

人类群体中的达尔文正选择

吴东东^{①③}, 张亚平^{①②*}

① 中国科学院昆明动物研究所遗传资源与进化国家重点实验室, 昆明 650223

② 云南大学生物资源保护与利用重点实验室, 昆明 650091

③ 中国科学院研究生院, 北京 100049

* 联系人, E-mail: zhangyp1@263.net.cn

2007-12-06 收稿, 2008-03-02 接受

国家自然科学基金(批准号: 30621092, 30430110)和云南省科技厅资助项目

摘要 概述了现代人群中存在的大量受达尔文正选择作用的基因, 如大脑发育相关基因、免疫基因、生殖相关基因及感觉受体基因等。这些基因的进化特性将会为人类进化、各种疾病机理等研究提供重要线索。随着群体遗传学数据以及比较基因组数据的增多, 越来越多的证据表明, 达尔文正选择作用在人类起源和进化上起着不可或缺的作用。本文还介绍了检测正选择作用的常用方法, 分析了所面临各种干扰因素, 并对进一步的研究作了展望。

关键词

自然选择

正选择

群体遗传学

比较基因组学

达尔文和华莱士于 1858 年提出的达尔文自然选择学说^[1], 作为 19 世纪最重要科学发现之一, 可以说是进化史上的里程碑; 其主要内容包括过度繁殖、生存斗争、遗传变异、适者生存等^[1]。随后虽然自然选择理论被广泛接受, 但是关于自然选择在生物进化历程中的作用地位一直争论不休。1925 年, 摩尔根提出了突变理论, 认为进化最重要的因素是有利突变的产生, 而自然选择只不过用来保存有利突变而删除有害突变。但是后来在实际实验中发现的突变绝大多数是有害突变^[2]。接下来突变理论逐渐被新达尔文主义所替代, 主要代表人物有 Fisher, Wright 和 Dobzhansky 等人。他们认为, 自然选择在群体遗传多态性的产生和维持过程中起主要作用, 所以新达尔文主义又被称为选择理论^[2]。20 世纪 60 年代, 蛋白质测序和电泳技术被广泛用于进化研究, 如在人和果蝇中发现大量遗传多态性、通过比较血色素和细胞色素的氨基酸序列发现在不同哺乳动物支系具有相近的氨基酸替代率等, 这些现象都是新达尔文理论无法解释的^[3]。针对这些问题, Kimura^[4]和 King 及 Jukes^[5]分别于 1968 和 1969 年提出著名的中性进化理论, 对经典的达尔文主义提出巨大挑战,

该理论认为, 大多数突变是有害的, 这些突变在群体中固定还是被删除是随机的, 即受随机遗传漂变决定, 而且突变速率与群体大小无关^[4,5]。在随后的 30 多年里, 自然选择在分子进化中的作用地位争议颇多, 而新达尔文理论和中性理论争论的焦点在于自然选择和遗传漂变在分子进化中的作用地位。非同义和同义替代速率差异是检测自然选择的主要依据, 基于此数据, 越来越多的发生在蛋白水平包括整体蛋白水平上的自然选择现象被揭示^[6]。例如, Smith 和 Eyre-Walker^[7]使用 McDonald-Kreitman 检验统计分析预测大约 45% 的氨基酸替代是自然选择作用的结果, 平均每 45 年就会有一个有利替代被固定; 另外一个研究表明, *Drosophila melanogaster* 和 *Drosophila simulans* 之间非同义替代率和同义替代率的比值为前者多态性比值的两倍^[8]。达尔文自然选择在分子进化中的重要性得到越来越多的认可, 因而, Ohta 对中性理论进行了修改, 提出了“近-中性”进化理论(near-neutrality)^[9~11], “突变-漂变-选择”三者分子进化中同时作用, 自然选择在进化中的作用得到提升。如今, 随着群体遗传学和比较基因组数据的增多, 大量基因组水平正选择的检测表明,

在种群进化历史中自然选择事件比以前想象的更多^[12~17]。关于自然选择在进化历程中具体起多大的作用仍难以准确估计。

正选择研究在近期得以迅速发展,促进了人类起源和进化的探讨,为各种疾病的机制的研究提供重要信息。本文将重点概述近期达尔文正选择在人群中的研究进展、所遇到的常见问题及解决方法。

1 人类进化过程中受正选择作用的基因

达尔文自然选择塑造并维持了现代人群中高度的表型多态性,包括外观、疾病抵抗力以及药物代谢能力的差异等,使其适应不同的生存环境。同时正是达尔文正选择作用成就了在很多方面较其他灵长类高等的人类,如认知能力的获得等。在与此相关的许多基因及其他各种免疫和生殖相关的基因等中都发现受到达尔文自然选择的作用。

1.1 大脑发育相关基因

人类和近亲黑猩猩基因有非常高的相似性,但表型差异甚多。认知是人类区别于其他高等灵长类动物的一个重要特征,而大脑正是这种能力的生理基础。大脑容量在灵长类中特别是人类进化系支上迅速增加,一些与大脑发育相关的基因受到正选择的作用而快速进化^[18~24],这也是正选择研究的一个热点。

小脑症基因(microcephalin)是调控人类大脑发育的关键基因之一。该基因突变失活以后,会引发人类原发性小脑症,病人的脑容量仅为正常人的四分之一到三分之一,相当于早期类人猿的脑容量。ASPM(abnormal spindle-like microcephaly-associated)是另一个小脑症关联基因,位于染色体1q31位置,含有28个外显子,全长65 kb,编码10434 bp长的编码序列。ASPM蛋白序列包含N-端microtubule结合区域,calponin同源区域,IQ重复区域以及一个C末端区域^[18~24]。基于基本覆盖世界的人类群体遗传多态性数据分析表明,这些在大脑的发育中起重要作用的基因在现代人群中仍受着正选择的作用,而且这种作用仍在继续^[18,24]。语言表达同样属于人类特有的能力。FOXP2基因,最近鉴定与人类语言表达以及鸟类的歌唱能力有关。灵长类种间数据分析表明,FOXP2基因在人类系统发育支上由于受到正选择作用而进化速率加快^[25~27];与前述的两个大脑发育基因相似,FOXP2基因在现代人群中也受到近期正选择的作用。

