

新基因的起源与进化

李 昕 杨 爽 彭立新 陈 宏 王 文 *

(中国科学院昆明动物研究所细胞与分子进化重点实验室, 中德马普青年科学家小组, 昆明 650223; 西北农林科技大学动物科技学院, 杨陵 712100; 中国科学院研究生院, 北京 100039; 徐州师范大学生物技术研究所, 徐州 221116.

* 联系人, E-mail: wwang@mail.kiz.ac.cn)

摘要 随着基因组数据的大量积累, 人们愈加认识到各种有机体中基因数目的巨大差异. 这些差异的存在表明, 新基因如何产生不仅是一个重要的进化生物学问题, 也是生命科学中面临的一个基本问题. 对新基因起源机制的探索, 可以追溯到大半个世纪以前, 然而直到上世纪 90 年代第一个年轻基因——精卫基因(*jingwei*)的发现, 才使以实证方法研究新基因起源的分子机制成为可能. 此后的 10 多年中又陆续发现一些新的年轻基因的例子, 对这些基因起源与进化的研究极大地丰富了人们在这一领域的认识. 但目前有限的例子难以从整体的水平对基因组中新基因产生的速率以及新基因的产生对原基因组的影响等问题作出解答. 我们正致力于在基因组的水平寻找更多年轻基因的实例, 以期总结新基因起源与进化的一般规律.

关键词 新基因 起源 进化 分子机制 正选择

随着人类和其他一系列物种全基因组序列的测定, 人们发现不同生物在基因组大小及基因数目上存在巨大的差异, 如一种支原体 *Mycoplasma genitalium* 基因组大小为 5.8×10^5 bp, 仅含 470 个基因^[1], 而人的基因组大小为 3.0×10^9 bp, 基因数目约为 3 万多个^[2], 两者基因数目相差数十倍. 从横向上看, 正如我们在果蝇中所观察到的^[3-5], 即使分化时间很短的近缘物种间, 基因的种类和数目也不尽相同, 说明生物进化的过程伴随着基因组的大小及基因数目的不断变化. 由此引出一个根本性的生物学问题: 这些新基因是如何产生的? 对此问题的了解还有助于我们解决其他一些进化生物学的问题, 如种的形成和分子进化与物种进化的关系等. 此外, 可能还有应用科学上的意义. 例如, 知道了自然界怎么产生基因的规律后, 会对人类设计制造新的生物活性药物有指导作用.

人们对新基因起源这一问题的兴趣可以追溯到 20 世纪 30 年代, 尽管当时对遗传物质的本质还没有清晰的认识, Haldane^[6]和 Muller^[7]就已提出通过基因重复可以产生新的基因. 此后, 得益于分子生物学实验手段的进步和遗传学的发展, 人们进一步认识了基因的本质, 观察到大量的实验现象, 如染色体重复、基因家族和断裂基因等, 并在此基础上提出了一些新基因产生的假说^[8,9]. 20 世纪 80 年代中期以后, 大规模基因组序列信息的获得以及分子进化和群体

遗传学理论的成熟, 更使得在基因组水平的理论预测成为可能^[10]. 然而由于基因组中的大多数基因产生太早, 在漫长的进化时间中积累的大量突变早已湮没了大部分重要的进化信息, 无论是基因最初产生的分子机制或是随后在群体中扩散并最终固定下来的群体动力学过程, 都已无法直接观察和检测. 因此直到 20 世纪 90 年代以前, 有关这一问题的探讨基本上是设想性或理论性的. 人们迫切需要能够获得一些年轻的新基因起源的实例, 使人们能够以实验的手段近距离观察并阐明新基因起源的分子机制和进化的动力学过程.

1993 年, 华裔学者龙漫远(Long)等人^[3]发现了第 1 个年轻基因——精卫基因(*jingwei*), 从此新基因起源的研究进入了一个新的时期. 此后, 又有司芬克斯(*sphinx*)基因^[4]和猴王基因(*monkey king*)^[5]等大约 20 多个年轻基因被报道. 与那些古老的基因相比, 年轻基因可以提供给人们新基因进化早期的结构、序列信息, 有助于推断其起源机制及进化力量^[11].

