

復旦大學

本科毕业论文



论文题目：肠道沙门氏菌 *enterica* 亚种的群体结构和进化历史分析

姓 名：王悠扬

学 号：20307110361

院 系：生命科学学院

专 业：生物科学

指导教师：Daniel Falush

职 称：教授

单 位：中国科学院上海免疫与感染研究所

完成日期：2024 年 5 月 20 日

肠道沙门氏菌 *enterica* 亚种的群体结构和进化历史分析

完成人

王悠扬

指导小组成员

Daniel Falush 教授

目 录

摘要	I
Abstract	II
一、前 言	1
1.1 研究背景	1
1.1.1 肠道沙门氏菌概述	1
1.1.2 肠道沙门氏菌的群体结构与进化历史研究	1
1.2 主要研究内容及研究意义	2
二、材料与amp;方法	4
2.1 菌株基因组数据收集	4
2.2 序列比对和核心 SNP 位点识别	4
2.3 群体结构划分	4
2.3.1 基于 fineSTRUCTURE 的共同祖先分析	5
2.3.3 SNP 距离计算	5
2.4 进化树构建	5
2.5 进化历史分析	5
2.5.1 连锁不平衡衰减计算	5
2.5.2 主成分分析和 SNP 载荷	5
2.5.3 染色体涂绘	6
2.5.4 基于 TreeMix 的基因流动分析	6
三、研究结果	7
3.1 肠道沙门氏菌 <i>enterica</i> 亚种的群体结构	7

3.1.1 群体中的菌株来源	7
3.1.2 群体结构的初步划分	8
3.1.2 群体结构的验证	9
3.2 群体的进化历史分析	10
3.2.1 进化树的构建和标记	10
3.2.2 不同群组内部的连锁不平衡分析	11
3.2.3 PCA 和全基因组 SNP 载荷分析	12
3.2.4 基于染色体涂绘的祖源概况分析	13
3.2.5 基因流动分析	14
四、讨论	16
4.1 研究总结	16
4.2 局限性与展望	16
参考文献	18
附表 本研究使用的 <i>enterica</i> 菌株背景信息	20
致谢	41

Abstract

Salmonella is a common foodborne pathogen, and the leading cause of human infections is the *Salmonella enterica* subspecies *enterica*, which has rare recombination with other subspecies, while recombination within the subspecies is very common. Recent studies have shown an internal divergence of subspecies *enterica*. This study focused on the population structure and evolutionary history of subspecies *enterica*. We downloaded the genomes of 303 *enterica* strains from public databases and identified 478,732 core SNP sites from the aligned sequences. We then calculated the population structure of subspecies *enterica* by using fineSTRUCTURE and pairwise SNP distances. Based on the population structure, we divided *enterica* into eight clusters. Subsequently, we inferred the evolutionary history within the subspecies by linkage disequilibrium decay, PCA, SNP loading calculation, and chromosome painting. Our study showed that the evolutionary pattern of *enterica* couldn't be explained by a simple binary fission model, and it consisted of two clonal extreme clusters and six intermediate clusters, with three of the eight clusters experiencing strong internal genetic drift. This study helps us to better understand the evolutionary driving forces of *enterica* and other pathogenic microorganisms.

Key words: *Salmonella enterica*, population structure, genetic drift, PCA, evolution analysis

一、前言

1.1 研究背景

1.1.1 肠道沙门氏菌概述

沙门氏菌 (*Salmonella*) 是肠杆菌科的兼性厌氧型革兰氏阴性菌, 有鞭毛, 表现出较强的运动能力。作为一种常见的食源性致病菌, 沙门氏菌是仅次于弯曲杆菌 (*Campylobacter* spp.) 的主要人类肠胃病原菌, 因此沙门氏菌是受到研究最深入的微生物之一^[1]。近一个世纪以来, 根据沙门氏菌鞭毛蛋白或 O 抗原与特异性抗体之间的血清反应, 沙门氏菌被分为 2500 多个血清型, 其中一些血清型与致病性直接相关, 比如鼠伤寒血清型 (Typhimurium) 和肠炎血清型 (Enteritidis) 会导致肠炎等疾病^[2-4]。随着分子生物学的发展, 根据 DNA-DNA 杂交结果, 沙门氏菌属被分为邦格里沙门氏菌 (*Salmonella bongori*) 和肠道沙门氏菌 (*Salmonella enterica*) 2 个种, 肠道沙门氏菌又被分为 6 个亚种 (*enterica*、*salamae*、*arizonae*、*diarizonae*、*houtenae* 和 *indica*)^[5]。流行病学研究表明, 邦格里沙门氏菌主要感染冷血动物, 而 99% 的人类感染是由肠道沙门氏菌的 *enterica* 亚种造成的^[6, 7]。基于此, 公共数据库中收录了大量沙门氏菌感染事件中采集的 *enterica* 亚种的基因组信息, 另外 5 个肠道沙门氏菌亚种的基因组信息和相关研究则明显少于 *enterica* 亚种。

1.1.2 肠道沙门氏菌的群体结构与进化历史研究

群体遗传学关注进化驱动力对个体变异、群体结构和物种进化的影响。与真核生物相比, 细菌作为单倍体生物主要进行无性繁殖, 物种边界相对模糊, 造成其遗传多样性的主要进化驱动力包括突变 (mutation)、重组 (recombination)、遗传漂变 (genetic drift) 和自然选择 (natural selection), 重组事件发生在不同群体之间则造成了群体间的基因流动 (gene flow)^[8, 9]。

为了更好地揭示病原微生物的群体结构, 研究人员优化了传统的基于多位点酶电泳技术 (multilocus enzyme electrophoresis, MLEE) 的分型方法, 直接依据多个位点的核苷酸序列对细菌进行多位点序列分型 (multilocus sequence typing, MLST)^[10]。2012 年, Achtman 等使用了肠道沙门氏菌的 7 个管家基因

片段对其进行多位点序列分型，这种肠道沙门氏菌的分型方法被称为传统多位点序列分型（legacy MLST），该研究根据分型结果将肠道沙门氏菌的 *enterica* 亚种定义成 138 个爆发群（eBurst groups, eBG），每个爆发群中的菌株仅存在一个等位基因的差异^[4]。近年来，还出现了基于 3002 个基因位点的核心基因组 MLST（core genome MLST, cgMLST）、基于 51 个核糖体基因位点的核糖体基因 MLST（ribosomal gene MLST, rMLST）以及基于全基因组共 21065 个位点的 MLST（whole genome MLST, wgMLST），根据 rMLST，研究人员定义了对应的核糖体爆发群（ribosomal eBurst groups, reBG）^[11-13]。伴随基因组学的发展，现今肠道沙门氏菌的群体遗传学及群体结构研究常使用全基因组或泛基因组的单核苷酸多态性（single nucleotide polymorphism, SNP）信息，MLST 方法则多用于研究中的冗余菌株（遗传组成基本相同的菌株）筛选和初步分类^[3, 14]。

传统观点认为沙门氏菌是高度克隆性的细菌，然而基于 DNA 序列的分子进化研究挑战了这一观点，这是由于 *enterica* 亚种内部的某些基因呈现不一致的状态^[15]。近年来研究人员利用不同方法构建了不同的沙门氏菌系统发生树，这意味着沙门氏菌的进化模式比传统认为的不可逆二分裂更为复杂，无法通过单独的系统发生树进行追溯^[14, 16, 17]。沙门氏菌和其他原核生物很容易通过同源或非同源重组与外来 DNA 进行重组，这与有性生殖中的重组方式截然不同，*enterica* 亚种的内部的水平基因转移（horizontal gene transfer, HGT）事件十分常见，而其与其他亚种之间的重组则十分罕见^[15, 18, 19]。Criscuolo 等和 Alikhan 等基于全基因组序列的研究表明，肠道沙门氏菌的 *enterica* 亚种出现了内部分化，这可能说明这一亚种中存在较为活跃的进化事件^[3, 14]。

1.2 主要研究内容及研究意义

本研究针对肠道沙门氏菌造成人类感染最多且最经常被研究的 *enterica* 亚种（后文简称为 *enterica*），采用 fineSTRUCTURE、成对 SNP 距离及主成分分析（Principal Component Analysis, PCA）等方法，利用 SNP 信息在全基因组尺度上分析其群体结构，探究 *enterica* 是否已经分化成两个不同亚群。在此基础上，我们采用连锁不平衡（Linkage Disequilibrium Decay, LD）分析、SNP 载荷（SNP loading）计算、基因流动分析等多种群体遗传学方法分析 *enterica* 的进化历史，探讨形成现存群体结构的进化机制。

作为常见的病原微生物，*enterica* 的进化可能与其对宿主的适应有关^[19]，这意味着 *enterica* 的进化可能影响到其耐药性，因此加深对 *enterica* 进化机制的理解可以为防控 *enterica* 感染造成的疫情提供理论基础。此外对 *enterica* 群体结构和进化历史的研究有助于加深我们理解原核生物群体结构保持和变化的分子机制。