可以说,人类的智力是适应性进化的产物。垂体腺苷环化酶激活肽(pituitary adenylate cyclase-activating polypeptide, PACAP)是在神经系统中起重要作用的一种神经肽,介导大脑皮质的神经系统发育和调控,在染色体上所在区域与人群中的前脑无脑畸形症(holoprosencephaly)有相关性。功能上的重要性致使该肽在哺乳类动物中高度保守,但是系统发育分析表明,其前体蛋白UD和PRP区域在人类这一支上具有很高的正选择强度^[28]。这些与人类特异能力相关的基因的相关研究,如Clark等^[27]利用最大似然法从7645个直系同源基因中筛选出人类和黑猩猩之间特异的正选择基因,无疑为两物种间的差异以及人类进化研究提供了重要线索。

1.2 免疫系统相关基因

主要组织相容性复合物(major histocompatibility complex, MHC)座位是正选择基因的经典之例。主要组织相容性复合物主要是在免疫系统起抗原递呈的作用,起初在抗原结合区域发现具有较高的非同义替代率与同义替代率的比值(dN/dS)^[29]。该基因家族在物种间经历了“生和灭”的进化历程,具有很高的多态性,在这方面MHC同样受到平衡选择作用。宿主和外源病毒之间由于协同进化产生一个“军备竞赛”(arm race)的过程,病毒需要更新途径进入宿主而受到很强的选择作用,而宿主受到选择则是抵抗多变的外源物的入侵,这就是van Valen^[30]提出的Red Queen假说。许多抗病毒作用的基因如APOBEC3G^[31,32],CCR5^[33~38]等很好地见证了此假说。基因组水平的正选择检测也表明免疫系统经历着显著的正选择作用^[12,13,15,17]。

1.3 生殖相关基因

在灵长类中,许多与生殖相关的基因(如女性生殖蛋白:ZP2, ZP3, OGP;男性生殖蛋白:PRM1, PRM2)经历了快速适应性进化的过程,主要原因有精子间相互竞争(sperm competition)、性别冲突(sexual conflict)和隐密雌性选择(cryptic female choice)等^[39~41]。基因组水平正选择基因研究表明与生殖有关的基因类群受到近期的正选择作用,而且不同群体间的选择座位存在很大差异^[13]。

1.4 感觉受体相关基因

在长期进化过程中,动物能够在复杂的环境中选择对自身有利而避免有害的食物,味觉(甜、苦、咸、酸和辣)信号系统起着不可或缺的作用。哺乳动物的

苦味受体(bitter taster receptor, TAS2R或T2R)基因家族在适应性进化即正选择作用下经历了“生和灭”的进化过程,表现出高度的多态性^[42]。

TAS2R16 受体介导对一些自然界中有毒物质的敏感性。人群中K172N突变个体对这些物质的敏感性更高。通过对60个群体中997个个体分析表明,该多态性位点rs846664的衍生突变具有很高的频率,在人类走出非洲之前产生,伴随着人类迁移一直受到正选择的作用^[43]。PTC (TAS2R28), 两个分别具有感觉能力和没有感觉能力的单体型(hsA和hsG)在非洲、亚洲和欧洲人群中维持较高的频率,考虑到近期人类群体大小发生扩张的影响,通过群体扩张程度及时间的模拟,Wooding等人^[44]检测到PTC 基因受到平衡选择的作用,用以维持两个单型型的频率。关于人群中同时具有较高频率的有感觉PTC和没有感觉PTC能力的个体的原因并不清楚,可能是由于杂合优势^[44]。

同样,在哺乳动物中嗅觉受体基因超家族也经历了适应性进化和“生和灭”的过程。由于不同的生活环境,物种之间嗅觉受体基因数量及真假比例差异很大。在人和黑猩猩中均发现OR基因具有很低的多态性,含有大量低频率的等位基因。这种现象实际上在人群中是由于正选择的作用,而后者是由于受到净化选择作用,因为在黑猩猩中非同义替代速率显著小于同义替代速率^[45]。

随着大量基因组数据和多态性数据的产生,许多基因组水平的正选择基因检测结果表明正选择事件较以前想象的更多更频繁。比如,最近研究报道在非洲人群发现大量近期正选择事件^[13],揭示这些基因所受正选择作用的机制能够为人类起源进化,抗各种病毒的能力及机制提供重要信息,实际上在人群中很多遗传性疾病相关基因也受到正选择的作用。

1.5 受到正选择作用的遗传疾病相关基因

脊髓小脑性共济失调症(SCA) 2型(OMIM: 183090)主要病因是SCA-2 基因3'端蛋白编码区内的CAG重复序列存在扩增突变。研究表明,与(CAG)8CAA(CAG)4CAA(CAG)8 等位基因具有相关性的单体型在现代欧洲人群中受到近期的正选择作用^[46]。Duchenne型肌营养不良症(duchenne muscular dystrophy, DMD; OMIM: 300377)是一种以骨骼肌进行性变性、坏死为主要临床特征的致死性的X2连锁隐性遗传病,其发病机制由DMD突变导致。各种遗

传多态性数据特征,如较低的遗传多样性、种群间分化显著以及较长区域连锁不平衡等,都表明第7内含子受到正选择作用^[47]。OCA2, MYO5A, DTNBP1, TYRP1, SLC24A5等都是白化病相关性基因,基因组水平扫描发现它们在欧洲人群中都受到正选择的作用^[13]。其他受正选择作用的遗传疾病相关基因,如BRCA1^[48,49], SRY^[50]等。虽然这些遗传疾病相关基因受到正选择作用的机制目前并不清楚,但进一步的研究将会为疾病发病机理和治疗的研究提供重要线索。