通过对已发现的这些年轻基因的研究, 我们已得到了新基因起源与进化的一些基本认识. 对此, Long 等人^[11-13]已作了很好的总结. 但为了能够归纳和总结新基因发生的分子机制和进化过程的一般规律, 我们还有必要发现和研究更多的年轻基因. 随着基因组数据的快速积累, 目前这一领域发展迅速, 而国内对这一新兴研究方向还比较陌生. 本文将对这

一领域目前发展的概况作一介绍, 并就我们的理解提出一些待解决的问题及简要介绍今后的研究方向.

1 新基因产生的分子机制

有关新基因起源的分子机制, Long等人^[13]已作过系统的介绍, 其主要有基因重复(gene duplication)、外显子重排(exon shuffling)、逆转座(retrotransposition)、可移动元件(mobile elements)、基因水平转移(gene lateral transfer)和基因分裂与融合(gene fission and fusion)等. 下文简述几种主要机制.

1.1 基因重复(gene duplication)

基因重复是人们最早认识到的新基因产生机制. 经典理论认为, 通过重复产生的冗余拷贝, 由于不受或很少受到选择压力, 不断积累各种突变, 与原基因(parental gene)产生分化, 最终可能产生具有新功能的基因. 根据重复区域的大小, 基因重复可分为单个基因重复、部分基因组重复(segmental duplication)和整个基因组重复(genome duplication)即多倍体化. 单个基因和部分基因组的重复主要通过不等交换产生, 而基因组重复是有丝分裂或减数分裂过程中发生错误产生的. 根据前人的研究, 基因重复是新基因产生的重要来源之一. Lynch和Conery^[14]利用果蝇、酵母、线虫、鸡、鼠和人的全基因组信息对基因重复的频率做了保守的估计, 约为每基因每百万年0.01次. Blanc等人^[15], Bubin^[16], Ball等人^[17]和Li等人^[18]分别对酵母、线虫、拟南芥、果蝇和人的基因组序列进行分析, 发现由基因重复产生的基因家族所包含的基因数占整个基因组的百分比在上述5个物种中分别达到30%, 48%, 60%, 40%, 38%. Gu等人^[19]利用多个物种的基因组序列, 发现大规模的和小规模基因重复都对脊椎动物的基因组的进化有着重要影响.

1.2 外显子重排(exon shuffling)

外显子重排是指由来自不同基因的2个或多个外显子相互接合, 或基因内部的外显子产生重复而形成新的基因结构. 20世纪70年代, 在真核生物中发现断裂基因后, Gilbert^[10]提出, 通过内含子介导的重组, 不同基因的外显子可发生互换, 使得原基因结构发生变化, 可能产生新的基因. 随后发现的实例证实了这一理论^[20]. 人们现已发现外显子重排可以由异常重组^[21](illegitimate recombination)和返座子介导的外显子插入^[22]等产生. 此外, 相邻基因间序列的缺失产生的基因融合也可造成外显子重排^[23]. Patthy^[24]

通过对大量蛋白质家族结构域的分析, Long等人^[25]通过对内含子相位的分析以及Li等人^[18]对5个真核生物基因组的共享结构域的分析, 都发现真核生物中相当比例的基因是由外显子重排产生的. 这些基因组水平的分析以及大量发现的实例使得人们认识到外显子重排在真核生物的新基因产生中扮演着重要角色.

1.3 逆转座(retrotransposition)

逆转座是指转录产生的RNA通过逆转录合成cDNA插入到基因组的过程. 由于通过逆转座产生的新拷贝一般不含启动子和调控序列, 使得大部分产生的序列成为假基因. 然而, 在特殊情况下, 逆转座序列通过原基因不正常转录携带有启动子^[26], 或者插入到基因组后获得外源调控序列^[3,4,27,28]而具有表达活性, 进而可形成新的表达特异性或新的功能. 从这个意义上, Brosius^[29]称逆转座子为进化的“种子”. 由于真核生物基因组中具有丰富的逆转座序列(例如, LINE序列在人有10万个拷贝), 它们可介导产生逆转座基因, 因此逆转座作为新基因产生的一种机制越来越受到人们的重视^[30,31].

