序列的同源性,并据此制作反映进化上亲缘关系的树状图,即进化树。这样处理显然带有一定的主观性和不精确性,因为我们按照相同的原理可以对同一群病毒制作出很多进化树来,到底应该选择哪个进化树,没有一个客观和完全理性的标准;其次,病毒的进化并不完全以某一恒定的速度不断演变;另外,我们经常缺乏足够的样本。但是由于病毒进化研究的复杂性也只好如此处理。我们所能做的只是按照这个原则尽量改进具体的处理方法。这种处理在流感病毒、人免疫缺陷病毒、乳头瘤病毒等病毒的进化推算上,结果让人能够接受。由于现在已经有各种网络软件(两个重要的网址分别是:<http://evolution.genetics.washington.edu> 和 <http://www.sinauer.com/titles/frswofford.htm>)可以较为简易地进行这项工作,所以这儿只是说明一下大致的原理与步骤。

(1) 确定最佳对比:由于核酸序列中插缺几乎是没有规律可言的,而每一次插缺都给序列对比带来一定的困难,如何解决这个问题从而找出序列间最佳的对比,最直观、也是最基本的方法可能就是点阵做图(dot diagram comparing)了。它也是上述计算机检索相似序列的基本原理。

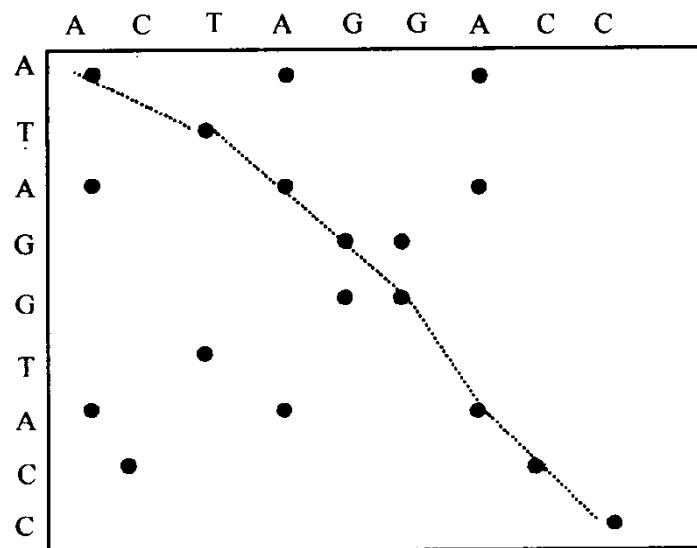


图 1-1 点阵做图法寻找最佳对比示意图

以图 1-1 为例,横向的是序列 1,纵向的是序列 2。序列 1 和序列 2 上各个位点实际上构成了一个矩阵。如果矩阵上每点对应序列 1 上的核苷酸和序列 2 上的核苷酸是一样的,则在该点的位置画一个点,如果不相同,则不做记号,这样就得到一个点阵图。两个序列连续相同的一段将在图上呈现 45°角的线段,叫做相似线。用一条虚线将相互最靠近的相似线连起来,这样我们不仅找到了序列最佳对比,而且从这条虚线的一些转折发现插缺所在位置。如果转折的部分倾向于水平,表明序列 1 有插入,或序列 2 有缺失;反之则相反。这种点阵做图法需要进行一定的修改,才能提高信噪比和运算效率。

(2) 计算遗传距离:一般按照双参数法计算:

$$K = -1/2 \ln[(1 - 2p - q)(1 - 2q)^{1/2}]$$

$K$  代表遗传距离, $p$  指序列间转换的核苷酸数量, $q$  指序列间颠换的核苷酸数量,这种计算存在的问题,如同义突变和错义突变在遗传上造成的影响并不相同,但是未被计算在内;又如亲缘关系较远的毒株在相同的环境下承受相似的选择作用而表现一定的趋同作用,导致计算的遗传距离显著低估;同一序列高变区和保守区进化速度不一样,若从整个序列来计算遗传距离也会造成低估。这些不足之处是计算方法应改进的地方,一般用户可以按照软件的指示去改进。