疟疾(malaria)是儿童死亡最重要因素之一,在非洲每年超过100万的儿童死于疟疾^[51]。与抗疟疾有关的基因很多受到正选择的作用(具体与抗疟疾有关的基因可参考文献[51])。编码β-球蛋白的HBB基因是最早鉴定的正选择基因,之后又通过各种统计学检测方法鉴定HBB各种等位基因在不同群体内的选择特性^[14]。葡萄糖-6-磷酸脱氢酶基因G6PD的一个变异体202A所在的个体具有抗疟疾感染的能力^[52], CD40L (TNFSF5)基因编码免疫系统中一个重要蛋白,启动子区域的一个变异体和抗疟疾感染具有相关性^[53]。长区域单体型检验表明这两个基因近期在人群中都受到正选择的作用。GYPA是一种编码细胞表面糖蛋白的基因,同样各种方法检验在人群受到正选择和平衡选择的作用^[54]。Duffy血型座位(FY)一直都被认为可能是受自然选择作用的座位,这是因为FY*A, FY*B和FY*O三个等位基因独特的地理分布分化^[55,56]。例如, FY*O在撒哈拉以南非洲(sub-Saharan African)人群中已固定;而在中国人群只存在FY*A等位基因; FY*B是个祖先状态,在大多数人群中已消失。群体多态性特征也表明, FY*O和A受到正选择的作用^[56,57],如非洲人群FY*O的多样性非常低,而且衍生等位基因的频率也很高。

为什么突变会产生不利的遗传性疾病的基因仍然受到正选择作用?表面上很矛盾,实际上这与基因多效性(gene pleiotropy)有关,突变在一种功能上是有利的,但在另一功能上是有害的^[58]。例如,上述在非洲发现抗疟疾的有利突变在另一方面又会造成镰刀型贫血症。另一原因是快速变化的环境所引起,就是说一种突变在历史上是有利的,但近期环境的变化会使这种突变变得有害。多巴胺受体D4 (dopamine receptor D4, DRD4)的遗传多样性与人类文化结构变化之间的关系就是个很好的例子^[59]。该基因在

第3号外显子内具有特征性48 bp数目可变性串联重复序列(variable number tandem repeat, VNTR)多态性,从2次重复(2R)到11次重复(11R)不等,其中7R与注意缺损多动障碍(attention deficient hyperactivity disorder, ADHD)有关联性。但Ding等^[59]基于群体遗传学证据认为7R在人类进化近期受到正选择的作用而其频率升高。考虑到7R与运动能力有关,7R的分布与人类迁移有相关性。同时早期7R携带者在雄性竞争的社会里,具有选择配偶、获取食物以及照顾下一代的优势。近期社会,人类迁移基本停止,体力运动能力的重要性下降,进而可能导致遗传病状^[60,61]。所以,从另一方面来说,基因的选择性研究会给其相关遗传性疾病机制的研究提供重要线索。

2 基因组水平正选择作用的检测

随着各种比较基因组数据,即人、黑猩猩、小家鼠、狗等基因组测序的接近完成,以及遗传多样性数据SNP数据库的公布,基因组水平正选择检测研究大量出现。基因组水平正选择检测的最大优势在于不受群体历史因素的影响,但很多是基于SNP数据库中的数据,因而受数据测量偏倚影响较大,即使各自使用不同的校正方法。比较基因组分析由于物种数很少,使用的最大似然方法检测正选择基因的假阳性一般较高。此外,不同报道的选择基因存在差异,这可能是由于不同的方法适宜检测不同时期的正选择事件,且使用的数据也不同。

3 正选择检测方法

基于所需数据的差异,检测方法可以分为3类,即基于种内多态性数据的方法、种间分歧数据及种内多态性和种间分歧数据。下面分别简单介绍几个常用的检测方法。

3.1 基于种内多态性数据的方法

自然选择可简单地分为3种,即正选择、负选择(或净化选择)和平衡选择。简单的说,在正选择条件下,有利突变能够快速在群体中固定,负选择是有害突变在群体中被删除的过程,平衡选择则能维持群体中两个或多个有利等位基因。正选择不仅影响受作用的基因或区域自身的遗传特性,而且与其邻近的区域因连锁也会具有相似的多态性,此作用即为基因搭乘(genetic hitch-hiking),又称为选择性清除(selective sweep)。具体地说,基因搭乘是指受正选择

作用的等位基因频率升高,会带动旁边连锁的等位基因频率的升高,选择性清除是指受选择的座位遗传多样性降低,周围的遗传多样性也会降低。实际上这两个概念所指的是同一个遗传现象,所以在实际中并不将其区分。这些就是正选择作用对等位基因频率图谱(frequency spectrum)分布的影响。频率图谱,简单地说,就是各频率等位基因占总等位基因数的比例。受正选择的等位基因同样具有较长的单体型区域,连锁不平衡强度较中性选择区域高,而且受正选择的座位的衍生等位基因的频率较高。关于自然选择作用产生的群体遗传多态性特征可参考文献^[16,62]。这些偏离中性进化的印记正是各种检测自然选择方法的凭据。

(i) 基于频率图谱的方法。Ewens抽样公式^[63]可以说是群体遗传学上的一块奠基石,基于此公式Watterson提出著名的Ewens-Watterson检验^[64]。通过多个座位的平均杂合度可以计算单个座位杂合度的理论分布作为无效模型,与观察到的单座位杂合度相比较即可检测是否偏离中性进化^[3]。这些检测只用于无限等位基因模型数据,如allozyme。同样很多方法是针对无限位点模型数据如核苷酸数据提出的。Tajima's D 检测方法^[65]在概念上与上述Ewens-Watterson蛋白质多态性杂合度的检测是相似的,就是比较分离位点数和核苷酸多样性之间的关系。在中性进化条件下两者是相同的,当两者之间差异显著时表明偏离中性选择。正选择和负选择作用都会产生负 D 值。当基因或区域的 $D > 0$,可能受到平衡选择的作用^[62,66]。比如PTC基因在亚洲人群和欧洲人群中Tajima's D 观察值分别为2.94 ($P(D>0) = 0.01$), 2.91 ($P(D>0) = 0.01$)^[44]。实际上 D 值并非符合正态分布,而是与 β 分布较接近,但使用 β 曲线模拟 D 值分布进行检验时也会产生误差^[3]。

