遗传 Hereditas (Beijing) 2016 年 2 月, 38(2): 137－143
www.chinagene.cn

综 述

细菌基因组同源重组：量化与鉴定

杨献伟，杨瑞馥，崔玉军

军事医学科学院微生物流行病研究所，病原微生物生物安全国家重点实验室，北京 100071

摘要： 同源重组(Homologous recombination)是塑造细菌群体多样性的重要原因之一。遗传物质通过同源重组在细菌不同种系间进行水平转移，打乱了克隆繁殖形成的竖向系统发育结构，从而为系统发育重建和种群结构判定带来困难。本文讨论了同源重组对系统发育分析和进化研究的影响，从实际应用的角度对量化重组程度和鉴定重组事件的常用软件及方法进行了综述，归纳了各软件工具和模型方法的优缺点，旨在对细菌重组分析和种群进化研究有所借鉴。

关键词： 同源重组；克隆；基因组；系统发育；种群结构

Homologous recombination among bacterial genomes: the measurement and identification

Xianwei Yang, Ruifu Yang, Yujun Cui

State Key Laboratory of Pathogen and Biosecurity, Institute of Microbiology and Epidemiology, Beijing 100071, China

Abstract: Homologous recombination is one of important sources in shaping the bacterial population diversity, which disrupts the clonal relationship among different lineages through horizontal transferring of DNA-segments. As consequence of blurring the vertical inheritance signals, the homologous recombination raises difficulties in phylogenetic analysis and reconstruction of population structure. Here we discuss the impacts of homologous recombination in inferring phylogenetic relationship among bacterial isolates, and summarize the tools and models separately used in recombination measurement and identification. We also highlight the merits and drawbacks of various approaches, aiming to assist in the practical application for the analysis of homologous recombination in bacterial evolution research.

Keywords: homologous recombination; clonal; genome; phylogenetic; population structure

真核生物生命周期中存在减数分裂过程，非姐妹染色单体通过交叉互换可以将不同染色体片段进行重组，不同等位基因得以重新组合，使后代呈现新的基因型。细菌属于原核生物，个体细胞以克隆方式进行无性繁殖，遗传物质通常是由亲代个体竖向遗传给子代个体，然而，越来越多的研究表明，在细菌种群进化过程中，重组也是重要的进化推动力之一；多种细菌种群内或种群间都发现了遗传物

收稿日期： 2015−09−08；**修回日期：** 2015−12−04
基金项目： 国家“十二五”科技重大专项(编号：2012ZX10004215)和国家自然科学基金(编号：31430006)[Supported by the National Key Program for Infectious Diseases of China (No. 2012ZX10004215) and the National Natural Science Foundation of China (No. 31430006)]
作者简介： 杨献伟，硕士生，专业方向：病原细菌基因组学与进化。E-mail: yangxianwei1988@163.com
通讯作者： 崔玉军，副研究员，硕士生导师，研究方向：病原细菌基因组学与进化。E-mail: cuiyujun.new@gmail.com
DOI: 10.16288/j.yczz.15-382
网络出版时间： 2015-12-30 17:12:47
URL: http://www.cnki.net/kcms/detail/11.1913.R.20151230.1712.003.html

质的重组现象[1]。细菌的重组机制主要包括：转换、转导、接合以及溶源性转变[2]。发生于细菌基因组上的重组事件通常是由同源性较高的序列介导，表现为同源序列替换，被称为同源重组(Homologous recombination)。在奈瑟菌(*Neisseria*)编码抗原和抗生素抗性的基因上首次观察到细菌基因组的同源重组现象[3, 4]。另一类重组发生于非同源片段之间，这种重组通常会引入全新基因或基因组岛的插入，也常被称为水平基因转移(Lateral gene transfer, LGT)。