二、材料与方法

2.1 菌株基因组数据收集

本研究共收集了 311 个肠道沙门氏菌 *enterica* 亚种菌株的全基因组序列，其中 295 个菌株来自公共数据库 EnteroBase^[20]，我们从 2018 年 Alikhan 等^[3]报道的沙门氏菌数据集（https://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:9641）中下载了全部的 *enterica* 菌株数据；另外 16 个菌株来自欧洲核苷酸档案库（European Nucleotide Archive, ENA, <https://www.ebi.ac.uk/ena/>），由 Criscuolo 等^[14]完成测序并上传。基于 SNP 距离的计算（见 2.3.2）表明，来自两个不同数据库的菌株有 8 株重合（SNP 距离等于 0），其余菌株之间的 SNP 距离均大于 2 kb，由此我们得到了包含 303 个菌株的非冗余 *enterica* 基因组数据集（图 1C）。

2.2 序列比对和核心 SNP 位点识别

我们使用基于 MUMmer v3.20^[21]的算法对 303 个菌株基因组进行序列比对和 SNP 位点识别。首先通过 MUMmer 软件的 nucmer 命令将所有菌株的基因组序列与 *enterica* 的参考基因组 ASM694v2（RefSeq 编号：GCF_000006945.2）序列进行比对，比对获得的基因组长度为 4.74 Mb。随后用 SNP-sites v2.5.2^[22]从比对后的序列中识别 SNP 位点，获得包含所有 SNP 信息的 VCF 文件。最后用 VCFtools v0.1.16^[23]筛选出 478732 个至少存在于 99%的基因组的 SNP 位点，我们将这些位点被定义为用于后续分析的核心 SNP 位点。

2.3 群体结构划分

2.3.1 基于 fineSTRUCTURE 的共同祖先分析

为了识别上述 *enterica* 菌株的群体结构，我们用 fineSTRUCTURE v0.1.0^[24]计算所有菌株的共同祖先矩阵（coancestry matrix），该软件将 303 个菌株 SNP 矩阵的变形文件为输入文件，首先基于 ChromoPainter（见 2.7）计算每个供体（donor）菌株向受体（recipient）菌株提供的 DNA 片段（chunk）数量，且全部菌株两两互为供体和受体，由此获得片段数量（chunkcounts）矩阵，随后利用马尔可夫链蒙特卡罗（Markov chain Monte Carlo, MCMC）算法对片段数量矩

阵进行聚类，得出群体结构。本文使用该软件的非连锁模型（unlinked model）和单倍体模式，将 MCMC 链长度和 burn-in 均设置为 50000 次迭代。最后我们综合 fineSTRUCTURE 输出的片段数量文件、MCMC 文件和 Newick 树文件，使用 R 语言绘图，根据聚类结果对菌株进行分组。

2.3.2 SNP 距离计算

为了验证 fineSTRUCTURE 得到的共同祖先矩阵的可靠性，我们使用 snp-dists v0.8.2（<https://github.com/tseemann/snp-dists>）计算菌株之间的成对 SNP 距离，该软件将两个菌株之间的每个 SNP 位点差异记为 1，并在全基因组范围内进行累加，由此获得 SNP 距离矩阵。使用 R 包 ComplexHeatmap^[25]将 SNP 距离矩阵绘制成热图并聚类分组。

2.4 进化树构建

我们分别用最大似然法（Maximum Likelihood, ML）和邻接法（Neighbor-Joining, NJ）构建进化树。首先将 2.2 获得的 SNP 矩阵文件转换成 FASTA 格式，使用软件 FastTree 2.1.10^[26] 构建最大似然树，使用 TreeBeST v1.9.2（<https://treesoft.sourceforge.net/treebest.shtml>）构建 NJ 进化树，将软件生成的两种 Newick 树文件输入在线工具 iTOL（<https://itol.embl.de/>）绘制无根进化树。

2.5 进化历史分析

2.5.1 连锁不平衡衰减计算

首先使用 PLINK v1.9^[27]将包含 SNP 信息的 VCF 文件转换成 Haploview^[28]软件的输入格式，用 Haploview 计算不同距离位点的连锁不平衡 r^2 值，位点最大距离设置为 30 kb。

2.5.2 主成分分析（Principal Component Analysis, PCA）和 SNP 载荷（loading）

我们使用 PLINK v1.9 对菌株做基于 SNP 的主成分分析，将原始数据分解为 10 个特征值。

为了探究不同 SNP 位点在进化中的贡献，我们利用 PCA 中贡献最大的 PC1 计算 SNP 载荷，首先将 PC1 的特征向量乘对应特征值的 1/2 次方，获得 PCA 载荷矩阵（<https://scentellegher.github.io/machine-learning/2020/01/27/pca-loadings-skill.html>），随后将每个 SNP 位点第二常见的等位基因赋值为 1，其他等位基

因赋值为 0，生成 bimbam 矩阵，将 PCA 载荷矩阵与 bimbam 矩阵相乘，即得到每个 SNP 位点的载荷。

我们还计算了不同距离范围内 SNP 载荷的皮尔森相关系数，以相关系数对这两个 SNP 距离的范围作图，获得 SNP 载荷的衰减图。

2.5.3 染色体涂绘

我们利用 ChromoPainter v0.0.4^[24] 将每个菌株基因组绘制成其他基因组序列的组合，即以来自不同群组的菌株基因组作为供体，本文中所有基因组作为受体进行计算，E-M 迭代值设置为 5，使用非连锁模型（unlinked model）得到供体菌向受体菌提供的片段数，使用 R 包 pophelper v2.3.1（<https://www.royfrancis.com/pophelper/articles/index.html>）绘制群体结构堆叠图。

2.5.4 基于 TreeMix 的基因流动分析

我们用 PLINK v1.19 生成 TreeMix 的输入文件，随后使用 TreeMix v1.13^[29] 推断不同群体之间的基因流动情况，将评估协方差矩阵的窗口长度设置为 500 个 SNP，预设可能的基因流动次数（m）为 0-5 次，每个 m 值重复计算 5 次。使用 R 包 optM^[30]用于寻找最佳 m 值。

三、研究结果

3.1 肠道沙门氏菌 *enterica* 亚种的群体结构

3.1.1 群体中的菌株来源

我们共下载了 311 个 *enterica* 亚种菌株的全基因组序列，这些菌株来自北美洲、亚洲、欧洲、非洲、南美洲的超过 30 个国家（图 1A，附表）。菌株分离于 1905 年至 2017 年间，主要来源为不同年龄段的人类、环境、食品及动物（图 1B 和 1D，附表）。来自 EnteroBase 的菌株由 Alikhan 等^[3]从每个 reBG 中选取的一个代表性菌株组成，即已经初步过滤了冗余菌株，我们将所有菌株与参考基因组进行比对，随后基于 SNP 距离对来自 EnteroBase 和 ENA 的菌株进行冗余菌株过滤，发现除了两数据库重合的 8 个菌株外，其余菌株间 SNP 距离均大于 2 kb，由此获得用于本文后续分析的包括 303 个 *enterica* 基因组的初级数据集（图 1C）。

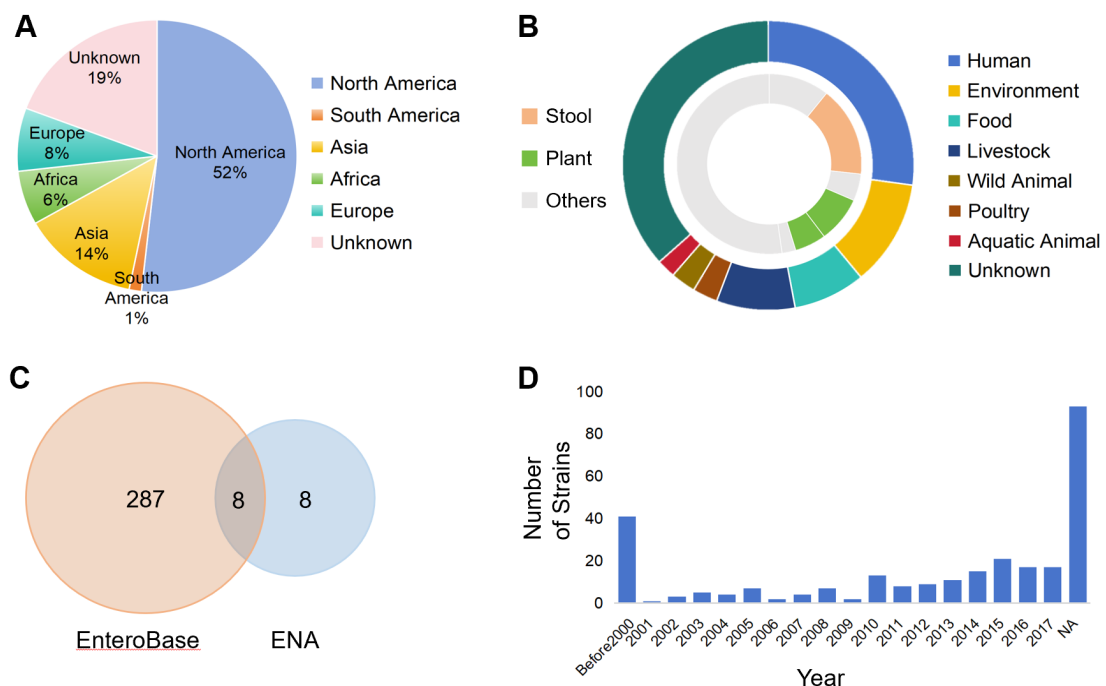


图 1. 本研究中菌株的来源信息基因与分离时间

A. 本研究中所有 *enterica* 菌株的来源地区；B. 菌株的宿主信息，内侧环中非灰色部分表示外侧环对应位置分组的子集，即环境中和食物中的植物组分均存在肠道沙门氏菌感染，部分人源 *enterica* 菌株分离自人类粪便；C. 来源于不同在线数据库的菌株数量；D. 菌株的分离时间