1.4 可移动元件(mobile elements)

可移动元件包括转座子和逆转座子. 过去人们认为它们是自私基因, 仅仅是为了增加其在基因组中的拷贝数. 然而, 现在人们认为它对新基因的产生也有着积极的贡献. 可移动元件可以插入到原基因的外显子和内含子中, 形成新的外显子, 使得基因结构发生变化, 可能导致新基因的产生. 哺乳动物中含有大量可移动元件(例如, 人的基因组中Alu序列有30万~60万个拷贝), 使得可移动元件的插入频繁发生. Nekrutenko等人^[32]通过对人的基因组分析后, 发现编码蛋白质的基因中有4%的外显子是通过可移动元件的插入产生的.

1.5 基因水平转移(gene lateral transfer)

基因水平转移是指遗传物质从一个物种通过各种方式转移到另一个物种的基因组中. 在原核生物中, 转化、转导、接合和转染等现象是频繁发生的. 因此, 基因水平转移对原核生物的基因组贡献是相当大的. Ochman等人^[33]发现一些细菌基因组的16%是通过基因水平转移获得的. 对于真核生物, 基因水平转移主要通过逆转录病毒介导, 并且对基因组影响不大. 这些通过水平转移产生的外源基因在选择

作用下,经过突变积累,功能分化,可能形成新的基因.因此,基因水平转移也是新基因的来源之一.例如,一种毛滴虫(*Trichomonas vaginalis*)通过水平转移获得了嗜血菌(*Haemophilus influenzae*)的一种裂解酶,此裂解酶通过插入获得了24个氨基酸构成的一段信号肽,使其由胞内酶变成了胞外酶^[34].

2 新基因在群体中的固定

对于新基因的起源来说,通过不同机制产生的新拷贝只是提供了进化的原材料,如同大部分的突变会在进化过程中丢失一样,这些新拷贝也可能面临同样的命运.按照中性理论的估计,一个突变在群体中被固定的概率只有 $1/2N_e$ (N_e 为有效群体大小)^[35],并且由于大量的突变为有害突变,即使固定下来的新拷贝也有很大一部分成为假基因,而只有其中一小部分能够保留原功能或成为具有新功能的基因.那么新的基因是如何在群体中固定下来的,在其进化的过程中又受到什么作用力量的支配呢?这是新基因起源及进化研究的另一个重要方面.

到目前为止,在已发现的新基因中,通过基因重复、外显子重排、逆转座及可移动元件等所产生的新基因占绝大多数,但对其固定过程中动态变化的模型研究较多的主要集中在基因重复.其研究最早可以追溯到1933年Haldane的突变模型.随后Fisher^[36], Nei^[37], Bailey等人^[38], Kimura和King^[39]以及Li^[40]等各自提出并发展了一系列新的模型,在这些模型中提出大部分的重重复基因只可能是通过无功能的形式保存下来.阐明新功能基因的模型到Ohta^[41]才发展起来,到Walsh^[42]才形成了较完整的体系, Walsh认为,在 $\rho S \gg 1$ 时(ρ 为有利突变对无功能突变率的比, $S = 4N_e s$, 其中 N_e 为有效群体大小, s 为选择系数),新功能基因可能被固定下来,概率为 $1 - (\rho S)^{-1}$,并提出正选择(positive selection)在进化过程中是一个重要的推动力.但是为了解释真核生物中存在大量具有亚功能重复基因的现象, Force等人^[43,44]提出了复制-退化-互补模型(duplication-degeneration-complementation, DDC model).该模型认为,许多基因可能含有多个功能区域,基因重复后不同区域的互补失活会迫使2个拷贝都必须保留下来,从而导致基因的亚功能化(subfunctionalization),并指出以这种形式固定的基因随亚功能区域的数目及其突变率增加.

以上几个模型在一定程度上描述了中性选择、正选择在进化过程中的作用. Walsh^[45]和Ohta^[46]认为中

性选择与正选择两者都会在新基因形成过程中起作用,特别在一个大群体中,选择将大大增加形成新基因的概率.

