

DOI: 10.3724/SP.J.1005.2009.00896

# DNA 水平上检测正选择方法的研究进展

林栲, 李海鹏

中国科学院-德国马普学会计算生物学伙伴研究所, 上海 200031

**摘要:** 达尔文的自然选择学说指出, 自然选择作用是物种进化的主要因素。而 1968 年 Kimura 提出的中性进化学说认为中性突变和随机漂变才是进化的主要动力。在接下来的 30 多年时间中, 人们尝试从各种角度来检测自然选择是否存在。随着 DNA 测序技术的发展, 大量的 DNA 序列信息为检验自然选择提供了丰富的数据。因为自然选择会影响 DNA 变异模式, 所以可以通过分析现有的 DNA 样本来推断过去是否发生了自然选择。另一方面, 种群历史等因素也会影响到 DNA 变异模式, 因此会对自然选择的检测产生干扰。文章主要介绍了中性检验基本的概念, 全面回顾了一些经典的检验方法, 并着重介绍了近几年新发展出的研究方向。

**关键词:** DNA 水平; 自然选择; 中性检验

## Advances in detecting positive selection on genome

LIN Kao, LI Hai-Peng

CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences, Shanghai 200031, China

**Abstract:** After Darwin's natural selection theory emerged, Kimura proposed the theory of neutral evolution in 1968, which considered neutral mutations and random genetic drift as the major evolutionary forces. In the following years, various kinds of methods were developed to test whether natural selection has ever happened. With the improvement of DNA sequencing technologies, large amount of DNA sequence polymorphism data is available, providing a mass of materials for testing the natural selection theory. Since natural selection would leave its footprint on the genome, we are able to refer the adaptive evolutionary history of a population. On the other hand, population demographic history and other evolutionary forces may affect DNA polymorphic pattern in similar way, which may interfere with tests. In this paper, we summarized some basic concepts of neutral test, and briefly introduced some classical methods. Focus was given to several recently developed methods.

**Keywords:** DNA level; natural selection; neutrality test

达尔文的自然选择学说被认为是 19 世纪最重要的科学发现之一, 为整个进化论的发展奠定了坚实的基础。然而, Kimura<sup>[1]</sup>在 1968 年提出了与达尔文进化论相矛盾的中性进化学说, 认为中性突变和随机漂变等随机事件是物种进化的主要动力, 在之后的 40 年里, 对于中性进化和自然选择理论的研究,

始终是进化生物学研究的一个主要内容。

随着 DNA 测序技术的发展, 人们开始从 DNA 序列的水平上寻找自然选择留下的痕迹, 并且取得了一系列的成果, 提出了一系列相关的检验方法。这些检验方法大多以中性假说为零假设, 然后计算相关的统计量, 检测这些统计量与中性假设的情况

收稿日期: 2009-01-15; 修回日期: 2009-04-19

基金项目: 浦江人才计划项目(编号: 08PJ4104)和中科院百人计划项目资助

作者简介: 林栲(1986-), 男, 本科, 专业方向: 群体遗传学。E-mail: linkao@picb.ac.cn

通讯作者: 李海鹏(1975-), 男, 博士, 研究员, 研究方向: 群体遗传学。E-mail: lihaipeng@picb.ac.cn

是否相符。越来越多的证据表明, 在一些模式生物以及人类基因组中, 一些区域在一定的时间段经历了自然选择。

在近几年的几篇综述<sup>[2~4]</sup>中, 大致回顾了这些检验方法在近30几年来的发展情况。因此本文将简略地介绍相关的理论基础, 总结一些比较经典的检验方法, 并着重介绍最近几年的研究热点。

## 1 理论基础

### 1.1 群体样本数据特征

在分析群体样本时, 一般基于两种类型的数据: 一种是等位基因的数据, 另一种则是DNA核苷酸序列, 最常见的数据形式是基于单核苷酸多态性(Single nucleotide polymorphism, SNP)样本的数据。不同的检测方法可能适用于不同类型的数据, 有的检测方法适用于两种类型的数据。下面主要就DNA核苷酸序列的数据进行说明。

