



鼠疫耶尔森氏菌的进化研究: 从系统发育学到系统发育基因组学

崔玉军, 宋亚军, 杨瑞馥*

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

* 联系人, E-mail: ruiyuyang@gmail.com

收稿日期: 2012-12-09; 接受日期: 2012-12-25

国家重点基础研究发展计划(批准号: 2009CB522600)、卫生部行业科研专项(批准号: 201202021)和国家科技重大专项(批准号: 2012ZX10004215 和 2013ZX10004221-002)资助

doi: 10.1360/052012-424



摘要 20 世纪 90 年代末起, 基因组学在细菌研究中应用越来越广泛, 尤其在进化领域, 取得了一系列革命性的发现. 本文以鼠疫耶尔森氏菌进化研究为例, 介绍了从利用基因组中少数特定片段(等位基因)多态性进行分析的传统系统发育学, 到基于大量菌株全基因组序列进行系统发育基因组学的研究发展历程, 回顾讨论了基因组学技术的进步为鼠疫菌进化研究领域带来的成果.

关键词

鼠疫菌
进化
基因组学
系统发育学
系统发育基因组学

鼠疫是由鼠疫耶尔森氏菌(*Yersinia pestis*, 以下简称鼠疫菌)引起的一种自然疫源性传染病, 以发病急、传播快、病死率高和传染性强为特征, 对人类社会影响巨大. 历史上曾先后发生过 3 次鼠疫世界大流行: “查士丁尼瘟疫(公元 6 世纪)”、“黑死病(14 世纪中叶~17 世纪中叶)”和“现代鼠疫(19 世纪末~20 世纪中叶)”, 导致至少 1.6 亿人死亡^[1]. 自 1990 年代起, 鼠疫疫情在全球范围内呈上升趋势. 2000 年起, 鼠疫被世界卫生组织(WHO)列为重新抬头的传染病^[2]. 在自然界, 鼠疫菌的主要宿主是啮齿类动物, 通过蚤叮咬等途径, 鼠疫菌可传播给人, 引发腺鼠疫、败血型鼠疫和肺鼠疫. 其中肺鼠疫发病迅猛, 可通过空气在人间直接传播, 如不采取治疗措施病死率高达 100%^[1]. 因此, 鼠疫菌也被认为是最危险的生物恐怖剂之一, 不仅严重影响人类的健康, 对国家安全、社会和谐与经济

发展也构成巨大威胁^[3].

鼠疫菌的进化是其通过遗传变异“重塑”自己的基因组, 从而与环境 and 宿主相互作用的一个动态过程^[4,5]. 通过进化研究了解掌握鼠疫菌遗传变异的规律、形成机理以及地域和时空分布特征, 将有助于制订有效的防治策略, 为鼠疫菌的分型防控、溯源以及保障国家生物安全奠定基础. 进化研究的策略一般包括以下几步: 首先要准确鉴定不同菌株之间发生的各种变异; 然后选择合适的聚类算法, 根据变异状态重建菌株的系统发育关系; 最后通过检测自然选择留下的痕迹, 了解进化的背后推动力和机制, 从而勾勒出整个进化历史. 20 世纪 90 年代兴起的基因组学, 极大带动了鼠疫菌进化研究的发展. 一系列分子生物学方法被开发出来用于获取各类基因组多态性数据, 这些方法应用高效液相色谱、毛细管凝胶电

引用格式: 崔玉军, 宋亚军, 杨瑞馥. 鼠疫耶尔森氏菌的进化研究: 从系统发育学到系统发育基因组学. 中国科学: 生命科学, 2013, 43: 23–30, doi: 10.1360/052012-424
Cui Y J, Song Y J, Yang R F. From phylogenetics to phylogenomics: a brief review to the evolutionary researches of *Yersinia pestis*. SCIENTIA SINICA Vitae, 2013, 43: 23–30, doi: 10.1360/052012-424

泳、Sanger 测序以及全基因组芯片等技术, 可以鉴定鼠疫菌基因组中部分等位基因发生的变异, 为初步推测鼠疫菌种群结构, 以及不同种群菌株间的传播扩散发挥了重要作用^[6-11]. 但由于无法考察全基因组范围的整体变异情况, 这些方法普遍存在分辨率低, 以及有系统发育挖掘偏倚(phylogenetic discovery bias)的缺点, 难以全面客观的反映整个物种的系统发育结构^[12,13].