同源重组对细菌种群的生存和进化具有重要意义。相对自发突变，同源重组能够为种群带来更快的变异速度，可迅速提高种群遗传多样性和环境适应能力[5]。新获得的变异位点使基因在功能或者调控层面发生变化，导致相应菌株呈现新的表型，如逃避宿主免疫反应、产生多耐药性以及提高致病性等[6,7]。因此，对同源重组事件的识别可以为细菌的生物学功能研究提供更多线索。另一方面，同源重组事件会打乱细菌种系之间的竖向遗传关系，为系统发育结构重建造成困难，并影响对进化规律的分析。因此，在细菌的群体遗传学分析中，首先需要对同源重组的程度进行量化，评估其对系统发育和进化分析的影响；并对所发生的重组事件和相应基因片段进行判定。本文讨论了同源重组(下文中的重组均指同源重组)对系统发育分析过程的影响，并对其度量方法和检测鉴定工具进行了梳理总结。

1 同源重组对系统发育分析的影响

为探索种群进化关系，在群体遗传学分析中通常基于细菌克隆增殖的性质，根据自发突变进行遗传关系回溯。这种方法被成功应用到遗传单态性细菌物种的分析中[8]。如在对上百株鼠疫耶尔森氏菌(*Yersinia pestis*)的进化分析中，利用全基因组范围的单核苷酸多态性(Single nucleotide polymorphism, SNP)分析建立了可靠的种群系统发育关系并与地域分布进行了关联[9]。然而，不是所有细菌物种都可以直接利用种群内变异信息进行系统发育结构重建。当种群内部发生重组时，种群的竖向遗传结构可能被干扰。如图 1 所示，在进化关系上，菌株 S1 和 S2 的遗传距离更近。当种群内无重组事件时，菌株间的进化关系能够通过遗传距离反映出来；当基因 B 中菌株 S2 与 S3 发生重组，S2 接受了 S3 中的基因，会使得菌株 S2 和 S3 的遗传关系更近，从而影响了真实的系统发育结构。

重组程度低的种群中，竖向遗传信号能够被较

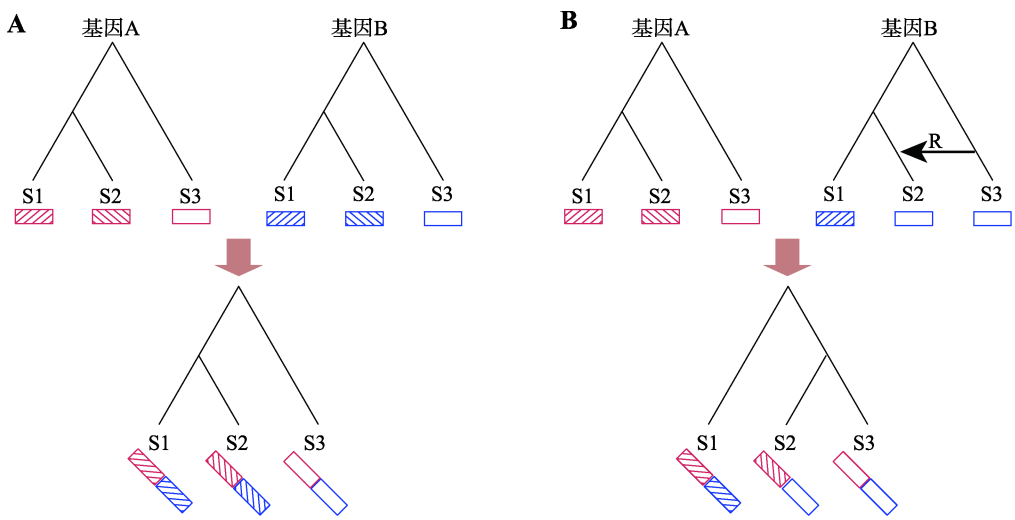

图 1 重组对系统发育重建的影响示意图

Fig. 1 The impact of homologous recombination on phylogeny rebuilding

图中表示了三株菌 S1，S2 和 S3 的进化关系，红色和蓝色矩形分别表示两个存在变异的基因 A 和基因 B。图 A 表示真实的系统发育关系，两个基因都未发生重组，利用两个基因序列进行系统发育结构重建能够还原出种群的原始结构；图 B 表示发生同源重组后的情况，由于在基因 B 中菌株 S2 获得了 S3 的基因片段(图中 R 表示发生重组事件)，拉近了 S2 与 S3 的遗传关系，使得真实系统发育关系无法被准确还原。