溯祖理论可以动态追踪序列在群体中的进化历程,根据在系统发育树中的分布位置,突变可分为外缘和内部突变,前者是近期产生的突变。在正选择或背景选择作用下外缘突变相对于内部会增多,基于这种差异,Fu等^[67-69]提出一系列相似的检测方法,但实际中针对不同的作用,各种方法检测能力不同。相对于前几种方法, F_s 适合群体增长或正选择产生的中性偏离的检验^[63],而Fu等的 D 和 F 检验更适合用于背景选择的检测。但是这些基于遗传多态性变异的正选择检测方法受群体数量增长影响较大。

Fay和Wu^[70]的 H 检验方法主要是基于基因搭乘会产生高频率的新的衍生等位基因,但是背景选择不会。这需要有外群,一般选择猩猩做外群。假设每个位点只会发生一次突变,最大似然或最大简约法等推算人和黑猩猩的祖先序列,而群体中其他不同的即为衍生等位基因^[70]。例如,人苦味受体TAS2R16基因座位的3个SNP: rs2692396, rs846664与rs1204014具有非常高的衍生位点频率。Fay和Wu的 H 值显著性小于0,表明此苦味受体基因发生了正选择事件^[43]。与前面几种方法相比,群体分化对 H 检验干扰较大,但受群体扩张因素影响较小^[70]。在正选择作用下,高频率的衍生等位基因会在很短时间内在群体固定,适宜检测小于约80000年前等位或位点的正选择事件^[14]。

(ii) 群体分化。1973年Lewontin和Krakauer^[71]提出首个基于群体内正选择能够加快基因在群体间的分化检测自然选择的方法。但是群体历史和结构对该方法干扰非常大,因而一段时间后该检测不再被使用。但是群体历史能够影响整个基因组水平的多态性,正选择作用只作用单个基因或区域,因而通过采集大量基因组水平的数据,受到选择的座位就会偏离其他中性分布的座位,这就是最近基于异常检验方法(outlier approach),利用群体分化参数 F_{st} 检测正选择座位的原理。Akey等^[72]就是利用这种方法基于基因组水平SNP的 F_{st} 经验分布,检测到174个候选正选择基因,这也是第一张基于SNP数据的人类正选择基因图谱。还有些例子如白介素受体IL4在免疫系统细胞介素信号通路中起着重要作用,其后动子区域多态性位点(rs2243250)在很多群体间就有显著性群体分化,成为正选择作用的标志,调节IL4表达:该位点为T的IL4的表达量比C高3倍^[73]。

(iii) 基于连锁不平衡和单体型结构的检测。连锁不平衡,作为两个座位间的相关性,强度可以被重组降低,同时单体型结构也会被破坏。但受正选择的座位或区域,有利突变会快速带动周围区域在群体固定以致重组来不及破坏,最终拥有较长区域的连锁不平衡和单体型结构,此长度与正选择的强度具有正相关性^[74]。因此单体型长度和相似性,以及连锁不平衡的长度、衰退速度等可以用于检测遗传标记是否偏离中性进化。相对于基于等位基因频率的检测方法,单体型分析对于正选择的检测更有力度,但后者对重组速率比较敏感,由于单型型的长度维

持时间非常短,相关的检验主要用于群体近期如人类不同群体特异的正选择作用。

长区域单体型(long range haplotype, LRH)检测所用的参数是扩展性单体型同质性(extended haplotype homozygosity, EHH),是指在含有中心单体型(core haplotype)样本中随机抽取到两个相同扩展性单型型的概率^[74]。人类基因组中,重组速率是高度异质性的,而且其变异非常迅速。因此可靠且令人信服的统计检验必须考虑此现象,LRH检验的另一个参数rEHH(relative EHH)就考虑到此问题。LRH方法对于一些较低等位基因或单体型频率(~10%)的正选择检测非常有用,能够精确到单个基因,而且受数据的测量偏倚(ascertainment bias)的影响较小^[14]。Sabeti等人^[33]使用此方法检验了CCR5-Δ32基因的选择性。CCR5是介导HIV感染的一种趋化因子受体,该基因由于受到正选择作用在人群有大量非同义突变发生。其中编码框中缺失了32bp的等位基因(CCR5-Δ32)在人群中有很高频率,该突变的纯合体具有抗HIV感染性,杂合体也能延迟感染^[33]。另一个典型例子是LCT座位,在欧洲人群中有一个单体型结构延伸至1Mb^[13]。与LRT检测相似的还有单体型相似度(haplotype similarity, HS)检测^[75]、iHS (integrated EHH)检测^[13]以及连锁不平衡的衰退(linkage disequilibrium decay)检测^[12]等。

3.2 基于种间分歧数据的检测

(i) 基因水平上的正选择检验。两条编码序列之间的非同义替代速率(nonsynonymous substitution rate, dN)与同义替代率(synonymous substitution rate, dS)比值一直被用于检验基因在种间的选择特性($\omega = dN/dS$, $\omega > 1$ 分别为正选择,中性选择和负选择)。但由于蛋白质不同位点功能存在差异,受到选择强度也会不同,同样基因在不同进化时期,所受选择也会有差异。针对这些原因,Yang等人^[76,77]基于最大似然检验建立了位点特异(site-specific model)、支特异(Branch-specific model)^[78,79]和支-位点模型(branch-site model)^[80]。这些模型原理都是基于构建嵌合式模型:一个是无效模型(所有位点 ω 均小于1),另一常规模型(允许有 $\omega > 1$ 的位点)与其对应,通过似然比检测(likelihood ratio test, LRT)检测两模型似然值对数的差异,如果卡方检验显著,则无效模型被拒绝。“位点特异”模型通过贝叶斯方法识别可能受到正选择的位点,其中贝叶斯方法有两种:NEB (native em-