2005 年开始, 第二代测序技术登上历史舞台, 使人们可以在短时间内, 以相对低廉的成本测定大量菌株的全基因组核酸序列, 为细菌进化研究提供了前所未有的高分辨率和数据完整性. 该技术在霍乱弧菌、耐药性金黄色葡萄球菌、肺炎链球菌等重要致病菌中得到应用, 很好的揭示了这些病原菌的传播和变异规律^[14-17]. 本课题组将新一代测序技术引入到鼠疫菌的进化研究中, 揭示了一些新的现象, 并对已有进化理论做出了补充和完善^[18]. 本文将回顾鼠疫菌从系统发育学(基于基因组少数特定位点变异进行分析)到系统发育基因组学(基于全基因组整体变异)的进化研究历程.

1 基于传统分子生物学方法的鼠疫菌进化研究

1.1 遗传多态性的产生和种类

鼠疫菌以克隆的方式, 通过细胞分裂进行无性繁殖. 在繁殖过程中, 其基因组会以一定的速率发生变异; 这些变异在群体中固定下来, 就形成了具有不同遗传多态性的鼠疫菌种群. 基因组遗传多态性产生的分子策略是多种多样的, 其表现形式可以归纳为以下 5 种类型: (i) 单碱基突变(或称为点突变); (ii) 基因片段拷贝数变化, 某些片段在基因组中出现多个拷贝, 在不同菌株中拷贝数可能发生变化, 这类变异又可分为 3 类: 串联重复(连续的重复片段)、散在重复(重复片段分散在基因组的不同位置, 互不相连)、以及介于 2 者之间的规律成簇的间隔短回文重复(clustered repeats interspaced short palindromic repeats, CRISPR); (iii) 基因片段水平转移, 通过质粒、噬菌体或者一些转座元件, DNA 片段可以通过水平转移(horizontal gene transfer, HGT)的方式在不同细菌基因组之间进行传递, 使得某些群体获得新的

基因片段; (iv) 基因缺失, 细菌生长的一些非必需基因可以在基因组中丢失, 基因片段的丢失在胞内寄生菌中体现最为显著, 使得基因组表现出缩减的趋势; (v) 基因组片段重排, 重排导致基因组片段的相对位置发生变化, 会造成所载荷基因的排列及相互关系改变, 可能为细菌带来新的生物学性状^[19].

下面按照出现时间列出的各种分子生物学方法, 分别针对了不同类型的基因组遗传多态性表现形式, 在已公布的鼠疫菌全基因组序列上寻找特异性变异靶标位点, 通过检测和分析这些位点的变异状态, 推测鼠疫菌的进化历史.

1.2 插入序列(insertion sequence, IS)方法

IS 序列是细菌基因组中可移动的 DNA 元件, 其在基因组中发生转座时, 会促进基因组的转置、重排, 改变基因的排列, 进而导致功能和表型的变化. 鼠疫菌基因组中存在 4 种 IS 序列(IS100, IS285, IS1541 和 IS1661), 各类 IS 序列均有多个拷贝^[20]. 1999 年, Achtman 等人^[21]以 IS100 为探针对 49 株鼠疫菌进行了限制性片段长度多态性(restriction fragment length polymorphisms, RFLP)分析, 通过构建邻接树, 将基于 IS100 多样性的鼠疫菌系统发育关系与传统的生物型(基于生化表型的鼠疫菌分型系统)进行了比较. 系统发育树显示, 3 个已知生物型从 1 个共同祖先生出来, 且形成相互隔离的不同分支. 根据各生物型菌株所在的分支长度及表现出的 IS 重排复杂性, 可以推测古典型和中世纪型菌株起源较早, 而东方型菌株多样性低, 分支长度最短, 说明其起源最晚. IS 序列介导的重排还可以将基因打断, 形成假基因. 本课题组在已测序基因组中鉴定出由 IS 重排造成的 24 个假基因位点, 使用 PCR 扩增方法考察了这些假基因在 260 株鼠疫菌中的分布^[22]. 结果显示, 田鼠型菌株(分离自青海和内蒙)以及新疆的分离株的系统发育地位比较古老. 与 Achtman 等人^[21]的结论不同, 本课题组发现中世纪型菌株和东方型菌株均衍生自古典型菌株, 该结果后来得到多项证据的支持.