好的保留下来。例如对 149 株甲型副伤寒沙门氏菌(*Salmonella enterica* serovar Paratyphi A)的全基因组分析发现，在总长度 4.07 Mb 的核心基因组中由重组引入的 SNP 位点仅有 88 个, 占总位点数的 1.9%。该种群基于所有 SNP 和去除重组位点后的 SNP 构建的两系统发育结构相一致[10]。而种群中重组程度提高，会使得竖向遗传信号比例降低，造成系统发育结构逐渐模糊。副溶血弧菌(*Vibrio parahaemolyticus*)是一个重组高发的物种，高频重组使得竖向进化关系完全被打乱，整个物种的系统发育关系呈现出辐射状结构，无法反映各个种系间的进化先后顺序和亲缘关系远近[11]。因此，对此类细菌物种进行进化分析时就必须考虑重组带来的影响，使用非系统发育的方法重建其种群结构。

2　重组程度的量化方法

如果在研究数据中通过不同基因序列子集构建的系统发育结构之间关系不一致时，可以采用分离式网络来对其进行定性化展示。分离式网络中应包含数据中出现的所有进化树，理想情况下网络复杂度应能反映重组的发生频度。构建分离式网络的方法有很多[12]，其中比较常用的是“分解法”[13]以及在此基础上的“邻接网络”[14]。软件 SplitsTree4[12]中整合了此类方法。该软件可以识别包括序列、遗传距离在内的多种数据，提供了多种方法进行网络构建和直观展示。虽然其计算速度很快，但是构建的网络结构并不能区分真实发生的重组和本身不稳定的进化关系(如选择压力作用下的趋同突变), 并且无法对重组进行准确量化。因此分离式网络适用于数据量较大且怀疑种群受重组影响时对数据的直观展示, 仅能为后续深入分析提供参考。

目前，对重组程度进行定量估计的常用方法主要分为两类：(1)基于对核酸序列变异的归纳统计；(2)利用系统发育重建进行估计。

2.1　基于对核酸序列变异的归纳统计

在此类方法中，连锁失衡(Linkage disequilibrium, LD)是一种比较常用的归纳统计方法, 可用来估计种群内的重组程度[11,15]。连锁失衡最初用于真核生物，在减数分裂中，染色体上距离越近的两个基因越不容易成为重组的断点，当两个基因同时遗传给子代时，称为连锁失衡。在理想的以克隆增殖方式进化的群体中，基因组上的非等位基因组合不会发生变化，均处于连锁失衡状态；重组发生后，群体内非等位基因可以出现不同组合，连锁失衡水平下降，甚至在相距较远的基因间，基因独立遗传给下一代，达到连锁平衡的状态。进行连锁失衡分析的常用软件是 Haploview[16]，软件结果中可以给出 D'和 r^2 的信息，这两者都是反映连锁失衡水平的指标，其数值越大, 表示连锁失衡水平越大, 相应的重组程度越低。

更为常用的种群重组发生率衡量参数是 ρ/θ, 其中 ρ 是重组事件的发生次数，θ 是自发突变(非重组所致变异)事件的发生次数。使用 LDhat[17]、DnaSP[18]等软件，基于对数据的归纳统计可以获得此类参数。其中 LDhat 采用哈德森复合似然法[19]使用有限位点模型[20]进行群体重组率的估计，可以处理单体型或基因型的数据，曾用于脑膜炎奈瑟菌(*Neisseria meningitidis*)的多位点序列分型(Multiple loci sequences typing, MLST)数据[21]以及大肠埃希氏菌(*Escherichia coli*)的全基因组数据[22]分析中。需要注意的是，软件 LDhat 并没有给出重组发生所需的时间，也没有对每个位点每一代的重组率进行估计。在这种情况下，仅能得到重组相对发生频率(ρ)，只有当种群的自发突变率已知的时候才能计算得到绝对重组率(ρ/θ)。因此，在种群的中性突变率相近(可使用 *Tajima's D* 检验[23]和 *F* 检验[24]进行判定)、随机交配水平相近以及种群结构相似的情况下，使用该软件进行不同种系间的重组水平比较才有意义。