(3) 进化树的构建:最简单的是算术平均对群法(UPGM)。如果 A、B、C、D 四株毒株间的进化距离分别为  $K(A,B)=4$ ,  $K(A,C)=9$ ,  $K(A,D)=20$ ,  $K(B,C)=7$ ,  $K(B,D)=18$ ,  $K(C,D)=17$ 。 $K(A,B)$  最小,因而可以把 A 和 B 合到一个分支上,再按照算术平均值计算  $K(AB,C)=[K(A,C)+K(B,C)]/2=8$ ,  $K(AB,D)=[K(A,D)+K(B,D)]/2=19$ ,这样 AB 这一支和 C 又可以合到一个更大的分支上,可以画出如图 1-2 的进化树。

因为病毒在各个分支上进化的速度不完全一样,而且假设了每个分支的长度都是一样的,所以肯定散失了部分信息,而且大量毒株进行比较时工作量很大。根据上面所述的数据,考虑到邻近序列的信息,有一种改进的画法,叫做 neighboring joining, NJ 法,结果参见图 1-3。

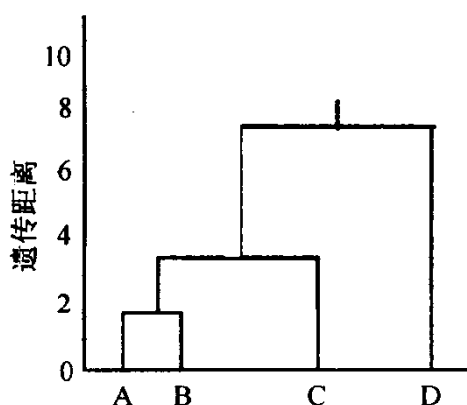


图 1-2 根据 UPGM 制作的进化树

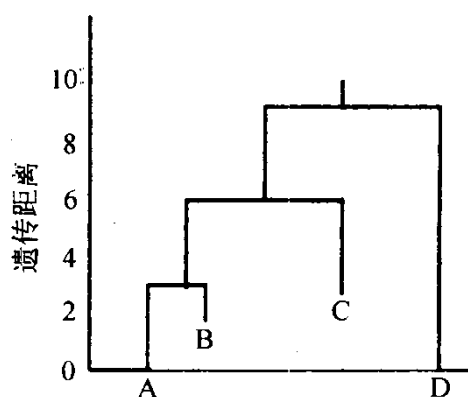


图 1-3 NJ 法制作的进化树

现在使用最多的是最节约(maximum parsimony, MP)法,这个方法和遗传距离法不同,它尽可能地使树的顶端到树根之间的距离最小。这种方法首先去除了进化上不变的位点,然后再去除只是在某一个序列中改变的位点,只有两个以上序列在某一位点上都发生了变化,才被认为具有进化意义。按照这种方法,根据表 1-3 的数据可以画出图 1-4 的进

化树。

表 1-3 有进化意义的位点(加重)

毒株	序列
A	TACTCGG
B	TTACACG
C	GTACACG
D	GAACTCG

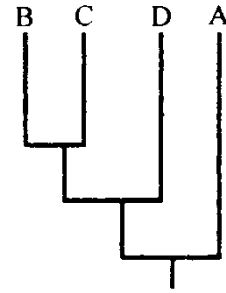


图 1-4 根据表 1-3 数据和最节约法制作的进化树

这种方法执行起来较为简易,是目前最常用的进化树制作软件 PAUP(已有 3.1 版本出现)的基础。但是这样处理,有时不止获得一个进化树,所以还要进一步制定一些评判标