1.3 差异区段(different region, DFR)方法

DFR 指在同一物种里, 某些菌株基因组中存在而在另一些菌株中缺失的基因片段, 其多态性显示了基因组片段水平转移和缺失的状态. Hinchliffe 等人^[23]最早通过比较基因组芯片杂交的方法, 在鼠疫

菌中发现了 16 个此类片段. Zhou 等人^[24]使用包含 4005 个基因探针的鼠疫菌全基因组芯片, 对 43 株鼠疫菌和 7 株假结核耶尔森氏菌的 DNA 进行杂交, 鉴定出 22 个 DFR 位点. Dai 等人^[25]通过抑止消减杂交技术发现了天山山地灰旱獭(*Marmota baibacina*)、长尾黄鼠(*Citellus undulatus*)鼠疫自然疫源地分离株所特有的、长约 383 bp 的 DNA 片段, 称之为 DFR23, 进一步完善了鼠疫菌 DFR 谱. 本课题组采用 PCR 技术分析了这 23 个 DFRs 在 909 株鼠疫菌自然分离株中的分布, 据此将中国鼠疫菌分成了 32 个基因组型. 将菌株的基因组型与地理分布进行关联分析后, 本课题组发现, 基因组型的分布具有明显的疫源地特异性, 多数疫源地拥有占优势数量的基因组型, 将其命名为主要基因组型(major genomovar), 主要基因组型可以作为某特定疫源地的特征基因组型; 同时, 种类多样, 但数量稀少的一些基因组型(命名为次要基因组型, minor genomovar)可以在同一疫源地与主要基因组型并存. 根据简约进化的理论, 本课题组推测了各基因组型鼠疫菌之间的进化关系, 并进一步阐明了第 3 次鼠疫世界大流行的起源和传播过程. DFR 数据表明, 引起第 3 次大流行的东方型菌株最早起源于中国云南的鼠疫自然疫源地, 然后扩散到广东、福建等地, 最后才传播到美国及世界其他地区^[8].

1.4 基因组板块(genome plate)重排方法

在鼠疫菌中, 大量 IS 序列及其他重复元件(如核糖体 RNA 操纵子)将整个基因组分隔成一系列相对位置可以改变的片段, 这些基因组片段在 Liang 等人^[26]的研究中被形象比喻为地质学上的“板块”. 在不同菌株间, 同一“板块”内部序列具有高度同源性和共线性, 而不同“板块”之间则存在大量的重排现象, 这使得鼠疫菌基因组以一种动态的状态存在. Darling 等人^[27]通过对两株假结核耶尔森氏菌和 6 株鼠疫菌全基因组序列之间比对, 鉴定出 78 个基因组“板块”; 然后利用贝叶斯方法分析了这些“板块”之间的重排历史, 并据此用最大简约法重建了 8 株菌的系统发育关系. 尽管未得到具有唯一拓扑结构的系统发育树, Darling 等人^[28]构建的网状树仍然与基于 SNP 构建的树所反映的鼠疫菌进化模式一致, 说明某些“板块”重排事件被固定在鼠疫菌的特定种系中, 这也提示可以通过基因组“板块”重排方法研究鼠疫菌的进化历史. 由于分析全基因组范围的“板块”重排时, 需要

使用达到完成图标准的基因组组装结果, 才能准确判定各“板块”之间的排列顺序; 此外, 某些 IS 序列在鼠疫菌中介导了非常高频的重排, 例如同菌株的不同批次培养物中某些位点的重排模式也存在不同, 在一定程度上影响了结果的重现性^[20]; 因此目前基于“板块”重排的鼠疫菌进化研究相对较少.