2.2　基于系统发育重建的方法进行重组量化

一次重组事件通常会影响一段 DNA 序列, 从而引入多个变异位点。因此, Guttman 和 Dykhuizen 等[25]提出 r/m 值来衡量重组程度。其中 r 是重组引入的变异位点数，m 是自发突变引入的变异位点数。ClonalFrame 软件[26]可以利用系统发育重建的方法估计 r/m、重组片段长度以及相应的 ρ/θ 值。该软件使用贝叶斯方法，综合考虑种系内各进化分支上所发生的突变和重组事件来构建系统发育结构；基于明确

的进化模型并使用马尔科夫-蒙特卡罗法(Markov chain monte carlo, MCMC)进行参数探索。ClonalFrame 软件的优势在于可以重构外源供体导入的序列，不需要分析对象中同时包含重组供体菌株后代和受体菌株后代；ClonalFrame 软件的另一个优势在于可以分别计算出各进化分支上发生的突变和重组事件；此外，通过 ClonalFrame 的计算，可以获得根据贝叶斯方法估计的一致性树，从而重现菌株间更为精确的系统发育关系。Vos 等[27]利用公共 MLST 数据库中的看家基因序列，使用 ClonalFrame 对 48 个物种的 r/m 值进行了分析和比较，发现不同物种间 r/m 值相差可达上千倍(0.02~63.6), 值的大小与物种的分类地位有关，且受到生存环境的强烈影响。

使用贝叶斯和马尔科夫–蒙特卡罗(MCMC)算法进行重组参数探索需要进行大量迭代运算，因此 ClonalFrame 软件使用时会耗费较长时间。运算的迭代次数较小时，通常无法获得可靠结果；而迭代次数较大或进行全基因组序列分析时，其所需运算时间太长(几个月甚至几年)，不利于实际应用。所以，ClonalFrame 常用于以少量基因序列为研究目标的 MLST 数据分析中。

3 重组事件的鉴定工具

重组鉴定主要是根据样本序列上不同片段区域的同源程度进行的，在遗传距离较远的菌株间出现同源性显著提高的片段被看作重组信号。用于检测此类信号的工具包括：PhiPack[28]、DualBrothers[29]和 RDP[30~32]等。

软件包 PhiPack[28]中包含了 3 个不同的重组鉴定方法——*Phi* 检验[28]、、最大 χ^2 检验[33]以及相邻相似性打分(Neighbour similarity score, NSS)[34]。程序可以识别 PHYLIP 格式和 FASTA 格式的输入序列，在进行重组鉴定时，该软件按照设定的滑动窗口大小，对输入序列分成多个窗口分别进行检验，最终给出 p 值以评估某一窗口内序列被判断为重组区域的可靠性。除包含重组鉴定方法外，PhiPack 软件还可以对结果进行绘图展示[35]。

PhiPack 软件的优点是重组鉴定效率高，可以用来处理较大、较为复杂的数据集，但其采用了较简单的模型方法，在重组片段的检出能力及重组断点检测的精确性上低于 DualBrothers 等复杂的方法。

DualBrothers 是一款基于“双重多检验点(Dual multiple change-point, dual MCP)”模型的重组检测方法，该模型通过检测多序列比对中，各位点的拓扑结构和进化速率的变化来提高精确性[29]。统计分析在贝叶斯框架下进行，使用可逆跳跃式 MCMC 采样来验证所有模型参数的后验概率分布。使用该方法可以鉴定出进化历史中的发生重组事件，并可对重组事件的断点进行定位，当数据集中含有重组供体或与供体相似的序列时，还可以对重组片段的来源进行鉴定。为使结果更加准确，在使用时需要满足两个条件：(1)所分析的数据集中含有无重组的参考序列；(2)具有可靠的菌株间系统发育关系作为参考。由于运行时计算资源和时间消耗较大，该软件通常用于处理 MLST 数据或病毒的基因组数据。

为使结果更加可靠，在方法的选择上存在困难时，通常会综合考虑多种方法的判定结果。软件包 RDP 中集成了多种重组鉴定的软件和方法，包括前述的 LDhat、最大 χ^2 检验等。目前，该软件的最新版本是 RDP4，以图形交互界面的操作方式运行于 Windows 平台下，暂不支持 Linux 操作平台；虽然使用方便，但运行效率较低，难以进行批量处理，不适合大规模数据分析。