3.1.2 群体结构的初步划分

为了初步了解 *enterica* 的群体结构，我们使用可以分析菌株间遗传相似度、并能对群体自动聚类的 fineSTRUCTURE 软件计算菌株间的共同祖先矩阵。以含有 303 个菌株的初级数据集作为输入菌株，大部分菌株彼此提供的片段数（此处将其定义为共同祖先值）在 2000 以下，仅个别菌株间的共同祖先值超过 10000，这些共同祖先值高的菌株压缩了颜色比例尺的分辨率，使得结果可视化后整体呈现较浅的颜色，此时群体结构并不清晰，仅能观察到沿对角线存在三个颜色相对均一的群组（图 2A）。随后我们将共同祖先值大于 5000 的菌株视作克隆菌株（图 2 中蓝色箭头所指），每个克隆仅保留一个代表性菌株，重新进行 fineSTRUCTURE 分析，重复此过程 5 次，得到逐渐清晰的群体结构（图 2B-F），此时群体剩余 280 个菌株，我们将其定义为最终的非冗余数据集。

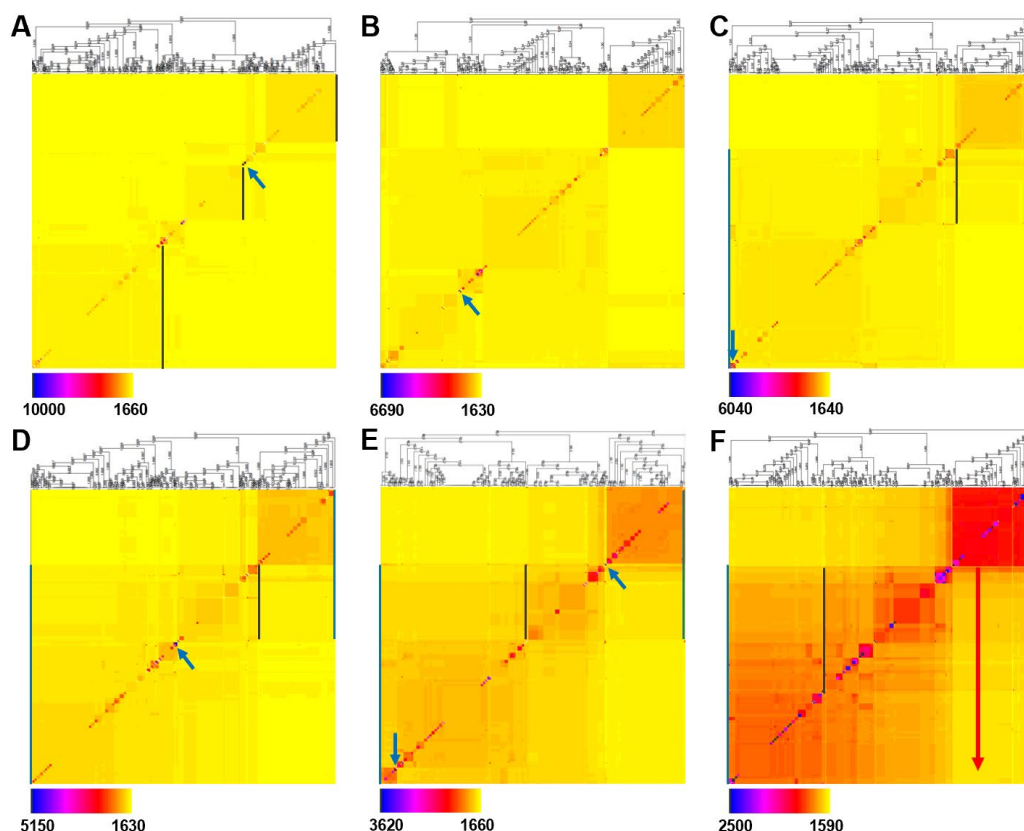


图 2. *enterica* 菌株的 fineSTRUCTURE 种群划分

A-F 为从 303 个菌株到 280 个菌株的 fineSTRUCTURE 迭代计算结果；每列代表一个供体菌株，每行代表一个受体菌株；热图上方为基于马尔科夫蒙特卡洛算法的聚类结果；热图颜色深浅表示供体菌株提供给受体菌株的 DNA 片段，数值越大代表两菌株共同祖先值越大，说明二者越相似；蓝色箭头示意每次迭代时需要选择性删除的克隆菌株；A 中黑线标记初始数据集中的三个群组；C-F 中两侧蓝线

标记相互重叠的两个大群体，中间黑线标记中间群体；F 中红色箭头表示共同祖先值梯度递减的方向

在上述 *fineSTRUCTURE* 迭代计算中，群体并未如我们预想的被从中一分为二^[3, 14]，而是出现了处于中间的群体。第二次迭代后的群体可以视作部分重叠的两个大群体（图 2C 两侧蓝色线标记），重叠部分记作中间群体（图 2C 中间黑线标记），随后的两次迭代中，大群体和中间群体的大小基本保持不变（图 2D 和 2E）。从第一次迭代起，共同祖先矩阵右上角始终有一个颜色相对均一的群组，该群组内部共同祖先值始终较高，我们将其记作群组 1。第五次迭代后，右侧大群体变大，中间群体随之变大，且可以观察到从群组 1 的边缘开始，中间群体菌株与群组 1 的共同祖先值梯度递减（图 2F 红色箭头），我们以此梯度为依据将对群体进行划分为七个群组（图 3A），由于每个群组中的菌株没有明显的地理分布特征，我们将其依次命名为群组 1-7。其中群组 1 共包含 74 个菌株，是克隆程度最高的群组；群组 2 仅包含 6 个菌株，但在群组 1 和其他群组之间起到重要的过渡作用；群组 3-7 依次包含 9、15、42、45、89 个菌株，即群体 2-7 群体大小递增。

3.1.3 群体结构的验证

为了验证上述 *fineSTRUCTURE* 群体划分的可靠性，我们又采用了成对 SNP 距离判断菌株之间的相似程度，SNP 距离越大说明两菌株间的差异越大，按照 *fineSTRUCTURE* 的最终聚类结果对 SNP 距离矩阵排序，发现 SNP 距离矩阵与共同祖先矩阵给出了类似的群体结构（图 3B）。群组 1 菌株内部相似度高，其余群组与群组 1 的相似度递减。值得注意的是，群组 6 对应的菌株在 SNP 距离热图上颜色不均一，可能说明群组 6 内部存在某种分化，于是我们对 SNP 距离热图进行聚类，发现群组 6 的一个子群体被单独隔离出来，与其余群组 6、群组 7 菌株差异较大（图 3C）。这一包含 13 个菌株的子群组在 *fineSTRUCTURE* 聚类中，也形成了一个靠近群组 5 的小簇（图 3A 和 3C），于是我们将这一子群组定义为群组 6a，即最终包含 280 个菌株的非冗余数据集被划分八个群组，分别为群组 1-7 及群组 6a。