Gu等人^[47,48]对基因重复后功能分化的问题做了大量的研究.基于位点进化速率的改变, Gu等提出了统计学的方法预测那些基因重复后有功能分化的拷贝,并且进一步检测出那些对功能分化有重要贡献的氨基酸位点.将此方法应用到一些蛋白质家族的分析,结果表明基因重复后的功能分化可能是一种普遍现象^[49,50].

但是,人们对于新基因产生中的实际群体动力学过程仍不得而知.目前,我们对年轻基因起源的研究正是希望通过发现更多保留大量进化信息、可检验的实例,认识这一问题的真实过程.现已发现的新基因都不同程度地观察到正选择的作用(表1),表明由选择驱动的快速进化在新基因的诞生过程中是一个普遍现象.例如,对叶猴中的胰核糖核酸酶基因(*RNASE1B*)的分析结果表明,其错义替换率(nonsynonymous substitution rate, 0.0310)显著地高于同义替换率(synonymous substitution rate, 0.0077),显示其进化过程受到了强烈的正选择作用,从而适应其在胃中消化细菌RNA的新功能^[51]. Moore等人^[52]在 *Arabidopsis thaliana* 基因组数据库中筛选出3个年轻基因分别产生于0.24, 0.5, 1.2百万年(Ma)前,数据分析表明其中2个基因在固定过程中受到正选择作用,并认为这最终决定其固定的命运.

3 已发现的年轻基因

表1总结了迄今已发现的年轻基因.其中, *sphinx* 基因是我们在果蝇中发现的第一个年轻的RNA基因.它的产生距今不超过2百万年,是一个非常好的近距离观察新基因,尤其是RNA基因起源与进化的实例.通过同源序列对比,我们发现其3'端外显子与ATP合成酶F链具有同源性,但它不含内含子部分,其两端有短的重复序列(TTCG),并且在3'末端有poly(A)序列,这些证据指示此外显子是由ATP合成酶F链逆转座插入产生的.而 *sphinx* 基因5'端的调控序列及外显子被推测是由原先已存在的基因所贡献,这2部分通过外显子重排形成一新的嵌合基因(图1).由于其序列上有多处导致移码的缺失、插入以及无义突变,此基因不可能是编码蛋白质的基因.根据其表达的数据,我们发现 *sphinx* 基因具有多种剪切形式,并且有的剪切形式具有性别表达特异性.对比 *sphinx* 基因

表 1 迄今已发现的年轻基因

基因名	年龄/Ma	所在种类	[文献]
<i>jingwei</i>	2.5	果蝇	[3]
<i>sphinx</i>	2~3	果蝇	[4]
<i>mkg</i>	1~2	果蝇	[5]
<i>Dntf-2r</i>	3~12	果蝇	[31]
<i>Sdic</i>	3	果蝇	[23]
<i>Cid</i>	3	果蝇	[53]
<i>Exuperantia1X</i>	<3	果蝇	[54]
<i>Finnegan</i>	20	果蝇	[55]
<i>POXP2</i>	0.1~0.2	灵长类	[56, 57]
<i>PmchL2</i>	5	灵长类	[58]
<i>PmchL1</i>	25	灵长类	[58]
<i>RNASE1B</i>	4.2	灵长类	[51]
<i>BC200</i>	35~55	灵长类	[28]
<i>PGAM3</i>	>25	灵长类	[30]
<i>Morpheus</i>	12~25	灵长类	[59]
<i>ECP</i>	31	灵长类	[60]
<i>CGβ</i>	34~50	灵长类	[61]
<i>Tre2</i>	21~33	灵长类	[62]
<i>FUT3/FUT6</i>	35	灵长类	[63]
<i>Arctic AFGP</i>	2.5	鱼	[64, 65]
<i>Antarctic AFGP</i>	5~14	鱼	[64, 65]
<i>4.5Si RNA</i>	25~55	鼠	[66]
<i>N-acetylneuraminate lyase</i>	<15	原生动物	[34]
<i>GD1</i>	0.24	拟南芥	[52]
<i>GD2</i>	0.5	拟南芥	[52]
<i>GD3</i>	1.2	拟南芥	[52]
<i>rps11</i>	<45	植物	[67]