重组鉴定的另外一个思路是根据基因组上不同区域的变异分布密集程度进行估计[36]。自发突变是随机的，除个别受自然选择压力的区域外，与近缘菌株基因组进行比对，应观察到变异在基因组上的分布是相对均匀的(变异位点的间隔服从指数分布)。但是，由于重组可将远缘菌株的序列引入，因此一次重组事件可能同时引入多个突变，形成突变密集区域。通过比较变异的分布间隔或发生密度，即可推测发生重组事件的区域。基于变异分布进行重组鉴定的软件包括 Gubbins[36, 37]和 RecHMM[10]。Gubbins 中假设未发生重组事件的区域中变异位点分布服从二项分布，以窗口滑动的方式对不同系统发育分支分别进行统计检验，判定不同窗口内的变异位点分布密度。运算速度快，但无法准确判断重组区域的断点。RecHMM 是基于 R 语言的程序，应用隐马尔科夫模型(Hidden markov model, HMM)，采用与 ClonalFrame 相似的算法，利用变异在各个进化分支上的分布进行重组区域的鉴定。与 ClonalFrame 需自己构建拓扑结构相比，该程序使用先验拓扑结构，因此计算速度更快，适用于全基因组序列分析中进

行重组鉴定。

4　重组分析工具的应用

对细菌种群进行重组分析时，考虑到计算资源需求及时间消耗, 可以先选择软件 SplitsTree 构建邻接树或者邻接网络，初步观察样本间的遗传关系，在宏观上把握数据特征，以设计更为精细的分析策略。例如对肺炎克雷伯菌的 MLST 研究中, SplitsTree 分析发现样本明显分化为 3 个主要群体，各个群体之间重组较少，而群体内部分支间重组频繁[38]。因此使用 LDhat 对 3 个群体分别进行了 ρ/θ 的计算和比较。当计算资源允许时，通常使用理论模型更为完善的 ClonalFrame 软件进行重组分析。如在沙门氏菌的种群结构研究中, Xavier 等[39]使用 ClonalFrame 进行计算，同时获得样本的系统发育关系、r/m 值以及各分支上发生的重组事件等遗传多样性结果。需要注意的是，对于重组频率极高的物种，重组片段与自发突变位点会交织在一起，此时很难对重组事件作出正确判定，使用重组事件鉴定工具进行分析时可能会报错。如副溶血弧菌的全基因组分析中，邻接树显示样本间克隆结构几乎完全消失[11]。使用 Haploview 计算 r^2 的结果表明，该物种重组水平已接近有性繁殖种群的重组理论值[40]。因此在后续分析中不再进行重组事件鉴定，而是采用有性繁殖种群的分析理念和工具探索其种群结构和进化规律。

在重组分析的实际应用中经常遇到的问题是：研究对象数据量过大，以致常规软件分析所需时间太长。此时可以根据数据特点和分析目的，对数据集进行适当缩减或拆分以减少资源及时间消耗。比如，若研究对象中存在遗传距离极为相近的克隆群菌株，可以每群只选择一株代表菌株进行重组分析，缩减数据集大小；若研究对象能够可靠划分为多个不同的种系，可以对每个种系分别进行重组分析，通过并行计算减少时间消耗。

5　结语与展望

本文中探讨的都是有效重组事件，即重组的发生引起了遗传物质的改变。在细菌种群中，还有一些重组的发生并未引起遗传物质的改变，被替换的重组片段与新进入的重组片段完全一致。这种不改变遗传信息的重组大多发生在遗传距离非常近的菌株间，是极难被观测和验证的。