对于一个群体的DNA样本, 人们一般关注以下几个特征: 多态性、变异的频率谱线、同义突变与错义突变的比例<sup>[5,6]</sup>、单倍体型与连锁不平衡度<sup>[3,7]</sup>、群体结构。当自然选择或其他因素作用于群体的时候, 可能会改变以上这些特征, 所以分析DNA样本的这些特征成为推断群体历史的主要入手点。

### 1.2 影响DNA样本形态的因素

在群体的发展历史中, 有着各种各样的力量在影响着群体的遗传组成, 这些力量的共同作用造就了今天所观察到的遗传形态。这些力量主要有突变、遗传漂变、种群历史、自然选择、迁徙、重组等等。突变是指遗传信息的改变。突变为物种的进化提供了原材料, 增加了遗传的多样性。

遗传漂变是指在群体中由一代传往下一代的过程中, 因为每个个体能够拥有的后代数量具有一定的随机性, 而每个个体在序列同一个位点上的碱基可能是不一样的, 于是造成同一个位点上的不同的碱基频率发生随机的波动。随机漂变的力量主要受有效种群大小影响, 一般来说, 有效种群越大, 随机漂变的效应越小。

种群历史主要包括种群扩增、瓶颈效应、奠基者效应、群体缩减、分割、群体间的基因交流等等。

自然选择作用于非中性突变上, 或者引起有利

突变的频率在群体中增长, 或者消除不利的突变, 或者以其他方式对遗传的多样性进行修饰。自然选择的类型主要有正选择、负选择和平衡选择, 其中正选择会伴随产生搭载效应(Hitchhiking); 负选择会伴随产生背景选择(Background selection)。

上面提到的这些力量单独或共同作用于群体, 会影响到DNA样本的各种特征<sup>[8~14]</sup>。

## 2 检验正选择的方法

### 2.1 一些经典的检验方法

在中性进化学说提出后, 一系列针对中性假说的检验方法逐渐被提出来, 这些方法都是以DNA样本的各种特征为切入点, 通过计算相关的统计量或者与相关的模型进行比较, 来检测群体是否经历了正选择。这些方法包括Watterson的纯合性测试(Homozygosity test)<sup>[15]</sup>, 采用卡方检验的HKA测试<sup>[16]</sup>, 基于突变频率分布的Tajima's D检验<sup>[17]</sup>、Fu and Li's D<sup>[9,18]</sup>、Fay and Wu's H<sup>[10]</sup>, 基于非同义突变和同义突变比例的Ka/Ks检验<sup>[5]</sup>和M.K.检验<sup>[6]</sup>, 基于群体结构的 $F_{ST}$ 检验<sup>[19,20]</sup>等等。这些检验方法大部分在周琦等<sup>[2]</sup>的综述中有所介绍。本文主要介绍一些较新的检验方法和当前较热门的研究方向。

### 2.2 似然率检验

Kim和Stephan提出的似然率检验(Likelihood-ratio test, LRT)是一种拟合优度检验(Good of fitness), 通过计算样本数据在搭载模型和中性模型下似然函数的比值, 来判断是否发生了选择扫荡(Selective sweep)<sup>[21]</sup>。其理论基础是搭载效应将会在选择座位处产生特异性的多态性低谷, 而通过相关的似然函数就可以检验一个DNA区域中出现的低谷是否超出了中性进化模型的形态。

在中性假设和搭载效应的假设下, 在一个分离位点能够观察到某个特定频率突变的概率, 是能够通过理论计算得到的。对于一段群体DNA序列, 计算出其中每个分离位点的突变频率分别在中性假设和搭载效应假设下观察到的概率, 然后将这些概率相乘, 就分别得到了中性假设和搭载效应假设的似然函数 $L_0$ 和 $L_1$ 。其中 $L_0$ 是群体突变率的函数,  $L_1$ 则是有效群体大小、突变率、重组率、选择强度和有利突变位置的函数。求出 $L_0$ 和 $L_1$ 的极大值(即极