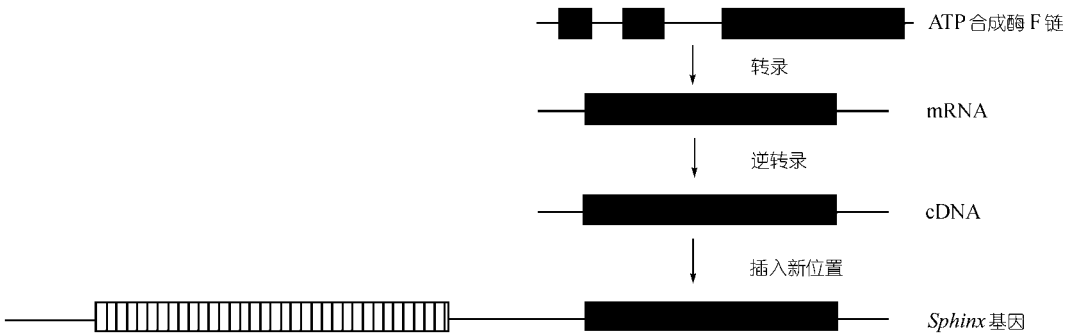


图 1 *Sphinx* 基因的形成
条纹框为新座位上原有的基因及其调控区域

的逆转座序列和 ATP 合成酶 F 链，发现其替换率显著高于中性序列。这些证据表明，*sphinx* 基因是有功能的 RNA 基因，并且其形成过程受正选择的驱动。

最近，我们又发现一个非常特别的年轻基因家

族——猴王基因家族(*mkg*)^[5]。在果蝇*Drosophila mauritiana*中，它在不到 2 百万年的时间里就产生了 3 个新的成员，这在进化的漫长时间尺度上无异于孙悟空拔毛变小猴一般神奇。更为难得的是，该基因家族

族第 1 次向我们展示了一个可观察的启动子产生速率和通过基因分裂产生新基因的进化过程。猴王基因家族的产生可分为 2 个阶段：在果蝇的 3 个近缘种 (*D. simulans*, *D. sechellia* 和 *D. mauritiana*) 分开前，祖先基因 (*mkgp*) 通过逆转座形成了一个新基因，并且新基因在 3 个种中分别形成了与祖先基因不同的启动子，其中 *sim-mkgr* 和 *sch-mkgr* 具有性别表达特异性；在 3 个物种分开后，*D. mauritiana* 中的 *mkgp* 又发生了一次逆转座，形成了另一个新基因 *mau-mkgr3*。而 *mau-mkgr3* 与 *mau-mkgr* 经过互补性的部分退化，分别继承了原始的 *mau-mkgr* 基因的 3' 和 5' 结构域的功能。这是第 1 次观察到的由互补性退化导致基因分裂而形成新基因的实例 (图 2)。

从上述基因的产生和进化过程，我们可以看出，一个新基因的产生是一个复杂的过程，常常综合了多种机制，如 *sphinx* 基因和猴王基因家族的产生就包括了逆转座、基因重复和外显子重排这些分子机制。而在这些基因中普遍检测到的快速进化，说明在新基因的进化过程中功能适应可能起着重要的作用。

4 总结与展望

年轻基因由于产生时间短，保留了大量进化过程中的重要信息，是研究新基因产生的理想材料。通过对目前已发现的年轻基因的分析，我们可以看出基因重复、外显子重排和逆转座等分子机制为新基因的产生提供了原材料，随后由于序列结构的改变导致新功能的产生，使生物体得以更好的适应环境，在正选择的驱动下这些新基因最终在群体中被固定下来。

从我们在果蝇中对年轻基因的研究^[3-5]，Lynch 和 Conery^[14] 对基因重复发生频率的估计，Patthy^[24]，Long^[25] 和 Li^[18] 等分析外显子重排对基因组贡献的研究，以及真核生物基因组大量存在的返座假基因和