表征方式中的特征较为一致，而中间群组在不同计算方法中表现出不同的特点，这些中间群组可能为我们进一步理解 *enterica* 群体结构的形成提供重要的信息。

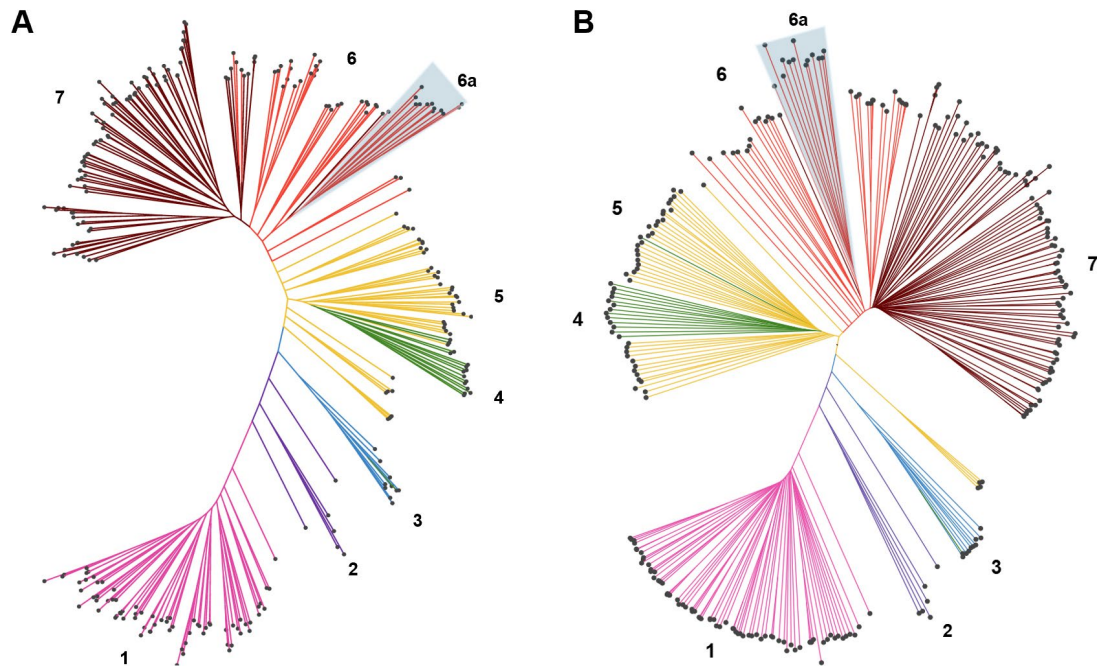


图 4. 非冗余数据集中全部菌株的无根进化树

A 为最大似然法进化树；B 为邻接法进化树；群组 1-7 的分支依次用粉色、紫色、蓝色、绿色、黄色、红色、棕色标记；蓝色三角区域内为群组 6a

3.2.2 不同群组内部的连锁不平衡分析

由于连锁不平衡可以衡量群体中的重组（recombination）-漂变（drift）平衡，且遗传混合（admixture）也会产生新的连锁不平衡^[8]，我们尝试通过计算群组内部的连锁不平衡衰减（LD decay）来初步探究群体的进化历史。

由于连锁不平衡受到种有效群大小（ N_e ）的影响^[31]（图 5A），我们将较小的群组 2 和群组 3 合并为一组，将进化树上相互融合程度较高的群组 4 和群组 5 合并为一组，从各组及总体数据集中各随机抽取 15 个菌株用于连锁不平衡衰减计算。

我们观察到，群组 2、群组 3、群组 1 和群组 7 的连锁不平衡值总体较高，群组 4、群组 5、群组 6 的连锁不平衡值在 0-250 bp 内快速衰减（图 2B-D）。由此我们认为群组 1、2、3、7 可能存在较高程度的遗传漂变或遗传混合事件，由于群组 1、2、3 的克隆程度较高，我们倾向于认为这三个群组中较高的连锁不平衡值是由遗传漂变造成的。

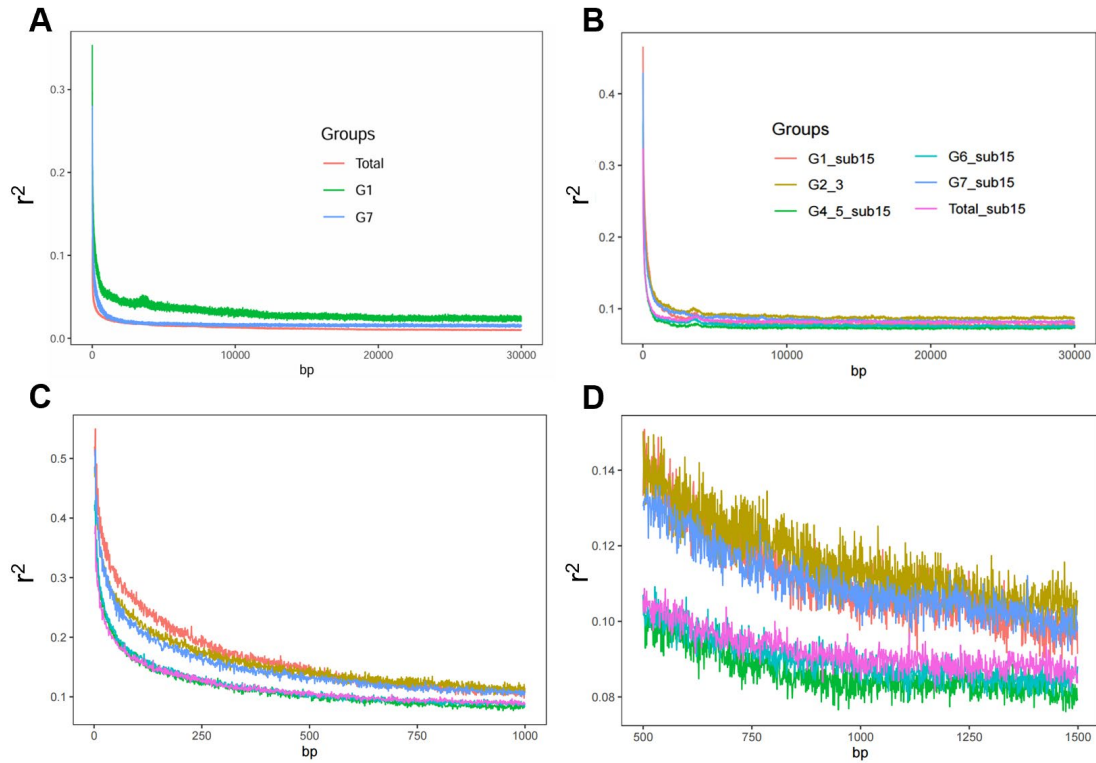


图 5. 群组内部连锁不平衡衰减分析

A-D 为连锁不平衡衰减曲线，横坐标为 SNP 位点之间的距离，纵坐标为表征连锁不平衡的 r^2 值；A 为群组 1、群组 7 和总数据集在 0-30 kb 范围内的连锁不平衡衰减曲线；B 为从各个群组中抽取 15 个菌株计算获得的 0-30 kb 内的连锁不平衡衰减曲线；C 为 B 中曲线的 0-1 kb 窗口；D 为 B 中曲线的 500-1500 bp 窗口

3.2.3 PCA 和全基因组 SNP 载荷分析

接着我们尝试通过 PCA 进一步研究种群结构的形成原因，此处我们选择了群体结构与 *enterica* 存在相似之处，且与 *enterica* 具有相似的无根进化树“骨架”的肺炎克雷伯菌 (*Klebsiella pneumoniae*) 作为 *enterica* 的对照 (图 6F)。在常规 PCA 中我们观察到群组 1、2、3 各自独立，群组 4-7 则彼此连续且靠近 (图 6A)。由于 PC1 对变量的解释程度最高 (图 6B)，我们利用根据先计算了 PC1 的载荷，随后通过矩阵乘法获得了全基因组范围内所有 SNP 的载荷 (图 6C)，与肺炎克雷伯菌相比，*enterica* 每个 SNP 的载荷值更大，且没有发现载荷值远高于其他位点的 SNP (图 6E)，这可能说明 *enterica* 的进化不是由某个 SNP 位点或基因受到选择导致的。SNP 载荷的衰减可能是一个与连锁不平衡衰减类似的表征值，我们发现 *enterica* 的连锁不平衡值和 SNP 载荷值相关系数都总体与肺炎克雷伯菌接近，但肺炎克雷伯菌的 LD 值衰减更快，且与 SNP 载荷相关系

数一样最终衰减至接近 0，而 *enterica* 的 LD 值在 SNP 距离超过 1000 bp 后始终高于 SNP 载荷相关系数，这意味着 *enterica* 比肺炎克雷伯菌具有更高的克隆性，*enterica* 的重组事件并没有完全克服其连锁不平衡效应，而相近的 SNP 载荷相关系数可能说明这一系数的衰减受到重组以外因素的影响。