进行重组鉴定时，重组事件的检出能力和重组断点判断的准确性与所选用的方法有关[41,42]。表 1

表 1　常用重组分析工具比较

Table 1　Tools used in the analyses of homologous recombination

	软件名称	重要参数	主要算法/模型	优缺点
量化工具	SplitsTree	建树方法；距离模型	邻接网络	优：直观；快捷 缺：无法对重组进行量化
	Haploview	D', r^2	统计检验	优：速度较快；可给出两两位点间连锁程度 缺：不能计算 r/m 值
	LDhat	θ, ρ	哈德森复合似然法，MCMC	优：应用有限位点模型量化重组 缺：耗时较长
	DnaSP	$R = 4Nr$[45]	平均核苷酸差异	优：操作简单，功能多样 缺：只支持 Windows 平台，批量数据处理不便
	ClonalFrame	r/m, ρ/θ	Bayesian, MCMC	优：给出一致性树；可靠性高 缺：耗时长，不宜用于全基因组数据
鉴定工具	PhiPack	Φ_w, Max χ^2, NSS	统计检验	优：速度快，效率高 缺：准确性偏低
	DualBrothers	par_lambda	Dual MCP	优：可定位重组断点 缺：计算资源和时间消耗大
	RDP	—	—	优：操作简单，集成多种软件 缺：只支持 Windows 平台，批量数据处理不便
	RecHMM	region.border, region.peak	HMM	优：优化了 ClonalFrame 的方法，速度快，精度高 缺：需要输入变异位点所处分支
	Gubbins	窗口大小	变异位点分布密度	优：速度快，重组位点判断可靠性高 缺：重组区域断点的推断不精确

对本文所涉及的重组分析工具的参数、模型及优缺点进行了比较。在选择软件或方法时，需根据实际情况进行综合考量，也可结合多个软件的结果进行判断。

测序成本的降低和测序效率的不断提高，为整个基因组范围全面的重组分析创造了条件，但迅猛增长的数据产出量与分析软件的相对滞后也形成了一定程度上的矛盾。部分软件或方法虽然在准确性上满足了要求，但是在计算资源或时间消耗上却不令人满意。因此，为适应大数据分析的需求，仍需要继续开发新的统计方法和相关软件。未来重组分析方法的开发将主要从两个方面入手：一是借助飞速发展的计算科学，设计新的算法，尤其是使用云计算的理念，摆脱单个服务器计算资源的限制，为重组计算带来质的改进；二是进一步了解重组的分子机制以及重组带来的染色体变异效应，从而发现更多重组信号标志，用以建立更加准确的重组分析模型，改善现有方法和软件工具。

参考文献(References)：

[1] Didelot X, Maiden MCJ. Impact of recombination on bacterial evolution. *Trends Microbiol*, 2010, 18(7): 315–322. [DOI]

[2] Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol*, 2005, 3(9): 711–721. [DOI]

[3] Feavers IM, Heath AB, Bygraves JA, Maiden MCJ. Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis*. *Mol Microbiol*, 1992, 6(4): 489–495. [DOI]

[4] Spratt BG, Bowler LD, Zhang QY, Zhou J, Smith JM. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal Neisseria species. *J Mol Evol*, 1992, 34(2): 115–125. [DOI]

[5] Vos M. Why do bacteria engage in homologous recombination? *Trends Microbiol*, 2009, 17(6): 226–232. [DOI]

[6] Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity: a darwinian view of the evolution of microbes. *EMBO Rep*, 2001, 2(5): 376–381. [DOI]

[7] Arnold ML, Sapir Y, Martin NH. Genetic exchange and the origin of adaptations: prokaryotes to primates. *Philos Trans R Soc Lond B Biol Sci*, 2008, 363(1505): 2813–2820. [DOI]

[8] Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol*, 2008, 62: 53–70. [DOI]

[9] Cui YJ, Yu C, Yan YF, Li DF, Li YJ, Jombart T, Weinert LA, Wang ZY, Guo ZB, Xu LZ, Zhang YJ, Zheng HC, Qin N, Xiao X, Wu MS, Wang XY, Zhou DS, Qi ZZ, Du ZM, Wu HL, Yang XW, Cao HZ, Wang H, Wang J, Yao SS, Rakin A, Li YR, Falush D, Balloux F, Achtman M, Song YJ, Wang J, Yang RF. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci USA*, 2013, 110(2): 577–582. [DOI]

[10] Zhou ZM, McCann A, Weill FX, Blin C, Nair S, Wain J, Dougan G, Achtman M. Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years of global spread of enteric fever. *Proc Natl Acad Sci USA*, 2014, 111(33): 12199–12204. [DOI]