可移动元件，我们可以看出，新基因的产生并不是一个稀有事件。生物进化的过程正是伴随着新基因不断产生的过程。

虽然对新基因的起源与进化的研究已经取得了一些成果，然而对整个基因组水平上新基因起源的规律，新基因产生对原基因组的影响 (如新基因与原基因间的相互作用和协同进化) 等方面还知之甚少，因此我们有必要去发现更多的年轻基因，总结其产生和进化的规律，并且应深入研究其功能，将基因结构的进化与功能的适应性联系起来，最终阐明新基因起源和进化的动力。

今天大规模的基因组测序工作仍在继续，不断有新的物种的基因组序列被公布。从这些庞大的数据中，我们将能找到大量关于新基因起源与进化的有用信息。然而针对近缘物种的大规模基因组测序短期内尚难以实现，为了能够快速地在整个基因组水平发现更多的年轻基因，经过周密的设计和几年艰苦的努力，我们已建立了一整套能够快速、有效地发现和研究新基因的研究体系。目前，我们实验室正以果蝇的 8 个近缘物种为实验材料来筛选其中的年轻基因。利用这套系统，我们已经确定了一定数量的候选年轻基因，并成功地获得了几个年轻基因。司芬克斯基因和猴王基因家族就是通过这套系统筛选出来的 2 个成功例子。此外，最近我们还利用该系统发现了 1 个年轻的有关细胞凋亡的基因家族。该家族也是在很短的时间内就产生了多个新的拷贝。根据我们现有的数据，这一基因家族纯粹是由基因重复然后功能分化而产生，它有望为研究基因重复这一最早被人们所提出的新基因产生机制和一个新基因出现后其所在功能通路的协同进化机制提供难得的材料。

在我们的研究计划完成之后，预期将有更多这

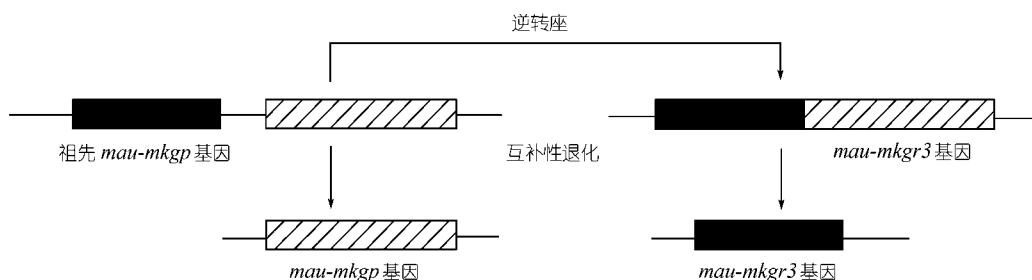


图 2 *mau-mkgr* 基因分裂示意图

黑框和斜纹框分别代表 2 个功能域

样的年轻基因被发现, 对这些基因起源与进化机制的研究, 将丰富我们关于新基因的起源与进化的知识. 并且我们期望这样的研究成果, 能够从基因组的水平初步探讨新基因起源的一般规律.

致谢 本工作受中德马普青年科学家小组经费、中国科学院生物局重要方向性项目(批准号: KSCX2-SW-121)和国家杰出青年科学基金的支持(批准号: 30325016)资助.