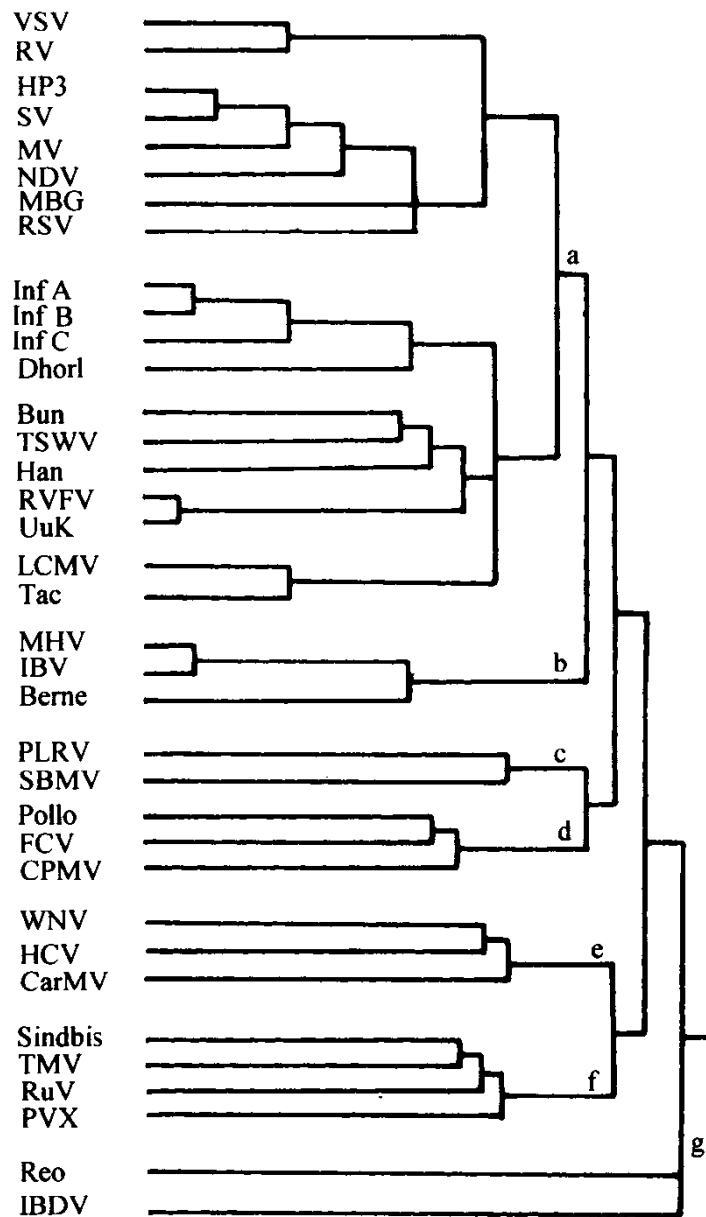


图 1-5 基于 RNA 依赖的 RNA 聚合酶序列建立的 RNA 病毒进化树

- a. 负链超级进化群 b. 冠状病毒样超级进化群 c. 南方菜豆花叶病毒样超家族 d. 微小 RNA 病毒样超级进化群 e. 黄病毒/香石竹斑驳病毒样超级进化群 f. 甲病毒样超级进化群 g. 双链超级进化群

准用以筛选。另外在趋同进化和以不同速度进化的两种情况下,这个方法显得不太合理。例如,用它来制作人免疫缺陷病毒 I 型 gp120 V3 高变区的进化树就不很理想。

另外一种进化树的制作方法是最可能(maximum likelihood, ML)法,这种方法也考虑到了序列上各个位点的进化意义,但是 ML 法还同时计算每个位点各种变化概率的大小。另外,ML 法还通过调节各个分支的长度来描述理论上的最可能的谱系关系,显然这在实际执行上要比 MP 法繁重得多。近年来 ML 法又有所发展。