大似然函数), 把二者的比值  $L_1/L_0$  作为检验统计量, 通过模拟实验得到中性模型下这个统计量的分布, 然后检验这个比值是否处于分布的尾端, 从而检验对应的 DNA 区域是否发生了正选择。

在求  $L_1$  的最大值时, 因为自由变量较多, 所以一般很难找到一个最优组合使得  $L_1$  达到最大。在实际运用中, 往往通过其他途径获得有效群体大小、突变率、重组率的信息, 而仅仅将选择强度和有利突变位置作为自由变量, 这样, 在检验正选择过程的同时, 求出了满足  $L_1$  达到最大时的选择强度和有利突变位置, 即得到了要考察的正选择参数的估计值。例如 Li 等<sup>[22]</sup>通过似然率检验来估计正选择发生的位点。

似然率检验考虑的因素很多, 例如将有效群体大小、选择强度、重组率、选择发生的时间等因素均加入到了溯祖模拟模型中。通过改变这些参数的值, 可以观察不同强度的因素对模拟结果的影响。

### 2.3 基于单倍体型比较的检验方法

在一段 DNA 序列中, 位点与位点之间存在着连锁的关系。在一段时间内, 这段区域上的位点将保持一定的连锁形态, 这叫做“单倍体型”。随着时间的流逝, 重组将会打破这样的固定组合, 造成特定的单倍体型逐渐消失。由于重组率与连锁距离有关, 所以重组会先打破遗传距离较远的位点间的连锁关系, 然后逐渐打破较近位点的连锁关系。所以, 如果某个单倍体型在群体中的连锁不平衡 (Linkage disequilibrium, LD) 所影响的距离较远, 那么这个单倍体型往往比较新。在中性条件下, 如果某个单倍体型是较新产生的, 那么它的频率往往较低, 而频率较高的单倍体型, 需要经历很长一段时间才可能因为受到随机漂变的影响达到较高的频率。

如果群体经历了正选择, 那么与有利位点连锁的周围位点会由于搭载效应而得到很快的频率提升, 所以包含有利位点的单倍体型一方面有着较高的频率, 另一方面由于经历的时间不长, 因此也有着较长的 LD 影响范围。这种特征为检测是否发生了正向选择提供了一个有效的突破点。

#### 2.3.1 LRH(Long-range haplotypes)

在 Sabeti 等<sup>[2]</sup>的研究中, 通过对 LRH 的分析来检测正选择的存在。他们选择基因组中的核心单倍体型(Core haplotypes)作为研究对象。所谓的核心单倍

体型就是指基因中存在的重组率较低的密集区域。计算它们的连锁不平衡度, 如果某个核心单倍体型的连锁不平衡程度高于具有其相同频率的一般单倍体型, 那么这个位置很有可能经历了正选择。关于连锁不平衡度, 他们定义了 EHH(Extended haplotype homozygosity)作为检测量。假如要测量距离核心单倍体型为  $x$  的区域的连锁不平衡度, EHH 的定义是: 两条随机选择的染色体从核心单倍体型到距离为  $x$  之间的区域是相同的概率<sup>[2]</sup>。

#### 2.3.2 HS(Haplosimilarity)

HS 检验是计算单倍体型相似性的检验方法。对于一批 DNA 样本数据, 观察其第一个多态位点, 记这个多态位点上频率较低的等位基因为  $X$ , 然后计算  $X$  所关联的染色体的 HS 值。计算方法是通过一个滑动窗口滑过整段染色体, 计算每个窗口中单倍体型的纯合度, 然后对所有的窗口取平均值。

$$HS = \frac{\sum_{t=1}^T \sum_i^k f_{it}^2}{T}$$

其中  $T$  是窗口的总数,  $k$  是一个窗口中不同单倍体型的个数,  $f_{it}$  是与  $X$  相关联的单倍体型的频率<sup>[23]</sup>。