pirical Bayes)^[76,81]和BEB (Bayes empirical Bayes)^[82],前者产生的假阳性非常高,推荐使用后者^[83]. 数据模拟以及实际数据分析表明,很多模型检测精确度和效率很低,现在普遍使用的嵌合模型是M1a (Nearly Neutral)和M2a (Positive Selection), M7 (β)和M8 (β & ω). M7中 ω 符合 β 分布(p, q), ω 值在(0,1)间,因而无正选择位点,为无效模型, M8模型中有 p_0 的位点符合 β 分布(p, q),而另一部分 $p_1=1-p_0$ 是正选择位点($\omega > 1$),是常规模型. 许多实际数据分析表明, M7对M8在检测正选择位点的能力不如M1a对M2a,前者很容易产生假阳性数据. 虽然此方法是检测正选择位点的有力工具,但此方法主要问题是其产生的假阳性. Anisimova等^[84,85]通过模拟给出下列建议: (1)序列相似性较高或序列数小(如, $S \leq 0.11$ 或 $T \leq 6$)时,正选择位点预测可靠性低;同样序列分歧度较大时会降低检测的精确度,序列的密码子数小于100时检测效率会降低; (2)增加序列数是增高精确度的最有效办法; (3)同时使用多个模型. “支-特异”模型,可以比较计算不同支上的 ω 值,因为基因在不同时间段可能所受选择压力不同,比如在重复发生后,通过增强正选择作用(也可能是净化作用放松),产生新的功能,但当新的功能形成并稳定后,就不允许产生非同义突变,此时和其他基因一样受净化选择作用, $\omega < 1$. “支-位点”模型包括模型A和B,它们将系统发育支分为前景支 (foreground lineage) 和背景支 (background lineage),其中前景支是我们所检测正选择的支,允许有正选择支和位点, background为其他支,不允许有正选择位点和枝. 将Model A与M1a (Nearly Neutral)进行似然比检测LRT (自由度 $df = 2$),此方法为test1. 也可将Model A与其无效模型(将潜在正选择位点的 ω 值设为1)比较,此为test2. 计算机模拟发现test1在检测正选择时假阳性非常高(20% ~ 70%),而且不能区分正选择和选择压力的放松,推荐使用test2^[83]. 同时BEB再计算后验概率时虽然可靠性比较高(当显著性为95%时,假阳性小于5%),但此方法较保守,不易检验出正选择位点^[83]. dN/dS检测方法必须解决的问题是区分正选择和选择压力的放松. 一般情况下,基因受到正选择作用时, dN升高,但dS不变;在选择压力放松时, dN和dS都会升高. 在很多情况下,正选择强度非常低,即便似然值检测也无法检测到正选择事件,这也是dN/dS无法解决的问题,这就期待更加灵敏的检测方法有效地区分蛋白

序列中正选择和负选择作用.

dN/dS在很多情况下用来衡量蛋白质的进化速率. 例如, Wang等^[86]发现人类大脑表达基因进化速率并不比其他灵长类动物高,揭示人类大脑快速进化很可能只是几个关键基因起作用.

(ii) 蛋白质水平上的正选择检验. 从另一方面说,上述对于所有氨基酸改变同等对待基因水平上的正选择检验误差比较大. 因此出现了基于氨基酸理化性质的统计学模型,即由系统发育树推算氨基酸改变的分布与氨基酸改变的中性期望分布相比较,检测可能与中性偏离的位点^[87,88]. TreeSAAP就是处理这方面问题的软件^[89].

3.3 基于种内多态性和种间分歧数据

著名的McDonald-Kreitman检验利用的就是种内多态性及种间分歧数据^[90]. 它将核苷酸位点分为两种: 固定位点 (fixed sites) 和多态性位点 (polymorphism sites). 在固定位点上,物种A所有序列都是一种核苷酸,而另一个物种都是另一种核苷酸;其他所有变化位点均为多态性位点. 这两种位点又可进一步分为同义替代和非同义替代位点,因而对于这4种位点可以用 2×2 列联表检验^[90]. 这种方法不受群体历史、重组等因素的影响,但却只能检测编码区序列. HKA检验是由Hudson等^[91]基于高速率的遗传座位比低速率的具有更高度的多态性建立的检验中性进化的统计学方法. 但是此检测方法要求对DNA序列进行长期观察,而且要求整个进化过程中有效群体大小恒定不变,这些条件常常不能满足,因而HKA检验并不常用^[3,91].

各种检测方法的理论基础差异很大,因而针对相同的数据不同的方法经常会得到不同的结果. 正是因为不同的原理,各种方法所适宜检测的正选择时间不同. 如McDonald-Kreitman方法基于种间和种内数据来比较编码区序列的非同义和同义替代数差异,适用于检测几百万年前的正选择事件^[14]. 事实上没有哪个单独的检测方法能够针对所有偏离中性的遗传学多态性数据.

4 各种因素对正选择检测的影响

Wright-Fisher群体模型作为标准中性模型,是个突变对生存率无影响、随机交配且群体大小恒定的群体. 因此以此作为无效模型检测选择作用肯定会受到影响这些参数的因素的影响. 比如群体大小扩张

或减少,重组速率的差异以及数据测量偏倚等都会产生干扰。毋庸置疑,排除这些干扰因素是鉴定受选择基因座位或区域的必须之举。

4.1 群体历史(demographic history)

实际上,受群体历史包括瓶颈效应、群体扩张、群体分化、群体间融合等的影响,都能形成类似于选择作用导致的群体遗传多态性特点。例如,群体数量的扩张和正选择作用同样可以产生很多低频率的等位基因;群体中维持两个或多个高频率的单体型有可能是群体再分(population subdivision)产生的结果,而不是平衡选择所致。因而在检测自然选择作用时必须考虑群体历史因素的影响。

ASPM和Microcephalin是大脑发育的重要调节因子,系统发育和最大似然分析表明,在灵长类中特别是人类和猿受到很强的自然选择作用。在现代人群中,这两个基因含有很高频率的单体型(ASPM单体型 63 的频率为 21%; Microcephalin 单体型 49 的频率是 33%),这可能是持续正选择作用的结果^[18,24]。但Currat等人^[92]通过溯祖模拟认为人类群体数量扩张后群体分化同样可以产生相似的结果。当然他们所用的参数是否符合人类真实历史还需要斟酌,而且他们的模拟未考虑重组速率的影响^[93]。但是群体历史与自然选择作用明显的差别在于,群体历史对于整个基因组都会产生影响,而后者只作用于单个基因或区域。在这方面最近大量基因组水平检测正选择作用就会显出优势。同时群体历史不会对种间分歧数据产生影响,因此McDonald-Kreitman 检测正选择时不受各种群体历史因素影响。当然排除群体历史因素的影响同样需要精确可靠的人类群体历史事件。