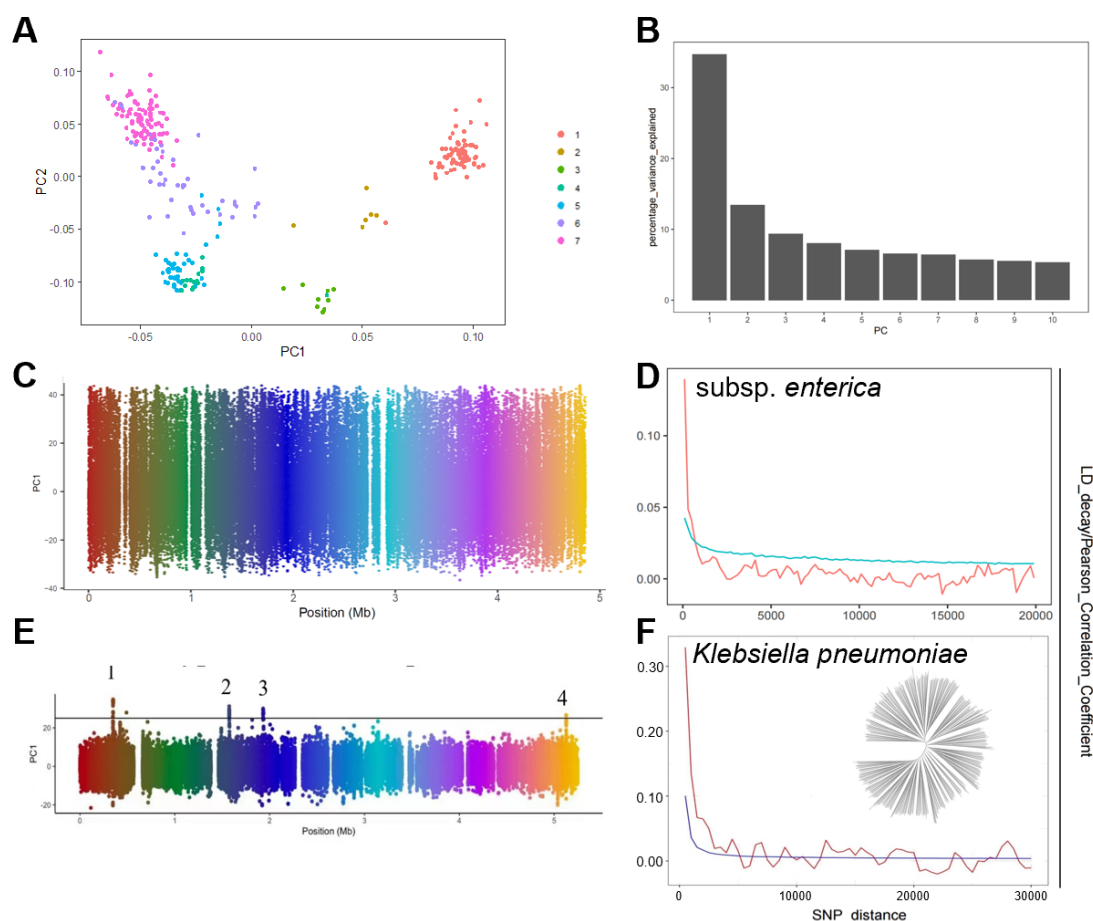


图 6. *enterica* 的 PCA、SNP 载荷与肺炎克雷伯菌的 SNP 载荷

A 为 *enterica* 完整数据集全基因组 SNP 的 PC1 和 PC2；B 为 PCA 分析中各个 PC 的特征值；C、E 分别为 *enterica* 肺炎克雷伯菌全基因组的 SNP 载荷值；D、F 为 *enterica* 与肺炎克雷伯菌载荷值相关系数衰减和 LD 衰减，其中红色线为载荷值相关系数衰减，蓝色线为 LD 衰减，F 中肺炎克雷伯菌的无根最大似然进化树

3.2.4 基于染色体涂绘的祖源概况分析

为了分析 *enterica* 群体进化的历史，我们对数据集中全部菌株进行了染色体涂绘，即计算每个菌株的基因组是由多少其他基因组片段组成的，以此推测每个菌株基因组的祖源占比。我们首先以 8 个群组（群组 1-7 和群组 6a）中全部菌株作为供体计算，我们发现不同群组的受体菌祖源情况不同，但是该结果中

的染色体组成明显受供体所在群组大小影响（图 7A）。于是，我们从每个群组中选取 6 个菌株作为供体，我们发现群组 1、2、3 在自身群组中的贡献显著大于对其他群组的贡献，而群组 4-7 和群组 6a 作为供体在不同群组的受体中贡献却比较类似（图 7B），所以我认为群组 1、群组 2 和群组 3 群体内发生了较大程度的遗传漂变，这符合我们在 3.2.2 中的推测。

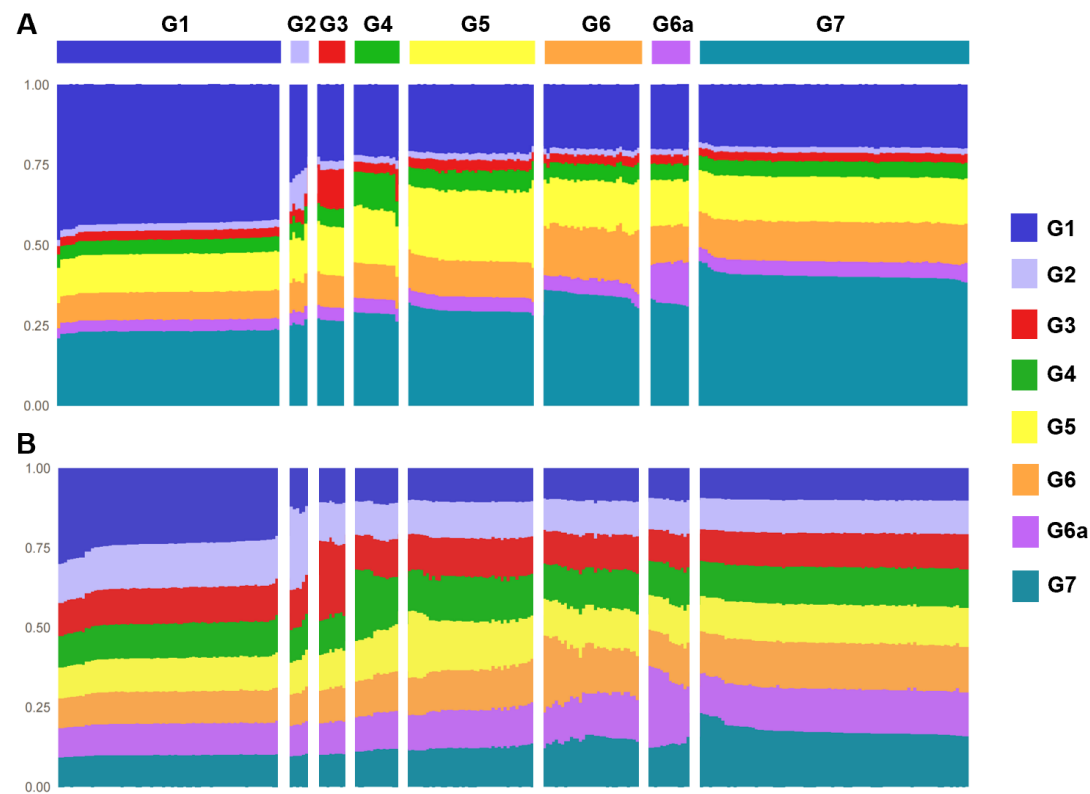


图 7. 非冗余数据集中所有菌株的染色体涂绘结果

每个柱为一个受体菌株，不同颜色代表供体菌所在的群组：A 为以全部菌株为供体的涂绘结果；B 为以每个群组中 6 个菌株为供体的涂绘结果

3.2.5 基因流动分析

最后，我们使用 TreeMix 推断不同群组之间的基因流动事件，无论发生几次基因流动事件，TreeMix 计算结果始终得出群组 1、2、3 的遗传漂变程度最大，这与我们此前的分析结论一致（图 8A-E）。依据 optM 对 TreeMix 结果的评估，四次基因流动事件的计算结果可能是最可靠的，综合不同基因流动事件次数的分析结果，我们认为最可能存在的是群组 6 (a)向群体 1 和群体 2 分支点处的基因流动（图 8E-G）。

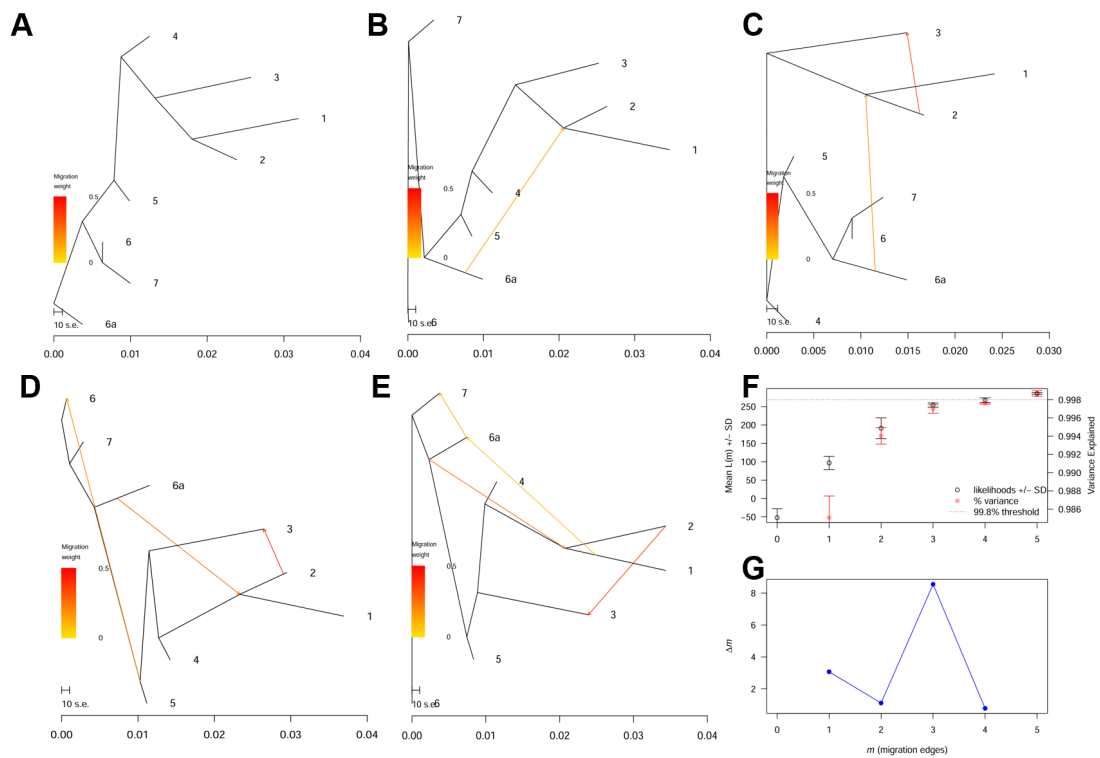


图 8. 基因流动分析结果

A-E 分别为基因流动事件设为 0-4 次的 TreeMix 计算结果；横坐标为遗传漂变参数，箭头颜色表示基因流动事件的强度；F 为 optM 计算的综合似然值 $L(m)$ 和标准差 (SD) (左轴，黑色圆圈) 以及被解释的方差比例 (右轴，红色圆圈)；G 为不同 m 值对应的 m 值变化率 (Δm)