参 考 文 献

- Fraser C M, Gocayne J D, White O, et al. The minimal gene complement of *Mycoplasma genitalium*. *Science*, 1995, 270: 397~403
- Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860~921[DOI]
- Long M, Langley C H. Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science*, 1993, 260: 91~95
- Wang W, Brunet F G, Nevo E, et al. Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*. *Proc Natl Acad Sci USA*, 2002, 99: 448~4453
- Wang W, Yu H, Long M. Duplication-degeneration as a mechanism of gene fission and the origin of *Drosophila* new genes. *Nat Genet*, 2004, 36(5): 523~527[DOI]
- Haldane J B S. The cause of evolution. London: Longmans and Green, 1932
- Muller H J. The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. *Genetics*, 1935, 17: 237~252
- Ohno S. Evolution by Gene Duplication. German: Springer-Verlag, 1970
- Gilbert W. Why genes in pieces? *Nature*, 1978, 271: 501
- Gilbert W. The exon theory of genes. Cold Spring Harbor Symposium on Quantitative Biology, 1987, LII: 901~905
- Long M. Evolution of novel gene. *Curr Opin Genet Dev*, 2001, 11: 673~680[DOI]
- Betran E, Long M. Expansion of genome coding regions by acquisition of new genes. *Genetica*, 2002, 115: 65~80[DOI]
- Long M, Betran E, Thornton K, et al. The origin of new genes: glimpses from the young and old. *Nat Rev Genet*, 2003, 4(11): 865~75[DOI]
- Lynch M, Conery J S. The evolutionary fate and consequences of duplicate genes. *Science*, 2000, 290: 1151~1155[DOI]
- Blanc G, Barakat A, Guyot R, et al. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell*, 2000, 12: 1093~1101[DOI]
- Bubin G M. Comparative genomics of the eukaryotes. *Science*, 2000, 287: 2204~2215[DOI]
- Ball C A, Cherry J M. Genome comparisons highlight similarity and diversity within the eukaryotic kingdoms. *Curr Opin Chem Biol*, 2001, 5: 86~89[DOI]
- Li W H, Gu Z, Wang H, et al. Evolutionary analyses of the human genome. *Nature*, 2001 409: 847~849[DOI]
- Gu X, Wang Y F, Gu J Y. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat Genet*, 2002, 31: 205~209[DOI]
- Sudhof T C, Goldstein J L, Brown M S, et al. The LDL receptor gene: A mosaic of exons shared with different protein. *Science*, 1985, 228: 815~822
- Anke A F, Rijk V, Wilfried W, et al. Exon shuffling mimicked in cell culture. *Proc Natl Acad Sci USA*, 1999, 96: 8074~8079[DOI]
- Moran J V, Deberardinis R J, Kazazian H H. Exon shuffling by L1 retrotransposition. *Science*, 1999, 283: 1530~1534[DOI]
- Nurminsky D I, Nurminskaya M V, Aguiar D D, et al. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature*, 1998, 396: 572~575[DOI]
- Patthy L. Genome evolution and evolution of exon-shuffling—A review. *Gene*, 1999, 238: 103~114[DOI]
- Long M, Souza D S J, Gilbert W. Evolution of the intron-exon structure of eukaryotic genes. *Curr Opin Genet Dev*, 1995, 5: 774~778[DOI]
- Mccarrey J R. Nucleotide sequence of the promoter region of a tissue-specific human retroposon: Comparison with its housekeeping progenitor. *Gene*, 1987, 61: 291~298[DOI]
- Martignetti J A, Brosius J. Neural BC1 RNA as an evolutionary marker: Guinea pig remains a rodent. *Proc Natl Acad Sci USA*, 1993 90: 9698~9702
- Martignetti J A, Brosius J. BC200 RNA: A neural RNA polymerase product encoded by a monomeric Alu element. *Proc Natl Acad Sci USA*, 1993, 90: 11563~11567
- Brosius J. Retroposons-seeds of evolution. *Science*, 1993, 251: 753
- Betran E, Wang W, Jin L, et al. Evolution of the Phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. *Mol Biol Evol*, 2002, 19 (5): 654~663
- Betran E, Long M. *Dntf-2r*, a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics*, 2003, 164: 977~988
- Nekrutenko A, Li W H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet*, 2001, 17 (11): 619~621[DOI]
- Ochman H, Lawrence J G, Groisman E A. Lateral gene transfer and the nature of bacterial innovation. *Nature*, 2000, 405: 299~304[DOI]
- Koning D A P, Brinkman F S, Jones S J, et al. Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol Biol Evol*, 2000, 17: 1769~1773
- Kimura M. The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press, 1983