4.2 基因重组速率(recombination rate)

在人类基因组中将近一半的核苷酸多态性是重组作用的结果,如在染色体着丝粒和端粒等低重组区多态性较低^[94,95]。经验也表明,检测到的正选择基因主要分布在低重组速率区域。很多方法都考虑到此问题,因此高精度重组速率结构图对于自然选择的检测是必须的^[94,95]。实际上统计学上根据群体多态性数据推算群体历史,基因重组,选择作用时,三者之间都会互相影响。比如同样道理,正选择会对基因重组热点区域(hot-spot)的检测造成很高的假阳性^[96]。

4.3 数据测量偏倚(ascertainment bias)

这是各种 SNP 数据库存在的主要问题,与 SNP

测定程序有关,首先在数量很少的几个个体中检测到 SNP 位点,在扩展到大样本中测量它们的频率,这样就会遗漏很多低频率的 SNP,而产生大量中等频率的 SNP。数据测量偏倚对上述的群体历史,重组速率分布和自然选择的检验都会产生影响,具体 SNP 测量偏倚和校正的数学理论可参考文献^[97]。但是理论处理误差仍然很大,在实际检测单个基因的自然选择事件时,数据库中的 SNP 只能作为参考,具体数据需要选择尽量覆盖一个群体的大样本,针对每个个体的基因整个区域都测序,这样就可获取无偏倚的 SNP 数据。

5 展望

正选择的研究仍然处在初始阶段,但此领域却在近期迅速发展,即使如今检测到大量的正选择基因,但这些仍然只是部分数据。研究也逐渐从比较基因组数据转向群体多态性数据,这可能是人们更加关心人类近期的进化动向。除了寻找前述大脑、免疫以及生殖等相关的基因,现代人群展现出的巨大表型差异会成为以后研究的重要目标。具体地说在现在人群不同群体之间存在很大的社会文化结构,特别是外观、饮食习惯等差异,这方面有皮肤黑色素表达、头发相关的基因,糖类、氨基酸代谢基因等。近期研究发现基因组中大量非编码序列具有重要功能,如顺式调控位点^[98]、microRNA^[99]等。比如在许多神经和营养相关基因的启动子区域发现在人类进化支上受到正选择作用^[98],但非编码序列的自然选择研究很少,有待于进一步的工作。在数据上,需要无偏倚的数据,当然这是理论上的追求,在实际中能做的就是样本大、覆盖面广等。同时各种基因组水平正选择的检测受到数据测量偏倚因素的影响较大,需要进一步的验证。近期基因组技术如测序、SNP分型等得以迅速发展,大批量数据的获得不再是该研究的瓶颈,因而复杂性状相关的多基因研究的现实可行性就很高,这也会成为研究的重点。在方法上,需要更加精确的统计学检测方法来解决各种因素的干扰,如群体历史、重组速率等,但此方面已近完善。如今正选择研究的难点也就是如何将选择作用与基因具体的功能联系起来,如何指导进一步的功能研究,这也是正选择研究的最终目的。如今通过正选择与疾病关联性相结合来研究基因的进化和相应功能应该是可行之举。

参考文献

- 1 Darwin C, Wallace A R. On the tendency of the species to form varieties and on the perpetuation of the species by natural means of selection. *J Proc Linnaean Soc London (Zoology)*, 1858, 3: 45—62
- 2 Nei M. Selectionism and neutralism in molecular evolution. *Mol Biol Evol*, 2005, 22: 2318—2342 [\[DOI\]](#)
- 3 Nei M, Kumar S. *Molecular evolution and phlogenetics*. New York: Oxford University, 2000
- 4 Kimura M. Evolutionary rate at the molecular level. *Nature*, 1968, 217: 624—626
- 5 King J L, Jukes T H. Non-Darwinian evolution. *Science*, 1969, 164: 788—798
- 6 Eyre-Walker A. The genomic rate of adaptive evolution. *Trends Ecol Evol*, 2006, 21(10): 569—575 [\[DOI\]](#)
- 7 Smith N G, Eyre-Walker A. Adaptive protein evolution in *Drosophila*. *Nature*, 2002, 415: 1022—1024 [\[DOI\]](#)
- 8 Fay J C, Wyckoff G J, Wu C I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature*, 2002, 415: 1024—1026 [\[DOI\]](#)
- 9 Ohta T. Near-neutrality in evolution of genes and gene regulation. *Proc Natl Acad Sci USA*, 2002, 99: 16134—16137 [\[DOI\]](#)
- 10 Ohta T. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst*, 1992, 23: 263—286 [\[DOI\]](#)
- 11 Ohta T. Population size and rate of evolution. *J Mol Evol*, 1972, 1: 305—314
- 12 Wang E T, Kodama G, Baldi P, et al. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc Natl Acad Sci USA*, 2006, 103: 135—140 [\[DOI\]](#)
- 13 Voight B F, Kudaravalli S, Wen X, et al. A map of recent positive selection in the human genome. *PLoS Biol*, 2006, 4(3): e72 [\[DOI\]](#)
- 14 Sabeti P C, Schaffner S F, Fry B, et al. Positive natural selection in the human lineage. *Science*, 2006, 312: 1614—1620 [\[DOI\]](#)
- 15 Nielsen R, Bustamante C, Clark A G, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*, 2005, 3: e170 [\[DOI\]](#)
- 16 Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*, 2005, 39: 197—218 [\[DOI\]](#)
- 17 Bustamante C D, Fledel-Alon A, Williamson S, et al. Natural selection on protein-coding genes in the human genome. *Nature*, 2005, 437: 1153—1157 [\[DOI\]](#)
- 18 Mekel-Bobrov N, Gilbert S L, Evans P D, et al. Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science*, 2005, 309: 1720—1722 [\[DOI\]](#)
- 19 Evans P D, Anderson J R, Vallender E J, et al. Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum Mol Genet*, 2004, 13: 489—494 [\[DOI\]](#)
- 20 Zhang J. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics*, 2003, 165: 2063—2070
- 21 Kouprina N, Pavlicek A, Mochida G H, et al. Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLoS Biol*, 2004, 2: e126 [\[DOI\]](#)
- 22 Wang Y, Su B. Molecular evolution of microcephalin, a gene determining human brain size. *Hum Mol Genet*, 2004, 13: 1131—1137 [\[DOI\]](#)
- 23 Evans P D, Anderson J R, Vallender E J, et al. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum Mol Genet*, 2004, 13: 1139—1145 [\[DOI\]](#)
- 24 Evans P D, Gilbert S L, Mekel-Bobrov N, et al. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science*, 2005, 309: 1717—1720 [\[DOI\]](#)
- 25 Zhang J, Webb D M, Podlaha O. Accelerated protein evolution and origins of human-specific features FOXP2 as an example. *Genetics*, 2002, 162: 1825—18352
- 26 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 [\[DOI\]](#)
- 27 Clark A G, Gnanowski S, Nielsen R, et al. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science*, 2003, 302: 1960—1963 [\[DOI\]](#)
- 28 Wang Y Q, Qian Y P, Yang S, et al. Accelerated evolution of the pituitary adenylate cyclase-activating polypeptide precursor gene during human origin. *Genetics*, 2005, 170: 801—806 [\[DOI\]](#)
- 29 Hughes A L, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 1988, 335: 167—170 [\[DOI\]](#)
- 30 van Valen L. A new evolutionary law. *Evol Theory*, 1973, 1: 1—30
- 31 Zhang J, Webb D M. Rapid evolution of primate antiviral enzyme APOBEC3G. *Hum Mol Genet*, 2004, 13: 1785—1791 [\[DOI\]](#)
- 32 Sawyer S L, Emerman M, Malik H S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol*, 2004, 2: e275 [\[DOI\]](#)
- 33 Sabeti P C, Walsh E, Schaffner S F, et al. The case for selection at CCR5-Δ32. *PLoS Biol*, 2005, 3: e378 [\[DOI\]](#)
- 34 Bamshad M J, Mummidi S, Gonzalez E, et al. A strong signature of balancing selection in the 5' *cis*-regulatory region of CCR5. *Proc Natl Acad Sci USA*, 2002, 99: 10539—10544 [\[DOI\]](#)