现在进化树的制作计算上有很多算法,感兴趣的读者可以到 <http://evolution.genetics.washington.edu> 网上下载有关文件。

进化树可以分为两种,即有根树和无根树。有根树表明了进化的方向和过程,而无根树却没有。图 1-2、图 1-3、图 1-4、图 1-5 是有根树;图 1-6 是无根树。无根树往往是因为缺乏其他佐证,难以确定参比对象起源时间。在有根进化树旁一般沿着分支的方向给出一个衡量变异大小的比例尺,而与分支方向垂直的线段一般只是为了隔开足够的距离,其长短没有进化意义。在无根进化树上一切所有线段的长短都表示进化距离的大小。本章为了简洁,多数没有标出比例尺。

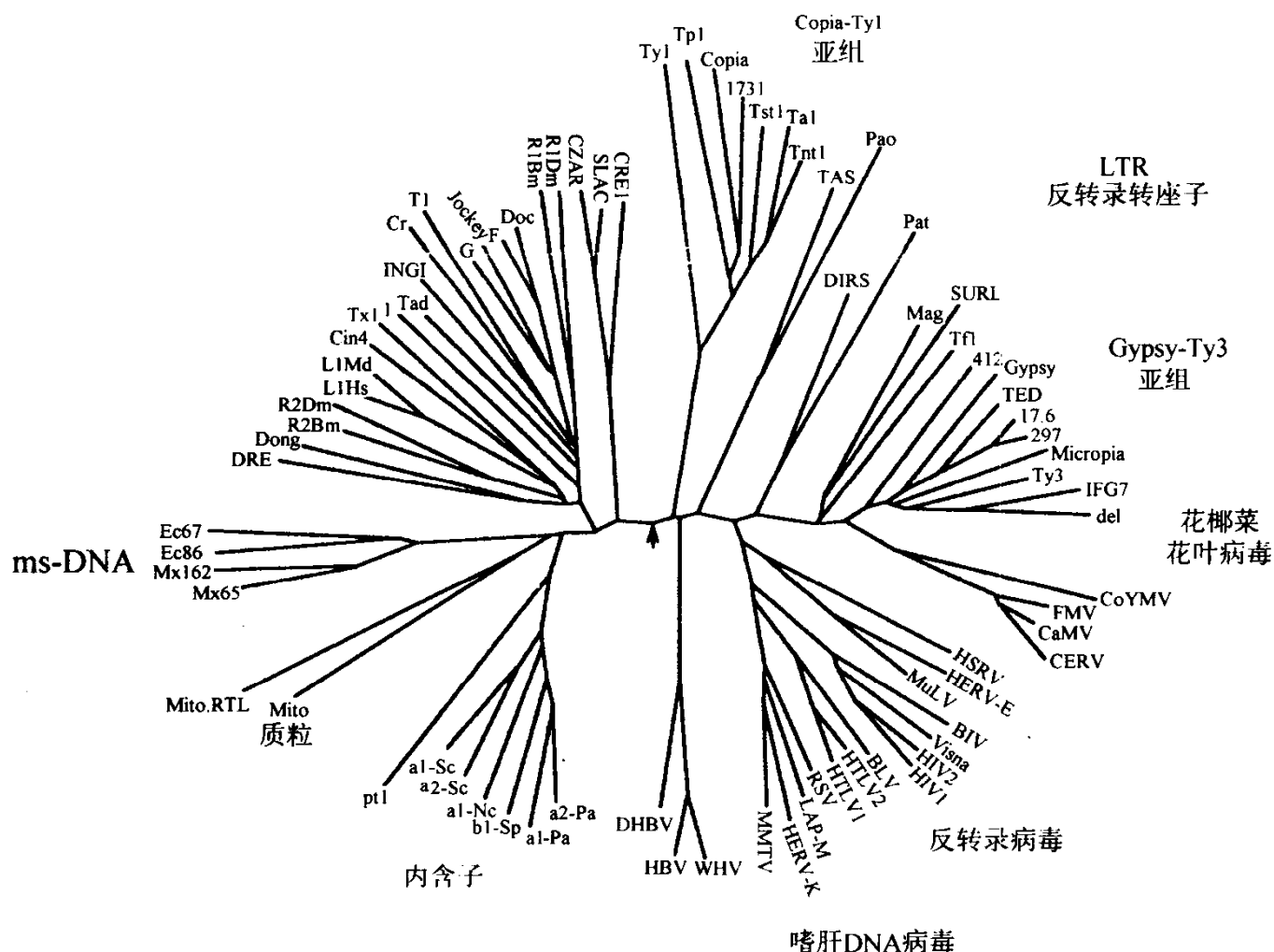


图 1-6 基于逆转录酶序列信息的反录成分进化树(Morse, 1994)