- bridge: Cambridge University Press, 1983
- 36 Fisher R A. The sheltering of lethals. *Am Nat*, 1935, 69: 446~455
- 37 Nei M. Accumulation of nonfunctional genes on sheltered chromosomes. *Am Nat*, 1970, 104: 311~322
- 38 Bailey G S, Poulter R T M, Stockwell P A. Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. *Proc Natl Acad Sci USA*, 1978, 75: 5575~5579
- 39 Kimura M, King J L. Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift. *Proc Natl Acad Sci USA*, 1979, 76: 2858~2861
- 40 Li W H. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics*, 1980, 95: 237~258
- 41 Ohta T. Simulating evolution by gene duplication. *Genetics*, 1987, 115: 207~213
- 42 Walsh J B. How often do duplicated genes evolve new functions? *Genetics*, 1995, 139: 421~428
- 43 Force A, Lynch M, Pickett F B, et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 1999, 151: 1531~1545
- 44 Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 2000, 154: 459~473
- 45 Walsh B. Population-genetic models of the fates of duplicate genes. *Genetica*, 2003, 118: 279~294[DOI]
- 46 Ohta T. Evolution by gene duplication revisited: differentiation of regulatory element versus protein. *Genetica*, 2003, 118: 209~216[DOI]
- 47 Gu X. Statistical methods for testing functional divergence after gene duplication. *Mol Biol Evol*, 1999, 16(12): 1664~1674
- 48 Gu X. Maximum-likelihood approach for gene family evolution under functional divergence. *Mol Biol Evol*, 2001, 18(4): 453~464
- 49 Wang Y F, Gu X. Functional divergence in the caspase gene family and altered functional constraints: Statistical analysis and prediction. *Genetics*, 2001, 158: 1311~1320
- 50 Gu X. Functional divergence in protein (family) sequence evolution. *Genetica*, 2003, 118: 133~141[DOI]
- 51 Zhang J, Zhang Y P, Rosenberg H F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating colobine monkey. *Nat Genet*, 2002, 30: 411~415
- 52 Moore R C, Purugganan M D. The early stages of duplicate gene evolution. *Proc Natl Acad Sci USA*, 2003, 100: 15682~15687[DOI]
- 53 Malik H S, Henikoff S. Adaptive evolution of *Cid*, a centromere-specific histone in *Drosophila*. *Genetics*, 2001, 157: 1293~1298
- 54 Yi S, Charlesworth B. A selective sweep associated with a recent gene transpositions in *Drosophila miranda*. *Genetics*, 2000, 156: 1753~1763
- 55 Begun D J. Origin and Evolution of a new gene descended from alcohol dehydrogenase in *Drosophila*. *Genetics*, 1997, 145: 375~382
- 56 Zhang J, Webb D M, Podlaha O. Accelerated protein evolution and origins of human-specific features: *Foxp2* as an example. *Genetics*, 2002, 162: 1825~1835
- 57 Enard W, Przeworski M, Fisher S E, et al. Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature*, 2002, 418: 869~872[DOI]
- 58 Courseaux A, Nahon J L. Birth of two chimeric genes in the *Hominidae* lineage. *Science*, 2001, 291: 1293~1297[DOI]
- 59 Johnson M E, Viggiano L, Bailey J A, et al. Positive selection of a gene family during the emergence of humans and African apes. *Nature*, 2001, 413: 514~519[DOI]
- 60 Zhang J, Rosenberg H F, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA*, 1998, 95: 3708~3713[DOI]
- 61 Maston G A, Ruvolo M. Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol Biol Evol*, 2002, 19 (3): 320~335
- 62 Paulding C A, Ruvolo M, Haber D A. The *Tre2(USP6)* oncogene is a hominoid-specific gene. *Proc Natl Acad Sci USA*, 2003, 100: 2507~2511[DOI]
- 63 Javaud C, Dupuy F, Mattah A, et al. The fucosyltransferase gene family: An amazing summary of the underlying mechanisms of gene evolution. *Genetica*, 2003, 118: 157~170[DOI]
- 64 Chen L, Devries A L, Cheng C H. Convergent evolution of anti-freeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc Natl Acad Sci USA*, 1997, 94: 3817~3822[DOI]
- 65 Chen L, Devries A L, Cheng C H. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc Natl Acad Sci USA*, 1997, 94: 3811~3816[DOI]
- 66 Gogolevskaya I K, Kramerov D A. Evolutionary history of 4.5S RNA and indication that it is functional. *J Mol Evol*, 2002, 54: 354~364
- 67 Bergthorsson U, Adams K L, Thomason B, et al. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, 2003, 424: 197~201[DOI]

(2004-03-22 收稿 2004-06-22 收修改稿)