上述过程是以第一个多态位点为基准进行的, 同样可以以第二个、第三个等多态位点为基准进行类似的计算。以某个多态位点为基准计算, 如果其相关 HS 值的水平高于同等频率下的其他多态位点, 那么在该多态位点上可能发生了正选择。

#### 2.3.3 iHS(iHH score)

iHS 是通过计算同一个 SNP 上旧的和新的等位基因的 iHH 比值并取对数得到的:

$$iHS = \ln \left( \frac{iHH_A}{iHH_D} \right)$$

其中 iHH 指对 EHH 的积分(Integrated EHH),  $A$  指旧的 (Ancestral) 等位基因,  $D$  指新的 (Drived) 等位基因<sup>[24]</sup>。

iHS 的基本原理和 LRH 很相似。当 iHS 为较大的正值时, 暗示长的单倍体型可能包含旧的等位基因, 而 iHS 为较大的负值时, 暗示长的单倍体型可能包含新的等位基因<sup>[4]</sup>。

#### 2.3.4 LDD(Linkage disequilibrium decay)

LDD 测试指连锁不平衡衰减测试。而 LD 是通过



计算FRC(Fraction of recombinant chromosomes)来体现的。具体方法是, 对于一个多态座位, 不考虑这个座位上的杂合子, 而在纯合子中观察其较少的等位基因和较多的等位基因, 考察所有的染色体, 将其中与较少等位基因关联的编成一组, 而将与较多等位基因关联的编成另一组, 然后分别在两组内这个位点周围一个事先预设好的窗口中, 计算重组频率与距离的关系, 也就是计算不同距离范围内相应的重组率。将这些重组率和相应的距离配对、列表, 和标准中性模型的这些值进行方差比较, 即计算出ALnLH(Average log likelihood)<sup>[25]</sup>。

在正选择发生时, 临近选择位点的ALnLH将高出一般的水平<sup>[4]</sup>。

## 2.4 基于微卫星多样性的检验

在基因组中, SNP 数据是 DNA 样本多态性一种主要标记, 但是这些 SNP 数据大多都是双等位基因, 单个 SNP 位点包含的多态性信息较少。而微卫星 DNA 是基因组中 DNA 一些小片段的连续重复, 在不同的个体中, 同一个位置的微卫星 DNA 重复次数可能存有差别, 这些重复次数的差别也是一种 DNA 的多态性。微卫星 DNA 重复次数的差别可以很大, 因此单个微卫星位点所包含的多态性信息非常丰富。近几年, 有的研究就采用微卫星位点作为多态性的标记来设计自然选择的检验方法。

因为微卫星序列一般是中性的, 所以一般用来研究发生在与它们连锁的位点上的正选择。与 SNP 位点类似, 当微卫星 DNA 与正选择的作用位点相邻时, 微卫星 DNA 的多态性会降低。

Schlötterer<sup>[26]</sup>提出的lnRV方法就是基于微卫星序列重复次数多态性比例的检验。由于不同位点微卫星序列的突变率不同, 所以如果只考察各个位点的微卫星序列重复次数, 得到的值之间不具有很好的可比性。但是如果考虑两个独立群体在相同位点微卫星序列重复次数的比值, 那么在中性情况下, 这个比值在各个群体中具有一定的一致性。这个比值被称作RV(The Ratio of the variance in repeat number), 对RV取对数即得到lnRV。lnRV在中性条件下满足正态分布。如果其中一个群体的某个位点经历了搭载效应, 那么这个位置的lnRV值将会落在中性情况下的分布尾端。