1.5 间区规律聚积的短回文重复(clustered regularly interspaced short palindromic repeats, CRISPR)方法

CRISPR 是一类结构特殊的重复序列, 由 21~47 bp 的同向重复序列和将它们隔离开来、大小与之近似的间区序列(spacer)组成^[29]. 鼠疫菌中总共有 3 个 CRISPR 位点, 本课题组与法国和俄罗斯研究人员合作, 在 109 株鼠疫菌的 CRISPR 位点中鉴定出了 29 个间区序列, 证明 CRISPR 位点多态性较高, 适用于鼠疫菌分型和进化研究^[30]. 随后收集了分离自中国 12 个疫源地以及前苏联和外蒙古 14 个源地的鼠疫菌共 125 株, 结合已有的 7 株全基因组序列, 充分分析了鼠疫菌 CRISPR 位点的遗传多样性, 新鉴定出 131 个 spacer, 并综合 3 个 CRISPR 位点上 spacer 阵列的排列规律, 构建了鼠疫菌进化模型. 该模型显示, 存在一条古老型鼠疫菌的传播路线, 该路线围绕塔克拉玛干沙漠和准葛尔盆地, 其 2 个分支之一穿越青藏高原、昆仑山脉, 到塔吉克斯坦; 另一条通过帕米尔高原、天山山脉、阿尔泰山脉到内蒙古高原. 根据模型推测, 鼠疫菌相对年轻的分支可能从沿该路线的某些特定区域衍生出来, 并传播到世界其他地区, 导致了新鼠疫自然疫源地的形成^[6].

1.6 多位点串联重复序列分析(multiple loci VNTR analysis, MLVA)方法

可变数目串联重复(variable number of tandem repeat, VNTR)是连续方式存在的重复序列, 在不同菌株中同一位点的重复数可能会发生变化. VNTR 在细菌基因组中分布极为丰富, 因此, 可以同时利用多个 VNTR 位点的变异信息, 即 MLVA 方法, 进行分型和进化分析. 2004 年, Girard 等人^[31]使用 43 个 VNTR 位点的 MLVA 方案, 对美国亚利桑那州一片约 350 km² 的鼠疫疫源地分离菌株进行了分析. 实验结果表明, VNTR 位点的体内和体外突变速率基本一致, 因此可以根据 VNTR 的实验室突变速率建立鼠

疫苗传播模型。模型表明, 鼠疫传播会经历 2 个主要阶段, 首先是种群数量增长造成的快速扩张期, 随后会进入传播速度减慢的一个持续期。这种模式可能适用于鼠疫在局部地区乃至全球范围的鼠疫生存和传播规律。2004 年, Pourcelet 等人^[32]通过对 5 株耶尔森氏菌(包括 3 株鼠疫菌和 2 株假结核菌)的研究发现, 有 25 个 VNTR 位点在鼠疫菌中表现出了多态性。由于其中 2 个位点在多株中国各疫源地代表性鼠疫菌株中发生缺失, 本研究组选定其余 23 个位点, 对 500 余株来自中国及世界其他疫源地的鼠疫菌进行了 MLVA 分析。所有菌株可以被分成 350 个不同的基因型, 表明 MLVA 方法具有比较高的分辨率, 许多无法用 DFR 或者 CRISPR 方法区分的菌株, 可以被该方法明确分开。该研究根据遗传多样性将鼠疫菌分为古老分支、过渡分支和现代分支 3 个大群, 并进一步确定了各类型鼠疫菌的系统发育地位: 田鼠型菌株(包括中亚地区分离的 *Pestoides* 菌株)属于鼠疫菌的最古老分支之一, 这与 CRISPR 方法得到的古老鼠疫菌传播结论一致, MLVA 结果也表明古老鼠疫菌的早期扩散形成了现在的疫源地格局; 新疆地区的鼠疫分离株属于古老分支和现代分支之间的过渡型菌株, 该地区非常有可能是鼠疫菌现代分支的发源地; 鼠疫菌现代分支又可以分为 2 个大亚群, 中世纪型菌株和东方型菌株分别从这 2 大亚群的古典型菌株(多数分离自中国或中亚地区的疫源地)中衍生出来后, 传播到世界各地。与基于 DFR 的推测相吻合, MLVA 分析同样支持东方型菌株最早起源于中国云南^[7]。

