



## 基因组所完成“人类多层次多人群自然选择数据库”构建

文章来源: 北京基因组研究所

发布时间: 2009-12-28

【字号: 小 中 大】

为了更好地了解人类不同种群的遗传差异和受到的自然选择情况,以及比较不同指标之间的印证关系。近日,中国科学院北京基因组研究所曾长青研究员实验室的程锋等人,通过选择目前SNP(单核苷酸多态性)分型数量最大和种群数最多的HapMap(国际人类基因组单体型图计划)分型数据作为研究基础,从基因组大片段,功能基因以及单个SNP位点等三个层次来研究人类不同种群的基因组遗传分化和所受自然选择的情况。根据使用多个不同指标(HET, Win\_HET, FST, Win\_FST, iHS, ES\_HET, ES\_FST, P\_iHS等)及策略来扫描选择信号,并把它们置于同一个框架下进行比较和验证,以求获得最大的信息。研究结果建立了“人类多层次多人群自然选择数据库”暨阳性自然选择数据库SNP@Evolution (<http://bighapmap.big.ac.cn/>)供国内外科研使用,自九月下旬相关文章在*BMC Evol Biol*发表以来,SNP@Evolution已受到来自全世界几十个国家和地区,上万次的访问和下载,为该领域的研究人员提供了一个发现选择信号的有用工具。

SNP@Evolution共分为数据查询和图形查询界面两个部分。包括了HapMap II期和III期的数据结果。II期共有3,619,226个SNP数据,以及21,859个基因的分析数据。共有1606个基因组大片段显示选择信号,660个显示分化信号。III期数据共包含1,389,498 SNPs,21,099个有效基因分析数据。在11个人群中找到了10,138个受选择的基因组片段,以及464个具有强分化的基因组片段。为了方便研究,SNP@Evolution的查询结果可以链接到其他数据库获取更多信息。

数据库链接:

<http://bighapmap.big.ac.cn>

文献记录:

Cheng Feng, Chen Wei, Richards Elliott, Deng Libin, Zeng Changqing. SNP@Evolution: a hierarchical database of positive selection on the human genome. *BMC Evolutionary Biology* 2009, 9:221.

原文链接:

<http://www.biomedcentral.com/1471-2148/9/221>

原文摘要:

Abstract

Background: Positive selection is a driving force that has shaped the modern human. Recent developments in high throughput technologies and corresponding statistics tools have made it possible to conduct whole genome surveys at a population scale, and a variety of measurements, such as heterozygosity (HET), FST, and Tajima's D, have been applied to multiple datasets to identify signals of positive selection. However, great effort has been required to combine various types of data from individual sources, and incompatibility among datasets has been a common problem. SNP@Evolution, a new database which integrates multiple datasets, will greatly assist future work in this area.

Description As part of our research scanning for evolutionary signals in HapMap Phase II and Phase III datasets, we built SNP@Evolution as a multi-aspect database focused on positive selection. Among its many features, SNP@Evolution provides computed FST and HET of all HapMap SNPs, 5+ HapMap SNPs per qualified gene, and all autosome regions detected from whole genome window scanning. In an attempt to capture multiple selection signals across the genome, selection-signal enrichment strength (ES) values of HET, FST, and P-values of iHS of most annotated genes have been calculated and integrated within one frame for users to search for outliers. Genes with significant ES or P-values (with thresholds of 0.95 and 0.05, respectively) have been highlighted in color. Low diversity chromosome regions have been detected by sliding a 100 kb window in a 10 kb step. To allow this information to be easily disseminated, a graphical user interface (GBrowser) was constructed with the Generic Model Organism Database toolkit.

Conclusion: Available at <http://bighapmap.big.ac.cn>, SNP@Evolution is a hierarchical database focused on positive selection of the human genome. Based on HapMap Phase II and III data, SNP@Evolution includes 3,619,226/1,389,498 SNPs with their computed HET and FST, as well as qualified genes of 21,859/21,099 with ES values of HET and FST. In at least one HapMap population group, window scanning for selection signals has resulted in 1,606/10,138 large low HET regions. Among Phase II and III geographical groups, 660 and 464 regions show strong differentiation.

打印本页

关闭本页