## 2.5 复合检验

在各种检验正选择的统计量中, 都有各自的优势或劣势, 例如 Fay and Wu' H 是基于高频突变丰度的检验, 它能够比较特异性地检验正选择, 而受群体历史和背景选择的干扰较少, 但是它只能检测到刚固定不久的正选择, 因为高频突变将随着时间流逝很快因为随机漂变作用而被固定。Tajima's D 在检验正选择的同时容易受到群体历史和背景选择的干扰, 但是 Tajima's D 所检验的低频突变丰度的信号能够在选择发生位点被固定后持续较长一段时间。一个比较容易想到的方法就是同时利用两种或多种检验方法, 使它们的优缺点得以互补, 从而能够较特异性地检验正选择。

Zeng 等<sup>[27]</sup>提出的DH检验就是直接结合了Tajima's D 和一个修正后的Fay and Wu's H检验, 其检验正向选择的特异性能力相对较高, 而对种群历史等其他因素的敏感度很低。后来考虑到Ewens-Watterson的EW检验对重组率的变化不敏感, Zeng等<sup>[28]</sup>又提出了Fay and Wu's H和EW结合的HEW检验以及DH和EW结合的DHEW检验, 它们相对于H检验或DH检验对重组率更不敏感。

## 2.6 全基因组扫描及假阳性问题

在单座位检验(Single-locus test)问题中, 判断这个座位是否发生了正选择往往受到种群历史的干扰, 而判断的结果也受到假设模型的影响, 所以为了解决这个问题, 一些研究者提出了多座位检验(Multi-locus test)的方法。多座位检验的基本思想是, 种群历史将会影响整个基因组的DNA变异模式, 而正选择只是特异性地作用于某个座位。在多座位比较中, 那些特征相异于普遍水平的座位就可能经历了正选择。多座位检验与单座位检验最大的区别就在于, 单座位检验是将统计值与假设模型的模拟分布比较, 而多座位是与其他所有座位的一般性水平比较, 所以多座位检验也叫做经验手段(Empirical approach)。

全基因组扫描的异常值方法(Outlier approach)是采取全基因组的全部或部分多态性座位进行多座位检验的方法。这种方法首先计算出所有考察座位的特征(多态性、突变频谱、连锁不平衡程度等)并得到一个经验分布, 然后取出这些分布尾端的值, 认为这些值对应的座位是经历了正选择的<sup>[29]</sup>。这种方法不需要构

建特定的模型, 原理比较直观, 但是存在很多问题。

最根本的问题在于, 目前还没有证据表明, 位于分布尾端的座位一定是经历了正选择的座位, 也没有证据表明经历了正选择的座位一定会位于分布的尾端。在经过全基因组扫描过后得到的座位还需要继续用其他的方法进行进一步验证。第二步检验的结果可能会存在确认偏倚(Ascertainment bias)问题<sup>[30]</sup>, 如果不考虑这个问题将可能得到假阳性的结果。另外, 不论一个群体是否经历了正选择, 这样计算出来的统计量分布都会存在一个尾部, 如果总是把这样的尾部当作正选择, 那么将可能产生假阳性。对于非平衡的群体, 其经过了选择扫荡后的座位, 在统计量分布尾端以何种程度出现也是未知的。为了解决这些问题, 一些修正的方案也逐渐被提出来。Thornton等<sup>[30]</sup>提出了修正确认偏倚问题和非平衡群体问题的方法, 指出如果不考虑确认偏倚的问题, 那么第二步检验用到的似然率检验方法可能会受到第一次扫描选择出来座位的影响很大。同时他认为在选择统计量的时候, 采用多样性和群体分化程度的统计量相对于选择突变频谱更为有效。

在全基因组扫描中, 因为选择模式的不同和受种群历史的干扰, 也会产生假阳性问题。例如, 如果选择作用在隐性等位基因而不是共显性等位基因上, 或者选择作用在一个新的突变而不是一个已经存在的分离位点上, 或者选择的同时经历了瓶颈效应, 这些因素都会使造成假阳性的可能性增大<sup>[31]</sup>。