1.7 单核苷酸多态性(single nucleotide polymorphism, SNP)方法

作为一种遗传标记, SNP 具有分布广泛、遗传稳定、适于快速高通量检出等优点, 已经成为研究微进化的金标准。2004 年, Achtman 等人^[33]从 3 株鼠疫菌全基因组序列中, 筛选出 76 个同义 SNP(同义突变不改变所编码的氨基酸, 一般被认为中性突变, 能够更为精确地反应进化历程), 并最终选取 44 个同义 SNPs, 采用变性高效液相技术对 105 株鼠疫菌进行了分析。该研究确定了鼠疫菌的 3 大分支和 8 个种群, 并提出了沿用至今的鼠疫菌系统发育种群标准命名法(如 2.ANT3a, “2”表示该菌株属于 2 号分支, “ANT”在最初命名时表示菌株的生物表型, 但后来研究发现这样命名的基因型与表型并不完全吻合; 接下来的“3”

表示 2.ANT 型菌株中的第 3 群, “2.ANT3”就构成了 1 个鼠疫菌种群的名称, 最后的字母“a”表示该种群中的某菌株编号, 用于区分具体菌株)。虽然该研究明确了整个鼠疫菌物种的进化框架结构, 但仍存在变异位点少, 分辨率低的缺点。为了弥补这一缺点, 本研究组与 Achtman 等人^[34]合作, 重新对 17 株多样性更为广泛的鼠疫菌全基因组序列进行了比较分析, 鉴定出 933 个 SNP, 并通过质谱方法考察了 286 株鼠疫菌全球分离株在这些位点的变异情况。这项研究进一步验证了 2004 年获得的进化框架的准确性, 同时发现了更多的鼠疫菌种群; 并且证明, 鼠疫菌可能起源于中国或者中国邻近地区, 在历史上通过多次大规模扩散事件, 传播到欧洲、南美洲、非洲和东南亚, 形成国家特异性的鼠疫菌进化分支。研究表明, 所有分离自美国的鼠疫菌都起源于一次传播, 这次传播事件可以追溯到 19 世纪末, 一艘从中国开出, 经由夏威夷, 最终到达加利福尼亚州的商船。感染鼠疫菌的宿主动物可能随这艘商船来到美国, 鼠疫菌在当地定植下来并形成广泛的自然疫源地。该研究还通过 82 株马达加斯加分离株的 SNP 多样性, 首次计算了鼠疫菌的自然突变率, 并使用该突变率推算了各进化分支的出现时间。

2 基于新一代全基因组测序的鼠疫菌进化研究

新一代高通量测序技术的迅速发展, 使得全面考察所有目标菌株在全基因组范围的变异成为可能。为了更加深入的探查鼠疫菌的进化规律, 本研究组选取了 118 株鼠疫菌, 应用 Illumina 第二代测序技术进行深度测序, 并将测序结果与以前发表的 15 株鼠疫菌全基因组序列和 1 株假结核菌株序列进行了系统发育基因组学分析^[18]。

使用全基因组范围的 SNP, 可以获得更高解析度的鼠疫菌种系发育树。通过发育树的拓扑结构, 我们鉴定出鼠疫菌的 2 个全新主要分支, 与以前的 3 个分支一起共同构成 5 大种系分支(Branch 0~4)的种群结构。系统发育树各分支的分化时间以及黑死病人残骸中得到的古 DNA 基因组^[35]变异均表明, 4 个年轻种系分支的形成可能与中世纪鼠疫大流行密切相关。将系统发育关系与菌株地理分布相关联(系统发育地理学分析, phylogeographical analysis), 可以对鼠疫菌

的历史传播情况进行推测. 分析结果显示, 中国古代商路(丝绸之路、唐蕃古道和茶马古道)与鼠疫菌的地理分布存在惊人的一致性, 说明历史上人类商贸活动在鼠疫的传播中发挥了重要作用. 由于青藏高原东部地区菌型复杂, 包含了 5 大分支中 4 个分支的菌株, 且包含目前发现的最古老鼠疫菌种群(0.PE7), 提示该地区可能是鼠疫菌物种的发源地.