四、讨论

4.1 研究总结

本研究采用不同算法对 280 个 *enterica* 菌株的群体结构进行了划分，与此前研究认为的不同^[14]，*enterica* 并非从中间直接分为两个独立的亚群，而是在两个克隆性较高的群组之间出现了多个中间群组，这可能意味着其群体进化是渐变发生的。我们根据获得的群体结构将 *enterica* 分为 8 个群组，通过计算群组内部的连锁不平衡和遗传漂变寻找不同群组的特征。此外，我们对群体结构存在相似特征的 *enterica* 和肺炎克雷伯菌进行了 SNP 载荷的比较，为理解类似群体结构的形成机制提供了思路。

4.2 局限性与展望

在进行 fineSTRUCTURE 分析和染色体涂绘时，我们均采用了软件的非连锁模型，这是由于连锁模型需要预估基因组不同位置的重组率，而本研究并未进行重组热点的计算，使用非连锁模型意味着我们损失了连锁不平衡信息，后续我们将计算较为可靠的 *enterica* 重组率，使用连锁模型验证本文中对群体的划分和染色体涂绘计算。同时，本研究进行 PCA 计算时也没有对存在连锁不平衡的 SNP 位点进行筛选，这可能使计算结果出现系统性偏差，后续我们会使用筛选后的 SNP 位点重新进行 PCA 和 SNP 载荷计算。

本文中计算 SNP 载荷的方法是课题组的新尝试，目前只利用非冗余数据集的全体菌株计算了 SNP 载荷，不能充分体现出不同群组的 SNP 载荷特征的区别，目前的计算方法将 PC1 载荷与 SNP 的 bimbam 矩阵相乘，损失了单独菌株的信息，我们正在尝试通过增加计算步骤表征来自不同群组的菌株对 SNP 载荷的贡献。我们目前发现了 *enterica* 与肺炎克雷伯菌的 SNP 载荷存在明显的差异，然而受限于我们对两种微生物群体结构和进化历史的理解，目前造成这种差异的原因尚未可知，这将是我们的后续工作的重点。

本文对 *enterica* 进化历史的研究并不充分，目前通过不同方法得到验证的信息为群组 1、2、3 内部存在较高程度的遗传漂变，然而对基因流动的分析仅为软件的预测结果，仅群组 6a 向群组 1 和群组 2 的基因流动事件在多次预测中稳定出现，且我们目前不能证明预测结果是否符合真实情况。后续我们会通过

建立模拟群体，复现目前获得的群体结构，并由此推测可能真实存在的基因流动事件。

本文中的全部分析都在全基因组尺度上进行，我们无法得知是否存在针对某个基因或某个 SNP 位点的自然选择推动了进化的发生，后续我们将进行全基因组关联分析（genome wide association studies, GWAS）和群体分化系数（ F_{st} ）计算，以判断是否有 SNP 在群体分化中起到较大贡献，进一步研究这些 SNP 所在的基因功能，探讨针对这些基因的群体分化驱动力。

本研究及后续工作将揭示 *enterica* 的具体进化历史和进化驱动力，说明本文所述群体结的成因，解释强克隆性细菌的无根进化树“骨架”出现的机制。

参考文献

1. Lamas A, Miranda JM, Regal P, et al. A comprehensive review of non-*enterica* subspecies of *Salmonella enterica*[J]. Microbiol Res, 2018, 206:60-73.
2. Franklin K, Lingohr EJ, Yoshida C, et al. Rapid Genosertotyping Tool for Classification of *Salmonella* Serovars[J]. J Clin Microbiol, 2011, 49(8):2954-2965.
3. Alikhan NF, Zhou ZM, Sergeant MJ, et al. A genomic overview of the population structure of *Salmonella*[J]. PLoS Genet, 2018, 14(4).
4. Achtman M, Wain J, Weill FX, et al. Multilocus Sequence Typing as a Replacement for Serotyping in *Salmonella enterica*[J]. PLoS Pathog, 2012, 8(6).
5. Tindall BJ, Grimont PAD, Garrity GM, et al. Nomenclature and taxonomy of the genus *Salmonella*[J]. Int J Syst Evol Microbiol, 2005, 55(1):521-524.
6. Giammanco GM, Pignato S, Mammina C, et al. Persistent endemicity of *Salmonella bongori* 48:z(35):--in Southern Italy: molecular characterization of human, animal, and environmental isolates[J]. J Clin Microbiol, 2002, 40(9):3502-3505.
7. Sánchez-Vargas FM, Abu-El-Haija MA, Gómez-Duarte OG. *Salmonella* infections: An update on epidemiology, management, and prevention[J]. Travel Med Infect Dis, 2011, 9(6):263-277.
8. Jobling M, Hollox E, Kivisild T, et al. Human Evolutionary Genetics, 2nd ed.[M]. Garland Science, 2013.
9. Robinson DA, Falush D, Feil EJ. Bacterial Population Genetics in Infectious Disease[M]. John Wiley & Sons, 2010.
10. Spratt BG. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet[J]. Curr Opin Microbiol, 1999, 2(3):312-316.
11. Maiden MCJ, van Rensburg MJJ, Bray JE, et al. MLST revisited: the gene-by-gene approach to bacterial genomics[J]. Nat Rev Microbiol, 2013, 11(10):728-736.
12. Jolley KA, Bliss CM, Bennett JS, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain[J]. Microbiology (Reading), 2012, 158:1005-1015.
13. Nadon C, Van Walle I, Gerner-Smidt P, et al. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance[J]. Euro Surveill, 2017, 22(23):30544.
14. Criscuolo A, Issenhuth-Jeanjean S, Didelot X, et al. The speciation and hybridization history of the genus *Salmonella*[J]. Microb Genom, 2019, 5(8).
15. Brown EW, Mammel MK, LeClerc JE, et al. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens[J]. Proc Natl Acad Sci U S A, 2003, 100(26):15676-15681.
16. Desai PT, Porwollik S, Long F, et al. Evolutionary Genomics of *Salmonella*

- enterica* Subspecies[J]. mBio, 2013, 4(2).
17. Trujillo S, Keys CE, Brown EW. Evaluation of the taxonomic utility of six-enzyme pulsed-field gel electrophoresis in reconstructing *Salmonella* subspecies phylogeny[J]. Infect Genet Evol, 2011, 11(1):92-102.
 18. Falush D, Torpdahl M, Didelot X, et al. Mismatch induced speciation in *Salmonella*: model and data[J]. Philos Trans R Soc Lond B Biol Sci, 2006, 361(1475):2045-2053.
 19. Sheppard SK, Guttman DS, Fitzgerald R. Population genomics of bacterial host adaptation[J]. Nat Rev Genet, 2018, 19(9):549-565.
 20. Zhou ZM, Alikhan NF, Mohamed K, et al. The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity[J]. Genome Res, 2020, 30(1):138-152.
 21. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes[J]. Genome Biol, 2004, 5(2).
 22. Page AJ, Taylor B, Delaney AJ, et al. *SNP-sites*: rapid efficient extraction of SNPs from multi-FASTA alignments[J]. Microb Genom, 2016, 2(4).
 23. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools[J]. Bioinformatics, 2011, 27(15):2156-2158.
 24. Lawson DJ, Hellenthal G, Myers S, et al. Inference of Population Structure using Dense Haplotype Data[J]. PLoS Genet, 2012, 8(1).
 25. Gu ZG, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data[J]. Bioinformatics, 2016, 32(18):2847-2849.
 26. Price MN, Dehal PS, Arkin AP. FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments[J]. PLoS One, 2010, 5(3).
 27. Chang CC, Chow CC, Tellier L, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets[J]. Gigascience, 2015, 4.
 28. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps[J]. Bioinformatics, 2005, 21(2):263-265.
 29. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data[J]. PLoS Genet, 2012, 8(11).
 30. Fitak RR. *OptM*: estimating the optimal number of migration edges on population trees using *Treemix*[J]. Biol Methods Protoc, 2021, 6(1):bpab017.
 31. Tenesa A, Navarro P, Hayes BJ, et al. Recent human effective population size estimated from linkage disequilibrium[J]. Genome Res, 2007, 17(4):520-526.