[11] Cui YJ, Yang XW, Didelot X, Guo CY, Li DF, Yan YF, Zhang YQ, Yuan YT, Yang HM, Wang J, Wang J, Song YJ, Zhou DS, Falush D, Yang RF. Epidemic clones, oceanic gene pools, and Eco-LD in the free living marine pathogen *Vibrio parahaemolyticus*. *Mol Biol Evol*, 2015, 32(6): 1396–1410. [DOI]

[12] Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*, 2006, 23(2): 254–267. [DOI]

[13] Bandelt H-J, Dress AWM. A canonical decomposition theory for metrics on a finite set. *Adv Math*, 1992, 92(1): 47–105. [DOI]

[14] Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol*, 2004, 21(2): 255–265. [DOI]

[15] Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Masignani V. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol*, 2010, 11(10): R107. [DOI]

[16] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 2005, 21(2): 263–265. [DOI]

[17] McVean G, Auton A. LDhat 2.1: a package for the population genetic analysis of recombination. *Department of Statistics, Oxford, OX1 3TG, UK*, 2007. [DOI]

[18] Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 2003, 19(18): 2496–2497. [DOI]

[19] Hudson RR. Two-locus sampling distributions and their

application. *Genetics*, 2001, 159(4): 1805–1817. [DOI]

[20] McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, 2002, 160(3): 1231–1241. [DOI]

[21] Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC. The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis. *Mol Biol Evol*, 2005, 22(3): 562–569. [DOI]

[22] Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, El Karoui M, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, le Bouguénec C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Médigue C, Rocha EPC, Denamur E. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet*, 2009, 5(1): e1000344. [DOI]

[23] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 1989, 123(3): 585–595. [DOI]

[24] Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*, 1993, 133(3): 693–709. [DOI]

[25] Guttman DS, Dykhuizen DE. Clonal divergence in Escherichia coli as a result of recombination, not mutation. *Science*, 1994, 266(5189): 1380–1383. [DOI]

[26] Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. *Genetics*, 2007, 175(3): 1251–1266. [DOI]

[27] Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J*, 2009, 3(2): 199–208. [DOI]

[28] Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. *Genetics*, 2006, 172(4): 2665–2681. [DOI]

[29] Minin VN, Dorman KS, Fang F, Suchard MA. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics*, 2005, 21(13): 3034–3042. [DOI]

[30] Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, 2000, 16(6): 562–563. [DOI]

[31] Martin DP, Williamson C, Posada D. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, 2005, 21(2): 260–262. [DOI]

[32] Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, 2010, 26(19): 2462–2463. [DOI]

[33] Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol*, 1992, 34(2): 126–129. [DOI]

[34] Jakobsen IB, Easteal S. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. *Comput Appl Biosci*, 1996, 12(4): 291–295. [DOI]

[35] Drouin G, Prat F, Ell M, Clarke GD. Detecting and characterizing gene conversions between multigene family members. *Mol Biol Evol*, 1999, 16(10): 1369–1390. [DOI]

[36] Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD. Rapid pneumococcal evolution in response to clinical interventions. *Science*, 2011, 331(6016): 430–434. [DOI]

[37] Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res*, 2015, 43(3): e15. [DOI]

[38] Guo CY, Yang XW, Wu YR, Yang HY, Han YP, Yang RF, Hu LP, Cui YJ, Zhou DS. MLST-based inference of genetic diversity and population structure of clinical *Klebsiella pneumoniae*, China. *Sci Rep*, 2015, 5: 7612. [DOI]

[39] Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D, Donnelly P. Recombination and population structure in *Salmonella enterica*. *PLoS Genet*, 2011, 7(7): e1002191. [DOI]

[40] Park L. Linkage disequilibrium decay and past population history in the human genome. *PLoS One*, 2012, 7(10): e46603. [DOI]

[41] Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA*, 2001, 98(24): 13757–13762. [DOI]

[42] Chan CX, Beiko RG, Ragan MA. Detecting recombination in evolving nucleotide sequences. *BMC Bioinformatics*, 2006, 7: 412. [DOI]

(责任编委: 刘钢)