以前研究认为, 鼠疫菌的进化与环境适应有关, 即受到自然选择压力影响. 但鼠疫菌全基因组以及系统发育树上各个分支节点 SNP 的 d_N/d_S 计算结果均接近 1, 说明从全基因组水平来讲, 鼠疫菌以中性进化为主. 对单个基因选择压力检测结果显示, 在核心基因组的 3450 个基因中, 仅有 22 个(0.6%)基因上存在正向选择的信号. 这进一步支持鼠疫菌以中性进化为主的结论, 证明遗传漂变是鼠疫菌进化的主要推动力^[18].

基于全基因组变异构建的鼠疫菌系统发育树存在一个明显特征, 即各分支长度非常不一致, 显著偏离泊松分布, 这提示鼠疫菌在进化过程中的变异累积速率可能发生了变化. 利用贝叶斯方法的统计分析结果也证实了该观点, 表明鼠疫菌在进化过程中发生了多次变异累积速度变化事件. 本研究组还发现, 突变累积速率与各菌株到共同祖先的节点数目存在强的正线性相关, 说明突变累积速率的变化不仅影响系统发育树的分支长度, 同时还影响了种系分化的速度. 鉴于鼠疫菌以中性进化为主, 这项观测补充完善了分子钟理论, 说明即使在中性进化的前提下, 细菌的突变累积速率在进化过程中仍然可能发生变化. 我们推测这种变化可能与鼠疫菌的种群大小变化有关, 在暴发期有效种群数目显著增大, 造成变异累积加速; 而静息期种群大小保持稳定, 突变累积速率相对降低.

3 展望

随着基因组学的发展进步, 鼠疫菌的进化研究也在不断取得新的认识和成果. 目前, 该物种的基本种群结构已经明确, 各种基因型与地理分布的关系

也认识的比较清楚. 许多进化研究中鉴定出的变异位点, 已经成为基层实验室在分离、鉴定、保藏鼠疫菌株时的标准分型方法. 其中 DFR 和 MLVA 方法在多次鼠疫疫情暴发时, 成功应用于菌株溯源, 有力保障了疫情的预防控制工作^[36,37]. 但在鼠疫菌研究领域仍然有一个缺憾: 前苏联地区存在面积广大、历史悠久的鼠疫自然疫源地, 且西伯利亚地区一度被怀疑为鼠疫的起源地. 但该地区鼠疫菌分离株仅有 3 株进行了全基因组测序^[18,38], 因此该地区鼠疫菌株的遗传多样性无法得到全面分析. 以后需要通过更积极的国际交流合作, 弥补鼠疫菌进化研究中这缺失的一环.

传统的基于基因组特定片段多态性的分析方法, 在认识鼠疫菌的种群结构和进化历史上, 起到了重要作用. 虽然其内在的缺陷, 包括低分辨率或者无法得到稳定的系统发育关系, 是方法本身所不可避免的, 但这些方法操作相对简单, 容易在常规分子生物学实验室普及. 另外, 目前已经建立了一些国际数据库, 用于存储和管理这些传统方法所产生的生物学数据, 方便了国际不同实验室之间的标准化数据比对与交流^[39]; 同时也有利于鼠疫疫情暴发或者遭受生物恐怖袭击时, 利用数据库中的数据对菌株来源进行追踪. 因此传统系统发育分析技术和方法在以后一段时间内将仍然会得到广泛应用.

基于所有目标样本全基因组序列的系统发育基因组学分析, 在进化研究的解析度和准确度上, 有着无可比拟的优势. 目前, 第三代测序技术日趋成熟, 并在实践中得到成功应用^[40]. 该技术将全基因组测序时间进一步缩短到几个小时, 成本可以控制在几百元人民币甚至更低. 可以想象, 随着技术普及和成本继续下降, 将来新一代测序仪必然会走入常规的分子生物学实验室, 而基于全基因组序列的系统发育学将成为进化研究的主流. 随之而来的问题是, 将有越来越多菌株的全基因组序列被测定, 对这些海量数据进行分析时, 巨大的计算资源消耗以及缺乏相应的统计分析软件会成为分析的瓶颈. 因此, 需要通过计算生物学和生物信息学手段, 对这些问题提前加以研究, 以应对将来的系统发育基因组学挑战.