此外, 对于基于高频突变的Fay and Wu's H以及基于连锁不平衡的统计量, 都会在正选择完成后不久很快丧失其特征, 因为高频突变会很快被固定, 而连锁不平衡会很快被重组打断<sup>[32]</sup>。如果采用基于高频突变变化或基于连锁不平衡的统计量来进行全基因组的扫描, 可能会遗漏一些可能经历过正选择的座位。

另一个策略是用基因组水平的多态位点估计群体数量历史变化参数, 再将估计的参数作为原假设(Null hypothesis), 通过似然率测试(Likelihood ratio test)来检测正选择<sup>[33]</sup>。

### 3 结论与展望

由于多种因素的共同作用形成了现有的 DNA 变异模式, 而且不同的因素可能造成同样的结果,

所以往往很难通过一种检验方法检测出某一个特定的事件。通过多种方法的联合使用, 可能会有效地检验出某一个特定的事件。

现有的各种检验一般都有各自的理论模型与之对应, 而计算这些统计量的分布时通常就采用对相应模型进行模拟的方法来实现。但是这些模型都包含了各种各样的假设, 当这些假设与真实情况相背离的时候, 相应的检验方法可能会作出不正确的预测。全基因组扫描虽然是基于经验手段的, 不需要通过构造模型来决定统计值是否具有特异性, 但是多种因素可能造成假阳性结果, 还需要进一步完善。

总体来说, 在检测正选择的时候, 可能需要同时考虑很多因素, 以下任何一个因素的改变都可能使得预测的结果发生偏差:

(1) 种群历史、背景选择和平衡选择: 在这些检验方法中, 一个比较普遍的问题是如何区分自然选择和种群历史的作用。一个思路是, 自然选择往往作用于某些特定的座位, 而种群历史影响整个 DNA 序列。如果某个区域具有偏离了中性进化的特征, 那么要判断这是因为经历了自然选择还是因为种群历史的变化, 可以通过将待检测的区域与整个 DNA 序列的相关特征(例如多态性)进行比较, 观察其是否和普遍的水平一致, 如果不一致的话, 那么其可能是经历了自然选择, 全基因组扫描就是基于这样一个思路。另一个思路就是尽量考虑自然选择所形成的特异性特征, 而这些特征往往是种群历史因素不能形成的, 似然率检验就是这样实现的。

背景选择也会对正选择有一定的干扰作用, 它们都能产生大量的低频突变, 但是正选择产生相对较多的高频突变, 这个特征是背景选择没有的<sup>[10]</sup>。

平衡选择在刚开始的时候和正选择很类似, 因为二者在刚开始作用于一个新的突变的时候都会使这种碱基的频率增加, 所不同的是平衡选择最终使新的碱基和旧的碱基的频率达到一个平衡, 而正选择最终会让新的有利碱基替换掉旧的碱基。

(2) 选择发生的时间: 选择是正在进行中的、刚刚完成的还是已经完成很久了, 其产生的结果可能是不同的。选择发生的时间不同, 所造成的DNA变异模式也会有所不同。正选择留下的痕迹可能随着时间的推移而减弱, 并且不同的痕迹所能持续的时间也不同。搭载效应产生的高频突变和连锁不平衡

就会很快被随机漂变和重组效应消除,而功能(非同义)突变与非功能(同义)突变之间的比例改变就能够持续相当长的时间<sup>[3]</sup>。在时间上,对于不同的物种来说,群体的真实历史也是需要考虑的一个因素。例如人类从非洲走出的时间大约在 50 000 到 75 000 年前,所以在人类的DNA数据中,如果考察从亚群体间差别程度(例如 $F_{ST}$ )来检验人类走出非洲后是否经历了某些自然选择,需要考虑到这个时间的因素。