### 三、病毒聚合酶核酸序列与病毒的进化关系

上面几节已经提到病毒各个基因在进化的过程中,可能表现出不同的演变过程、演变特征,因而有时难以从某一个基因来推断病毒的实际进化情况。相对而言,只有病毒的聚合酶基因是较好的病毒进化“分子种”,这是因为病毒聚合酶一般来说是病毒特有的基因,进化相对缓慢,能够记录较长的进化历史;另外,它在各种病毒中执行相似的功能,承受相似的选择压力,所以似乎又能忠实且一致地反映各种病毒进化关系。这样,病毒的聚合酶在病毒进化研究上具有重要的意义。

#### (一) 病毒聚合酶的基序

病毒的聚合酶可以分为四大类:RNA 依赖的 RNA 聚合酶、RNA 依赖的 DNA 聚合酶、DNA 依赖的 RNA 聚合酶、DNA 依赖的 DNA 聚合酶,以下分别简称为 RR 酶、RD 酶、DR 酶、DD 酶。RD 酶又叫做逆转录酶,DR 酶又叫做转录酶。

Kamer 和 Argos 最早研究了脊髓灰质炎病毒以及其他正链 RNA 病毒的 RR 酶,发现了 Gly-Asp-Asp 为核心的 5 个氨基酸(多半是疏水性氨基酸)组成的 GDD 基序,它们一般位于一个  $\beta$  发夹结构上。在很多 RD 酶的序列中也发现了 GDD 基序。之后,Poch 等人对 RR 酶和 RD 酶的基序进行了深入研究,发现了一些新的基序,参见表 1-4。这些基序一般在聚合酶的高级结构上占据重要的位置,并通过定点突变表明它们对酶的活性很重要。

表 1-4 一些病毒 RR 酶和 RD 酶共有的基序

毒株类别	基序 1	基序 2	基序 3	基序 4
HepBWo	TDLQWLSL D VSAAFYHI	IMGFR.KLPM G	VGLSPFLLAQFTSALAS	VFAYM DD LVLG DLGIHL.NVN K TK
HepBDV	VGMPRISL D LSQAFYHL	VYYFR.KAPM G	VGLSPFLLHLFTTALGS	TFTYM DD FLLC ELGIRI.NFD K TT
MMTV	KGWEIII D LQDCFFNI	RFQWK.VLPQ G	MKNSPTLCQKFVDKAIL	IVHYM DD ILLA KHGLVV.STE K IQ
HIV-1	KKKSVTVC D VGDAYFSV	RYQYN.VLPQ G	WKGSPAIFQSSMTKILE	IYQYM DD LYVG RWGLTT.PDKK HQ
IBDV	NSNTWYSI D LEK.GEAN	LQIKSYGQGS G	NAAT.FINNNLLSTLVL	IERSI DD IRGK LLGWSA.TYS K DL
BTV	GYTLEQII D FGY.GEGR	DLALIDTHLS G	ENST.LIANSMHNMAIG	EQYVG DD TLFY KCGHEA.SPS K AL

对于 DD 酶和 DR 酶也有一些基序。表 1-5 只列举了一些病毒 DD 酶部分基序,这些基序对蛋白质高级结构的形成很重要,有的还参与酶的催化活性和金属离子的结合活性。这些基序似乎在各种生物 DNA 聚合酶中都呈现保守性。有人甚至据此制作出各种生物 DNA 依赖的聚合酶的进化树来。但是,实际上这些基序是否是趋同进化而来的呢,或趋同进化是否参与了这种保守性的形成?值得怀疑。