参考文献

- 1 Perry R D, Fetherston J D. *Yersinia pestis*—etiologic agent of plague. Clin Microbiol Rev, 1997, 10: 35–66

- 2 Organization W H. Human plague in 2002 and 2003. *Wkly Epidemiol Rec*, 2004, 79: 301–306
- 3 Riedel S. Plague: from natural disease to bioterrorism. *Proc Bayl Univ Med Cent*, 2005, 18: 116–124
- 4 Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol*, 2008, 6: 431–440
- 5 Welch T J, Fricke W F, McDermott P F, et al. Multiple antimicrobial resistance in plague: an emerging public health risk. *PLoS ONE*, 2007, 2: e309
- 6 Cui Y, Li Y, Gorge O, et al. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE*, 2008, 3: e2652
- 7 Li Y, Cui Y, Hauck Y, et al. Genotyping and phylogenetic analysis of *Yersinia pestis* by MLVA: insights into the worldwide expansion of Central Asia plague foci. *PLoS ONE*, 2009, 4: e6000
- 8 Li Y, Dai E, Cui Y, et al. Different region analysis for genotyping *Yersinia pestis* isolates from China. *PLoS ONE*, 2008, 3: e2166
- 9 Maiden M C, Bygraves J A, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA*, 1998, 95: 3140–3145
- 10 Roumagnac P, Weill F X, Dolecek C, et al. Evolutionary history of *Salmonella typhi*. *Science*, 2006, 314: 1301–1304
- 11 Woese C R. Bacterial evolution. *Microbiol Rev*, 1987, 51: 221–271
- 12 Comas I, Homolka S, Niemann S, et al. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE*, 2009, 4: e7815
- 13 Pearson T, Busch J D, Ravel J, et al. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci USA*, 2004, 101: 13536–13541
- 14 Croucher N J, Harris S R, Fraser C, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science*, 2011, 331: 430–434
- 15 Harris S R, Feil E J, Holden M T, et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, 2010, 327: 469–474
- 16 Holt K E, Parkhill J, Mazzoni C J, et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella typhi*. *Nat Genet*, 2008, 40: 987–993
- 17 Mutreja A, Kim D W, Thomson N R, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*, 2011, 477: 462–465
- 18 Cui Y, Yu C, Yan Y, et al. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci USA*, 2012, 110: 557–582
- 19 Zhou D, Han Y, Song Y, et al. Comparative and evolutionary genomics of *Yersinia pestis*. *Microbes Infect*, 2004, 6: 1226–1234
- 20 Parkhill J, Wren B W, Thomson N R, et al. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, 2001, 413: 523–527
- 21 Achtman M, Zurth K, Morelli G, et al. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA*, 1999, 96: 14043–14048
- 22 Tong Z, Zhou D, Song Y, et al. Pseudogene accumulation might promote the adaptive microevolution of *Yersinia pestis*. *J Med Microbiol*, 2005, 54: 259–268
- 23 Hinchliffe S J, Isherwood K E, Stabler R A, et al. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Genome Res*, 2003, 13: 2018–2029
- 24 Zhou D, Han Y, Song Y, et al. DNA microarray analysis of genome dynamics in *Yersinia pestis*: Insights into bacterial genome microevolution and niche adaptation. *J Bacteriol*, 2004, 186: 5138–5146
- 25 Dai E, Tong Z, Wang X, et al. Identification of different regions among strains of *Yersinia pestis* by suppression subtractive hybridization. *Res Microbiol*, 2005, 156: 785–789
- 26 Liang Y, Hou X, Wang Y, et al. Genome rearrangements of completely sequenced strains of *Yersinia pestis*. *J Clin Microbiol*, 2010, 48: 1619–1623
- 27 Darling A E, Miklos I, Ragan M A. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet*, 2008, 4: e1000128
- 28 Chain P S, Hu P, Malfatti S A, et al. Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J Bacteriol*, 2006, 188: 4453–4463
- 29 崔玉军 李艳君 颜焱锋, 等. 规律成簇的间隔短回文重复: 结构、功能与应用. *微生物学报*, 2008, 48: 1549–1555
- 30 Vergnaud G, Li Y, Gorge O, et al. Analysis of the three *Yersinia pestis* CRISPR loci provides new tools for phylogenetic studies and possibly for the investigation of ancient DNA. *Adv Exp Med Biol*, 2007, 603: 327–338
- 31 Girard J M, Wagner D M, Vogler A J, et al. Differential plague transmission dynamics determine *Yersinia pestis* population genetic structure on local, regional and global scales. *Proc Natl Acad Sci USA*, 2004, 101: 8408–8413