(3) 选择发生的位置: 突变率和重组率是多少、多座位检验时各个座位的突变率和重组率是否一致,都可能会影响到检验的准确性。不同的突变率和重组率可能会使得检验统计量的临界值发生变化,而多座位检验时,如果各个座位的突变率和重组率不一致,那么可能产生假阳性的结果。另外,在一个位点发生了选择后,其所影响的周围中性位点的距离也是需要考虑的,距离越远,连锁程度越低,影响越弱。在空间上,较早期的检验,例如Tajima's D,都只是考虑了一段DNA序列区域内是否经历了自然选择,而没有考虑自然选择直接作用的位点,而较近期的似然率检验和单倍体连锁不平衡检验,都考虑到了选择直接作用位点及其周围的特异性变异模式(多态性低谷或连锁不平衡衰减),所以相对灵敏度更高,特异性更强。例如,HS检验的灵敏度与LRH差不多,但是远远高于Tajima's D等检验方法<sup>[23]</sup>。

(4) 选择作用的对象: 选择作用于新的突变还是已经存在的分离位点,产生的结果可能是不同的。选择作用于一个新的突变上时相对比较容易检验,而如果选择作用于已经存在的分离位点上,那么有的统计量,例如Fay and Wu' H 和基于连锁不平衡的信号就不是特别强烈了<sup>[32]</sup>。

(5) 有利等位基因的类型: 有利等位基因可能是隐性的,也可能是共显性的。隐性的有利等位基因产生后,比起共显性的有利等位基因的频率上升的速度相对较慢,选择的力量相对较弱,需要较长一段时间完成选择<sup>[31]</sup>。所以在考虑这个因素的时候实际上也是在考虑时间上的因素。

(6) 选择的强度: 选择的强度可能由多种因素决定,上面提到的等位基因是隐性或共显性也属于影响选择强度的一种因素。选择强度最根本还是取决于自然环境对表型的影响能力。选择强度越强,选择所经历的时间越短。

(7) 单个选择还是复发的选择: 在同一个座位只发生了一次选择事件还是多次,也会造成结果的不同。如果同一个座位经历了多次选择事件,那么在全基因组扫描中, Fay and Wu' H或基于连锁不平衡的统计量就不会表现出那么明显的特异性了<sup>[32]</sup>。

在未来的发展方向上,由于越来越多DNA数据的积累,人们将根据具体的数据应用相应的检验方法来推测各种历史事件的发生,人类HAPMAP<sup>[34]</sup>数据库的建立为这些方向的发展提供了丰富的原材料。另一方面,由于考虑的因素越来越多,各种模型将会越来越复杂,各种检验将会在的灵敏度和特异性上不断地努力,这个领域内的理论发展也将越来越具有挑战性。

## 参考文献(References):

- [1] Kimura M. Evolutionary rate at the molecular level. *Nature*, 1968, 217(5129): 624–626. [DOI](#)
- [2] 周琦, 王文. DNA 水平自然选择作用的检测. 动物学研究, 2004, 25(1): 73–80.
- [3] Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. Positive natural selection in the human lineage. *Science*, 2006, 312(5780): 1614–1620. [DOI](#)
- [4] Biswas S, Akey JM. Genomic insights into positive selection. *Trends Genet*, 2006, 22(8): 437–446. [DOI](#)
- [5] Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol*, 1985, 2(2): 150–174.
- [6] McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 1991, 351(6328): 652–654. [DOI](#)
- [7] Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 2002, 419(6909): 832–837. [DOI](#)
- [8] Tajima F. The effect of change in population size on DNA polymorphism. *Genetics*, 1989, 123(3): 597–601.
- [9] Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 1997, 147(2): 915–925.
- [10] Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics*, 2000, 155(3): 1405–1413.
- [11] Otto SP. Detecting the form of selection from DNA sequence data. *Trends Genet*, 2000, 16(12): 526–529. [DOI](#)