表 1-5 一些病毒 DD 酶共有的基序

毒株类别	基序 1		基序 2		基序 4	
Adeno-2	PRTERLFVTY	DVE TYTWMGAFGK	RFLELYIVGH	N ING FD EIVLAAQVIN	KYDIKETLD	YCALD VQVTAEL VEK
Adeno-7	PRTERLFLTY	DVE TYTWMGSFGK	RFLELYIVGH	N ING FD EIVLAAQVIN	RYDIIQETLN	YCALD VLVTAEL VAK
HSV	DLPAYKLMCF	DIE CKAGGEDELA	.QYGPEFVTGY	N IIN FD WPFLAKLTD	GPAQRGVIGE	YCLQD SLLVGQLEFK
CMV	SWPRYRCLSF	DIE CMSGEGGFPC	RYAPAFVTGY	N INS FD LKYILTRLEY	NAEGRAQVGR	YCLQD AVLVRDLENT
FPV	FEVKTTYLLE	DIE CQFDKKFPSV	EHRDFVITE	N GNN FD LRYITNRLEL	NLEIALDMER	YCLHD ACLCKYIWDY

## (二) RNA 病毒的超级进化群

RNA 病毒占有所有病毒种类 80% 以上, 包括很多生物学特性、形态学特征截然不同的病毒, 就从基因组而言, 可以分为正链 RNA 病毒、负链 RNA 病毒、双链 RNA 病毒, 它们还被进一步划分成各个科和属。到目前为止只有一个病毒的“目”被国际病毒分类委员会 (ITCV) 认可, 它是由弹状病毒科、丝病毒科、副黏病毒科组成的“单负目” (mononegaviral), 设置这个“目”还表明这 3 个科之间存在一定的进化关系。这种进化上的相关性特别表现在它们基因组结构上的相似性, 也表现在某些副黏病毒的基因序列更像弹状病毒。

现在, 由于对各个 RNA 病毒科的基因组, 特别是病毒 RNA 聚合酶序列都有所了解, 从这些序列分析中让人感到大多数真核生物 RNA 病毒, 不管是植物病毒还是动物病毒, 都可以归到一些 RNA “超级进化群” (supergroup) 中, 包括小 RNA 病毒样超级进化群 (picorna-like supergroup)、 $\alpha$ -样超级进化群 (alpha-like supergroup)、黄病毒样超级进化群 (flavi-like supergroup)、冠状病毒样超级进化群 (corona-like supergroup) 等。表 1-6 列举了部分超级进化群的特征, 而图 1-5 则是根据 RNA 依赖的 RNA 聚合酶的序列推算的进化树。

表 1-6 部分超级进化群的共有特征

超级进化群	囊括的科属	特征	超级进化群	囊括的科属	特征
负链	副粘病毒	核酸的极性为负	黄病毒样	黄病毒	核酸的极性为正
	弹状病毒	核酸末端互补		丙型肝炎病毒	5'-cap, 3'-polyA
	正粘病毒	5'-pyrophosphate		瘟病毒	无亚基因组 mRNA 保守的基因组合
	布尼亚病毒	相似的基因组合	冠状病毒样	冠状病毒	核酸的极性为正
沙粘病毒	无 polyA, 有包膜	土拉病毒		5'-cap, 3'-polyA	
小 RNA 病毒样	小 RNA 病毒	核酸的极性为正	动脉炎病毒	亚基因组 mRNA 保守的基因组合	
	杯状病毒	5'-Vpg, 3'-polyA	双链	双 RNA 病毒	双链 RNA
	线虫传多	无亚基因组 mRNA		隐病毒	分节段基因组
	面体病毒	mRNA		Crysto	5'-cap
	豇豆花叶病毒	多蛋白剪切		分病毒	3' 无 polyA
	欧防风黄点病毒	保守的基因组合	呼肠(孤)病毒		
马铃薯 Y 病毒		全病毒			