- 32 Pourcel C, Andre-Mazeaud F, Neubauer H, et al. Tandem repeats analysis for the high resolution phylogenetic analysis of *Yersinia pestis*. BMC Microbiol, 2004, 4: 22
- 33 Achtman M, Morelli G, Zhu P, et al. Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc Natl Acad Sci USA, 2004, 101: 17837–17842
- 34 Morelli G, Song Y, Mazzoni, C J, et al. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. Nat Genet, 2010, 42: 1140–1143
- 35 Bos K I, Schuenemann V J, Golding G B, et al. A draft genome of *Yersinia pestis* from victims of the Black Death. Nature, 2011, 478: 506–510
- 36 祁芝珍, 戴二黑, 周冬生, 等. 青海省 2004 年人间鼠疫分离株基因分型及流行病学意义. 中华流行病学杂志, 2006, 27: 316–318
- 37 Wang H, Cui Y, Wang Z, et al. A dog-associated primary pneumonic plague in Qinghai Province, China. Clin Infect Dis, 2011, 52: 185–190
- 38 Garcia E, Worsham P, Bearden S, et al. Pestoides F, an atypical *Yersinia pestis* strain from the former Soviet Union. Adv Exp Med Biol, 2007, 603: 17–22
- 39 Le Fleche P, Hauck Y, Onteniente L, et al. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC Microbiol, 2001, 1: 2
- 40 Rohde H, Qin J, Cui Y, et al. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. N Engl J Med, 2011, 365: 718–724

From Phylogenetics to Phylogenomics: a Brief Review to the Evolutionary Researches of *Yersinia pestis*

CUI YuJun, SONG YaJun & YANG RuiFu

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

Since 1990s, genomics has been serving a more and more important role in bacterial researches, especially in the evolutionary studies of pathogens. Numerous novel findings in this area are achieved with the accumulation of the flooding genomics data. Here we focused on the history of evolutionary researches of an important human pathogen, *Yersinia pestis*, which undergoes dramatic changes from the traditional phylogenetics analysis based on polymorphisms of specific alleles in genomes, to the phylogenomics studies involving whole genome-wide variations of large amount of strains.

***Yersinia pestis*, evolution, genomics, phylogenetics, phylogenomics**

doi: 10.1360/052012-424



崔玉军, 博士, 副研究员. 目前主要从事细菌基因组学、群体遗传学和生物信息学研究. 在 *New England Journal of Medicine*, *Proceedings of the National Academy of Sciences of the United States of America* 和 *Clinical Infectious Disease* 等杂志发表 SCI 论文 20 余篇. 主持国家自然科学基金、卫生部行业科研专项、国家科技重大专项等多项课题.



杨瑞馥, 博士, 研究员. 国家杰出青年科学基金获得者, 国家重点基础研究发展计划项目首席科学家. 主要从事细菌基因组学、进化与致病机制和微生物法医学基础数据库的研究. 近年来, 在 *New England Journal of Medicine*, *Nature*, *Nature Genetics* 和 *Proceedings of the National Academy of Sciences of the United States of America* 等杂志发表学术 SCI 论文 100 余篇, 他引超过 1600 次, H 因子 25, 主编专著 6 部. 学术兼职: 中国微生物学会副理事长, 中国微生物学会分析微生物学专业委员会主任委员, 卫生部疾病防控专家委员会委员; 卫生部突发事件卫生应急专家咨询委员会委员; 卫生部食品安全委员会委员、国家和军队处置生物恐怖袭击事件专家咨询组成员; *Journal Biodefense and Bioterrorism* 编委, 《人兽共患病杂志》、《微生物学学报》和《中华预防医学杂志》副主编.