- 35 Maayan S, Zhang L, Shinar E, et al. Evidence for recent selection of the CCR5-Δ32 deletion from differences in its frequency between Ashkenazi and Sephardi Jews. *Genes Immun*, 2000, 1: 358—361 [\[DOI\]](#)
- 36 Wooding S, Stone A C, Dunn D M, et al. Contrasting effects of natural selection on human and chimpanzee CC chemokine receptor 5. *Am J Hum Genet*, 2005, 76: 291—301 [\[DOI\]](#)
- 37 Zhang Y W, Ryder O A, Zhang Y P. Intra- and interspecific variation of the *CCR5* gene in higher primates. *Mol Biol Evol*, 2003, 20(10): 1722—1729 [\[DOI\]](#)
- 38 Li H P, Zhang Y W, Zhang Y P, et al. Neutrality tests using DNA polymorphism from multiple samples. *Genetics*, 2003, 163: 1147—1151
- 39 Swanson W J, Yang Z, Wolfner M F, et al. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci USA*, 2001, 98: 2509—2514 [\[DOI\]](#)
- 40 Swanson W J, Vacquier V D. The rapid evolution of reproductive proteins. *Nat Rev Genet*, 2002, 3: 137—144 [\[DOI\]](#)
- 41 Wyckoff G J, Wang W, Wu C I. Rapid evolution of male reproductive genes in the descent of man. *Nature*, 2000, 403: 261—263 [\[DOI\]](#)
- 42 Shi P, Zhang J, Yang H, et al. Adaptive diversification of bitter taste receptor genes in mammalian evolution. *Mol Biol Evol*, 2003, 20: 805—814 [\[DOI\]](#)
- 43 Soranzo N, Bufer B, Sabeti P C, et al. Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. *Curr Biol*, 2005, 15: 1257—1265 [\[DOI\]](#)
- 44 Wooding S, Kim U K, Bamshad M J, et al. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *Am J Hum Genet*, 2004, 74: 637—646 [\[DOI\]](#)
- 45 Gilad Y, Bustamante C D, Lancet D, et al. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet*, 2003, 73: 489—501 [\[DOI\]](#)
- 46 Yu F, Sabeti P C, Hardenbol P, et al. Positive selection of a pre-expansion CAG repeat of the human *SCA2* gene. *PLoS Genet*, 2005, 1(3): e41 [\[DOI\]](#)
- 47 Nachman M W, Crowell S L. Contrasting evolutionary histories of two introns of the duchenne muscular dystrophy gene, *Dmd*, in humans. *Genetics*, 2000, 155: 1855—1864
- 48 Huttley G A, Eastal S, Southey M C, et al. Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. *Nat Genet*, 2000, 25: 410—413 [\[DOI\]](#)
- 49 Fleming M A, Potter J D, Ramirez C J, et al. Understanding missense mutations in the BRCA1 gene: An evolutionary approach. *Proc Natl Acad Sci USA*, 2003, 100: 1151—1156 [\[DOI\]](#)
- 50 Wang X, Zhang J, Zhang Y P. Erratic evolution of SRY in higher primates. *Mol Biol Evol*, 2002, 19: 582—584
- 51 Kwiatkowski D P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet*, 2005, 77: 171—192 [\[DOI\]](#)
- 52 Ruwende C, Hill A. Glucose-6-phosphate dehydrogenase deficiency and malaria. *J Mol Med*, 1998, 76: 581—588 [\[DOI\]](#)
- 53 Sabeti P C, Usen S, Farhadian S, et al. CD40L association with protection from severe malaria. *Genes Immun*, 2002, 3: 286—291 [\[DOI\]](#)
- 54 Baum J, Ward R H, Conway D J. Natural selection on the erythrocyte surface. *Mol Biol Evol*, 2002, 19: 223—229
- 55 Zimmerman P A, Woolley I, Masinde G L, et al. Emergence of FY*A null in a plasmodium vivax-endemic region of Papua New Guinea. *Proc Natl Acad Sci USA*, 1999, 96: 13973—13977 [\[DOI\]](#)
- 56 Hamblin M T, Thompson E E, Rienzo A D. Complex signatures of natural selection at the duffy blood group locus. *Am J Hum Genet*, 2002, 70: 369—383 [\[DOI\]](#)
- 57 Hamblin M T, Rienzo A D. Detection of the signature of natural selection in humans: Evidence from the duffy blood group locus. *Am J Hum Genet*, 2000, 66: 1669—1679 [\[DOI\]](#)
- 58 Clark N L, Swanson W J. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet*, 2005 1(3): e35 [\[DOI\]](#)
- 59 Ding Y C, Chi H C, Grady D L, et al. Evidence of positive selection acting at the human dopamine receptor *D4* gene locus. *Proc Natl Acad Sci USA*, 2002, 99(1): 309—314 [\[DOI\]](#)
- 60 Harpending H, Gregory C. In our genes. *Proc Natl Acad Sci USA*, 2002, 99(1): 10—12 [\[DOI\]](#)
- 61 Hughes A L. Strength in numbers. *Nature*, 2002, 417: 795 [\[DOI\]](#)
- 62 Bamshad M, Wooding S. Signature of natural selection in the human genome. *Nat Rev Genet*, 2003, 4: 99—111 [\[DOI\]](#)
- 63 Ewens W J. The sampling theory of selectively neutral alleles. *Theor Popul Biol*, 1972, 3: 87—112
- 64 Watterson G A. Heterosis or neutrality? *Genetics*, 1977, 85: 789—814
- 65 Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 1989, 123: 585—595
- 66 Nielsen R, Williamson S, Kim Y, et al. Genomic scans for selective sweeps using SNP data. *Genome Res*, 2005, 15: 1566—1575 [\[DOI\]](#)
- 67 Fu Y X, Li W H. Statistical tests of neutrality of mutations. *Genetics*, 1993, 133: 693—709
- 68 Fu Y X. New statistical tests of neutrality for DNA samples from a population. *Genetics*, 1996, 143: 557—570
- 69 Fu Y X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 1997, 147: 915—925