附表 本研究使用的 *enterica* 菌株背景信息

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
1	EA3175AA	SRR2140715	FDA290104	Environment	Plant	Cumin Seeds	2004/06	Asia	Syria	Caracas
1	EA3620AA	SRR1015725	FL_FLDACS-57071	Food	Plant	Ground Ginger	1905/06	North America	USA-Florida	NA
1	EA4049AA	SRR1002807	NY_FSL S10-1092	Environment	Soil/Dust	Produce Field- Drag Swab	2010/10	North America	USA-Florida	Saphra
1	EA5009AA	SRR2007638	FDA905293-C2-1	Food	Plant	Chili Powder, 55Lb Sack	2015/04	Asia	Pakistan	Sundsvall
1	EA5468AA	SRR1980611	FDA904406 1-1	Food	Plant	Ground Chili Pepper	2015/04	Africa	Nigeria	Pomona
1	EA9557AA	SRR1805578	FDA481989	Environment	Plant	Serrano Peppers	2008/08	North America	Mexico	Tucson (predicted)
1	FA3977AA	SRR1481653	MOD1_Sal_866	Wild Animal	Reptile	Corn Snake (Small)-Swab	2010/03	North America	USA-Alabama	Kokomle (Predicted)
1	EA9892AA	SRR1772941	2013K-0048	NA	NA	NA	NA	North America	USA	NA
1	FA5258AA	SRR1212359	FSE0077	Food	Shellfish	Frz. Cooked Shrimp	2010/12	Asia	India	Teko
1	FA5926AA	SRR1105674	FSF0006	Wild Animal	Reptile	Gecko	2011/09	Asia	China	Eastbourne (Predicted)
1	FA6089AA	SRR1045868	AZ_TG68523	Environment	Water/River	Kitchen Drain	2011/02	North America	USA-Texas	Eastbourne
1	FA6145AA	SRR1035587	FSW0043	Food	Plant	Halva with Pistachio	2011/04	Asia	Turkey	Carnel
1	FA6307AA	ERR304836	ASERH120234	NA	NA	NA	NA	NA	NA	Curacao (Predicted)

注：此表中“NA”表示信息未知（not available），“ND”表示不确定（not determined），“Frz”表示冷冻的（Frozen），“CSF”表示脑脊液（Cerebrospinal Fluid）。

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
1	FA7416AA	SRR1534878	AZ-TG74720	Poultry	Poultry	Ground Turkey	1905/06	North America	USA-Minnesota	Brandenburg
1	FB9014AA	SRR5927373	2015AM-0954	Human	Human	Age: 40-49; Stool	2015/06	North America	USA	Miami (predicted)
1	IA2712AA	SRR2585561	FDA208317-51	Food	Plant	Ground Cumin	2002/12	Asia	Turkey	Onderstepoort (Predicted)
1	JA8455AA	SRR2911801	FDA428106	Environment	Plant	Cantaloupe	2007/06	North America	USA-South Carolina	Florida
1	JA9577AA	SRR2975856	2015K-0686	ND	ND/Others	Stool	2013/01	North America	USA	Oranienburg
1	KA2159AA	SRR3091390	POR1-0182	Livestock	Swine	<i>Sus scrofa</i> ; Feces	2015/04	North America	USA-Texas	Edinburg (Predicted)
1	KA3383AA	GCF_000341425	CFSAN001992	Human	Human	<i>Homo sapiens</i> ; Clinical	2012/10	North America	USA-Arizona	Javiana
1	KA3427AA	GCF_000940975	USDA-ARS- USMARC-1903	Environment	Animal- Related	Livestock- Associated Habitat	2006/10	North America	USA	Montevideo
1	KA3493AA	SRR957989	ATCC 35640	Environment	Water/River	Environmental- Water; Creek	NA	South America	Argentina- Misiones	Abacetuba
1	KA3494AA	SRR955256	CFSAN001080	Environment	Plant	Marjoram	NA	NA	NA	Bredeney
1	KA3512AA	GCF_000020745	CVM19633	NA	NA	NA	NA	NA	NA	Schwarzengrund
1	LA0730AA	SRR3497866	PNUSAS001990	Human	Human	Age: 50-59; Stool	2014/05	North America	USA	Bonariensis (Predicted)
1	LA3665AA	SRR3708002	2016K-0488	NA	NA	NA	NA	North America	USA	Pomona (Predicted)

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
1	LA4053AA	ERR1082845	BS_17322-sc-2322420	NA	NA	NA	NA	Africa	Gambia	NA
1	LA5313AA	SRR1753727		NA	NA	NA	NA	NA	NA	Telelkebir
1	LA6123AA	SRR1118693		Food	ND/Others	Food	1905/07	Africa	Ethiopia	Chandans
1	LA8795AA	ERR998630		NA	NA	NA	NA	NA	NA	NA
1	MA0478AA	SRR3930410	PNUSAS002509	Human	Human	Age: 10-19; Stool	2016/03	North America	USA	Gaminara
1	MA3151AA	ERR1556131	SL1344	NA	NA	NA	NA	NA	NA	Poona (Predicted)
1	MA6966AA	SRR4100117	PNUSAS003498	ND	ND/Others	Stool	2004/08	North America	USA	Give
1	NA0974AA	SRR4418295	PNUSAS004556	NA	NA	NA	2016/09	North America	USA	Give
1	OA3141AA	ERR1743772	L01166-15	NA	NA	NA	NA	NA	NA	Flutern (Predicted)
1	PA7338AA	NA	NCTC4840	NA	NA	NA	NA	NA	NA	Poona
1	PA7363AA	NA	NCTC6024	NA	NA	NA	NA	NA	NA	Sandiego
1	PA7371AA	NA	NCTC7832	NA	NA	NA	NA	NA	NA	Nottingham
1	PA7403AA	NA	NCTC5743	NA	NA	NA	NA	NA	NA	Oranienburg
1	PA7404AA	NA	NCTC6482	NA	NA	NA	NA	NA	NA	Carrau
1	PA7605AA	GCF_000188955	507440-20	Environment	Soil/Dust	Environmental; Environmental Swab	2010/02	North America	USA- Rhode Island	Montevideo
1	PA8587AA	SRR5176561	PNUSAS006319	Human	Human	<i>Homo sapiens</i> ; Age: 30-39; Stool	2015/12	North America	USA	Poona (Predicted)
1	QA5738AA	SRR5238974	OSF059775	ND	ND/Others	Finished Spinach	1905/07	North America	USA-California	NA

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
1	QA8796AA	SRR5335975	PNUSAS008823	Human	Human	Age: 70-79; Urine	2016/08	North America	USA	NA
1	RA0878AA	SRR5366175	PNUSAS010349	Human	Human	Age: 40-49; Stool	2017/01	North America	USA	Bredeney (predicted)
1	RA1483AA	SRR5379187	PNUSAS009813	NA	NA	NA	NA	North America	USA	Montevideo (predicted)
1	RA4414AA	SRR5398006	ADRD-1033	Wild Animal	Reptile	Veterinary Clinical/Research Wild Animal; Cloacal swab	2003/01	North America	USA-Oklahoma	Oranienburg
1	SA2775AA	SRR5514435	FSIS1700451	Livestock	Swine	Iguana Animal-Swine- Sow	1905/07	North America	USA-Texas	Gaminara
1	SA8630AA	SRR5574832	PNUSAS013297	Human	Human	<i>Homo sapiens</i>	NA	North America	USA	Give (Predicted)
1	SA8755AA	SRR5576554	PNUSAS013418	Human	Human	Age: 50-59; Stool	2017/01	North America	USA	Durban (Predicted)
1	SA9125AA	SRR5583018	08-0376	NA	NA	NA	NA	North America	USA	Jangwani (Predicted)
1	TA0958AA	SRR5605898	PNUSAS011936	Human	Human	CSF	2016/12	North America	USA	Kisarawe (Predicted)
1	TA2276AA	SRR5642187	PNUSAS014475	Human	Human	Age: 20-29; Stool	2016/11	North America	USA	Muenster (Predicted)
1	TA5546AA	SRR5740071	2015AM-0411	Human	Human	Age: 0-4; Stool	2015/03	North America	USA	Cotham (Predicted)

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
1	TA6144AA	SRR5756389	PNUSAS014406	Human	Human	Age: 80-89; Stool	2017/05	North America	USA	NA
1	TA6270AA	SRR5759633	2015AM-1601	NA	NA	NA	NA	North America	USA	Oranienburg (Predicted)
1	TA7269AA	SRR5760379	2015AM-1537	Human	Human	Age: 30-39; Stool	2015/06	North America	USA	Arechavaleta (Predicted)
1	TA9344AA	SRR5823440	PNUSAS016938	NA	NA	NA	NA	North America	USA	Charity (Predicted)
1	UA0742AA	SRR5839906	2015AM-1132	Human	Human	Age: 5-9; Stool	2015/07	North America	USA	Rubislaw (Predicted)
1	UA1891AA	SRR5861558	2015AM-0931	Human	Human	Age: 40-49; Fluid	2015/06	North America	USA	NA
1	UA3008AA	SRR5876392	PNUSAS017807	Human	Human	Age: 30-39; Stool	2017/06	North America	USA	Javiana (Predicted)
1	UA5399AA	SRR5926934	2015AM-1607	NA	NA	NA	NA	North America	USA	Rubislaw (Predicted)
1	UA6078AA	SRR5935626	PNUSAS018678	Human	Human	Age: 20-29; Stool	2017/06	North America	USA	Ball (Predicted)
1	UA6938AA	SRR5948413	PNUSAS020783	NA	NA	NA	NA	North America	USA	NA
1	UA7888AA	SRR5966304	2015AM-1363	NA	NA	NA	NA	North America	USA	Montevideo (Predicted)
1	VA9064AA	SRR6114366	PNUSAS023256	Human	Human	Age: 70-79; Stool	2017/06	North America	USA	Javiana (Predicted)
1	WA0646AA	SRR6128850	2015AM-2074	NA	NA	NA	NA	North America	USA	Panama (Predicted)