- [12] Kim Y, Stephan W. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*, 2000, 155(3): 1415–1427.
- [13] Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*, 1993, 133(3): 693–709.
- [14] Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, 1992, 356(6369): 519–520. [\[DOI\]](#)
- [15] Watterson GA. The homozygosity test of neutrality. *Genetics*, 1978, 88(2): 405–417.
- [16] Hudson RR, Kreitman M, Aguade M. A test of neutral molecular evolution based on nucleotide data. *Genetics*, 1987, 116(1): 153–159.
- [17] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 1989, 123(3): 585–595.
- [18] Fu YX. New statistical tests of neutrality for DNA samples from a population. *Genetics*, 1996, 143(1): 557–570.
- [19] Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 1973, 74(1): 175–195.
- [20] Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res*, 2002, 12(12): 1805–1814. [\[DOI\]](#)
- [21] Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 2002, 160(2): 765–777.
- [22] Li H, Stephan W. Maximum-likelihood methods for detecting recent positive selection and localizing the selected site in the genome. *Genetics*, 2005, 171(1): 377–384. [\[DOI\]](#)
- [23] Hanchard NA, Rockett KA, Spencer C, Coop G, Pinder M, Jallow M, Kimber M, McVean G, Mott R, Kwiatkowski DP. Screening for recently selected alleles by analysis of human haplotype similarity. *Am J Hum Genet*, 2006, 78(1): 153–159. [\[DOI\]](#)
- [24] Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*, 2006, 4(3): e72. [\[DOI\]](#)
- [25] Wang ET, Kodama G, Baldi P, Moyzis RK. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc Natl Acad Sci USA*, 2006, 103(1): 135–140. [\[DOI\]](#)
- [26] Schlötterer C. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, 2002, 160(2): 753–763.
- [27] Zeng K, Fu YX, Shi S, Wu CI. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*, 2006, 174(3): 1431–1439. [\[DOI\]](#)
- [28] Zeng K, Shi S, Wu CI. Compound tests for the detection of hitchhiking under positive selection. *Mol Biol Evol*, 2007, 24(8): 1898–1908. [\[DOI\]](#)
- [29] Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res*, 2006, 16(8): 980–989. [\[DOI\]](#)
- [30] Thornton KR, Jensen JD. Controlling the false-positive rate in multilocus genome scans for selection. *Genetics*, 2007, 175(2): 737–750. [\[DOI\]](#)
- [31] Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? *Genome Res*, 2006, 16(6): 702–712. [\[DOI\]](#)
- [32] Przeworski M. The signature of positive selection at randomly chosen loci. *Genetics*, 2002, 160(3): 1179–1189.
- [33] Li H, Stephan W. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet*, 2006, 2(10): e166. [\[DOI\]](#)
- [34] The International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005, 437(7063): 1299–1320. [\[DOI\]](#)

## •综合信息•

### 欢迎订阅《生物技术》杂志

《生物技术》杂志是中文核心期刊、中国生物学核心期刊,被生物学数据库(BIOSIS)、生物学文摘(BA)、中国科学引文数据库、核心期刊要目总览、中国核心期刊遴选数据库等近 20 多家国内外重要检索系统、数据库和文摘类期刊收录。是涉及生物工程、生物技术、微生物及生命科学相关领域的综合性学术刊期刊。主要刊登生物技术工程、微生物、医药、食品、农林、食药真菌、环保及生物学相关领域的研究论文。

《生物技术》(ISSN1004-311X, CN23-1319/Q)为双月刊,全年 6 期,2010 年定价 10.00 元,全年 60.00 元。国内外公开发行,国内邮发代号:14-225,全国各地邮局订阅;国外发行代号:BM5606。

地址:哈尔滨市道里区兆麟街 68 号《生物技术》编辑部

邮编:150010 <http://www.hljiam.cn>

联系人:奚新伟 电话:0451-84615121 传真:0451-84614297

Email: swjszxxw@163.com; swjszz@163.com