- 70 Fay J C, Wu C I. Hitchhiking under positive darwinian selection. *Genetics*, 2000, 155: 1405—1413
- 71 Lewontin R C, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphism. *Genetics*, 1973, 74: 175—195
- 72 Akey J M, Zhang G, Zhang K, et al. Interrogating a high-density SNP Map for signatures of natural selection. *Genome Res*, 2002, 12: 1805—1814[DOI]
- 73 Rockman M V, Hahn M W, Soranzo N, et al. Positive selection on a human-specific transcription factor binding site regulating IL4 expression. *Curr Biol*, 2003, 13: 2118—2123[DOI]
- 74 Sabeti P C, Reich D E, Higgins J M, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 2002, 419: 832—837[DOI]
- 75 Hanchard N A, Rockett K A, Spencer C, et al. Screening for recently selected alleles by analysis of human haplotype similarity. *Am J Hum Genet*, 2006, 78: 153—159[DOI]
- 76 Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 1998, 148: 929—936
- 77 Yang Z. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J Mol Evol*, 2000, 51: 423—432
- 78 Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*, 1998, 15: 568—573
- 79 Yang Z. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*, 1998, 46: 409—418[DOI]
- 80 Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*, 2002, 19: 908—917
- 81 Yang Z, Nielsen R, Goldman N, et al. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, 2000, 155: 431—449
- 82 Yang Z, Wong W, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol*, 2005, 22: 1107—1118[DOI]
- 83 Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*, 2005, 22: 2472—2479[DOI]
- 84 Anisimova M, Bielawski J P, Yang Z. Accuracy and power of bayes prediction of amino acid sites under positive selection. *Mol Biol Evol*, 2002, 19: 950—958
- 85 Anisimova M, Bielawski J P, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol*, 2001, 18: 1585—1592
- 86 Wang H Y, Chien H C, Osada N, et al. Rate of evolution in brain-expressed genes in humans and other primates. *PLoS Biol*, 2007, 5(2): e13[DOI]
- 87 Xia X. What amino acid properties affect protein evolution? *J Mol Evol*, 1998, 47: 557—564[DOI]
- 88 McClellan D A, McCracken K G. Estimating the influence of selection on the variable amino acid sites of the cytochrome b protein functional domains. *Mol Biol Evol*, 2001, 18: 917—925
- 89 Woolley S, Johnson J, Smith M J, et al. TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics*, 2003, 19: 671—672[DOI]
- 90 McDonald J, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 1991, 351: 652—654[DOI]
- 91 Hudson R, Kreitman M, Aguade M. A Test of neutral molecular evolution based on nucleotide data. *Genetics*, 1987, 116: 153—159
- 92 Currat M, Excoffier L, Maddison W, et al. Comment on "ongoing adaptive evolution of ASPM, a brain size determinant in homo sapiens" and "Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans". *Science*, 2006, 313: 172[DOI]
- 93 Mekel-Bobrov N, Posthuma D, Gilbert S L, et al. Response to comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens" and "Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans". *Science*, 2006, 313: 172[DOI]
- 94 McVean G A, Myers S R, Hunt S, et al. The fine-scale structure of recombination rate variation in the human genome. *Science*, 2004, 304: 581—584[DOI]
- 95 Crawford D C, Bhangale T, Li N, et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet*, 2004, 36: 700—706[DOI]
- 96 Reed F A, Tishkoff S A. Positive selection can create false hotspots of recombination. *Genetics*, 2006, 172: 2011—2014[DOI]
- 97 Nielsen R. Population genetic analysis of ascertained SNP data. *Human Genomics*, 2004, 1: 218—224
- 98 Haygood R, Fedrigo O, Hanson B, et al. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet*, 2007, 39(9): 1140—1144[DOI]
- 99 Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet*, 2006, 38: 1452—1456[DOI]