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
1	WA1865AA	SRR6155354	PNUSAS023366	NA	NA	NA	NA	North America	USA	Give (Predicted)
1	ABEH02	NA	NA	NA	NA	NA	NA	NA	NA	NA
1	AFCL01	NA	NA	NA	NA	NA	NA	NA	NA	NA
1	AFCM01	NA	NA	NA	NA	NA	NA	NA	NA	NA
1	AFCP01	NA	NA	NA	NA	NA	NA	NA	NA	NA
1	AFCQ01	NA	NA	NA	NA	NA	NA	NA	NA	NA
1	AFYI01	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	FA1944AA	SRR1612322	FNW19a83	Food	Plant	Turmeric Powder	2009/03	Asia	Vietnam	Bangkok
2	FA3443AA	SRR1544155	FNW19I14	Environment	Plant	Pepper	2012/02	Asia	Sri Lanka	41:Z4,Z23:-
2	FA4286AA	SRR1206168	CFSAN008300	Environment	Plant	Cilantro	2014/01	North America	USA-Georgia	Buzu
2	LA9136AA	ERR1046254	CG15F-sc- 2280773	NA	NA	NA	NA	NA	NA	Orientalis (Predicted)
2	OA3142AA	ERR1743773	L01171-15	NA	NA	NA	NA	NA	NA	Nima (Predicted)
2	QA9304AA	SRR5341695	PNUSAS008672	ND	ND/Others	Blood	2016/07	North America	USA	Bovismorbificans
3	BA5302AA	SRR1966222	72311	Human	Human	<i>Homo sapiens</i> ; Human	2014/12	Europe	UK-England	Mississippi
3	JA4089AA	SRR1957944	58480	Human	Human	<i>Homo sapiens</i> ; Human	2014/09	Europe	UK-England	Kingston
3	LA1103AA	SRR3535129	2012K-1664	NA	NA	NA	NA	North America	USA	Mississippi (Predicted)
3	LA3518AA	SRR3667854	PNUSAS002449	Human	Human	Age: 50-59; Stool	2016/05	North America	USA	Teilekebir (Predicted)

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
3	LA6782AA	SRR1753846	99-0250	NA	NA	NA	NA	NA	NA	Ibadan
3	LA8015AA	ERR731412	C1533	Human	Human	Blood Cultures from Children	1905/07	Africa	Nigeria-FCT	Predicted Agama
3	OA3075AA	ERR1743710	S00742-16	NA	NA	NA	NA	NA	NA	NA
3	OA5313AA	ERR1743662	L01836-15	Companion Animal	Canine	Diarrhoea	1905/07	Europe	UK-England	Agama
3	TA2101AA	SRR5640207	PNUSAS014542	Human	Human	Age: 5-9; Stool	2017/04	North America	USA	NA
4	EA4332AA	SRR2050968	FDA896142-88	Environment	Soil/Dust	Environmental Swab	2015/05	North America	USA-California	Havana (Predicted)
Bovine:										
4	FA2999AA	SRR1554531	OSF023130	Animal Feed	Bovine	Foodon_03414374 Food Product for Animal: Foodon_03309997	2007/01	North America	USA-Washington	Senftenberg (Predicted)
4	KA2826AA	SRR3133012	FDA920656-1-1	Aquatic Animal	Fish	Animal Feed; Bovine Feed Mud Carp Fish Fillet	2015/08	Asia	Thailand	Albany (Predicted)
4	KA3416AA	GCF_000993725	FORC_015	Human	Human	<i>Homo sapiens</i> ; Blood	2013/10	Asia	South Korea	Mbandaka
4	LA8693AA	ERR998515	dj4695_12-sc-2235503	NA	NA	NA	NA	NA	NA	NA
4	LA9822AA	SRR3134923	CFSAN031603	Livestock	Swine	Swine; Feces	2005/09	Africa	Ethiopia-Addis Ababa	Havana (Predicted)

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
4	PA2810AA	SRR51115800	FDA657077-C002-001	Environment	Plant	Sesame Seed	2010/11	North America	Mexico	Llandoff (Predicted)
4	PA3178AA	SRR5077944	FSIS1608821	Livestock	Bovine	Comminuted Beef	1905/07	North America	USA-Minnesota	Putten (Predicted)
4	PA7375AA	NA	NCTC8261	NA	NA	NA	NA	NA	NA	Alachua
4	PA7392AA	NA	NCTC5793	NA	NA	NA	NA	NA	NA	Worthington
4	RA6230AA	ERR1910715	MS170045	Human	Human	Human	1905/07	Europe	Ireland	Agama
4	TA4797AA	SRR5705135	FSIS1701475	Livestock	Bovine	Animal-Cattle-Steer	1905/07	North America	USA-Nebraska	Altona
4	UA2609AA	SRR5868300	2015AM-1188	Human	Human	Age: 0-4; Stool	2015/04	North America	USA	NA
4	VA5670AA	SRR6039028	PNUSAS020465	NA	NA	NA	NA	North America	USA	Kentucky (Predicted)
4	VA5780AA	SRR6039954	FSIS11704267	Livestock	Bovine	Comminuted Beef	1905/07	North America	USA-Georgia	NA
5	BA3057AA	SRR1272778	MDH-2014-00079	Livestock	Swine	Porcine Tissue	2003/10	North America	USA-Colorado	Livingstone
5	BA5801AA	SRR1574307	OSF023633	NA	NA	NA	2007/04	North America	USA-Maryland	NA
5	CA4312AA	ERR160977	07-0005	NA	NA	NA	NA	NA	NA	Krefeld (predicted)
5	EA3695AA	SRR2086909	ADRDL-NGUA-29	Poultry	Poultry	<i>Gallus gallus</i> ;	2014/09	Africa	Nigeria	Zega (Predicted)
5	EA5511AA	ERR160971	07-1373	NA	NA	Poultry Tissue	NA	NA	NA	Yoruba
5	EA5620AA	SRR1973971	CFSAN006219	Livestock	ND/Others	Sausages	2007/04	South America	Colombia-Cordoba	Fresno

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血型
5	EA8198AA	SRR1963508	79114	Human	Human	<i>Homo sapiens</i> ; Human	2015/01	Europe	UK-England	Krefeld
5	EA9028AA	SRR1920229	FDA470741 2-1	Food	Plant	Red Chili Powder	2008/05	Asia	India	Molade
5	FA1793AA	SRR1616722	2011K-1411	NA	NA	NA	NA	North America	USA	NA
5	FA2457AA	SRR1576008	OSF023681	Environment	Soil/Dust	Vacuum Dust, Dirt/Debris	2012/10	North America	USA-Oregon	Tennessee (Predicted)
5	FA4937AA	SRR1258580	VA_WGS-00184	Environment	Plant	Sesame Seeds	2008/07	North America	USA-Virginia	Senftenberg
5	FA5548AA	SRR1177647	FNE0057	Food	Plant	Chili Powder	2011/04	Asia	Pakistan	Amsterdam
5	FA5882AA	SRR1107425	FL_EPI-006	Human	Human	<i>Homo sapiens</i> ; Stool from Hospitalized Patient	1905/07	North America	Haiti	Kentucky (Predicted)
5	IA3795AA	SRR2585436	FDA809350 C2-1	Environment	Plant	Tahina (Crushed Sesame Seed)	2013/05	North America	Canada	50:y:-
5	IA8040AA	ERR212672	Ke487	NA	NA	NA	NA	NA	NA	4,12:d:- (Predicted)
5	JA8386AA	SRR1177650	FNE0060	Aquatic Animal	Fish	Fish	2011/04	Asia	Bangladesh	Oranienburg
5	KA3387AA	GCF_001409135	10259/2511STDY 5462388	Human	Human	Stool	NA	NA	NA	Weltevreden
5	KA3496AA	GCF_000486405	TXSC_TXSC08- 19	Food	Fish	Fishmeal	2004/01	North America	USA-Texas	Tennessee
5	KA3511AA	GCF_000020885	SL483	Human	Human	NA	1905/06	North America	USA	Agona

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
5	KA3515AA	GCF_000430125	CFSAN002050	Environment	Plant	Fresh Alfalfa Sprouts	2012/09	North America	USA-Arizona	Cubana
5	KA3668AA	ERR1251823	S01603_15	NA	NA	NA	NA	NA	NA	Kedougou (Predicted)
5	KA9645AA	SRR3402104	FSIS1606290	Livestock	Swine	Animal-Swine-Market Swine	1905/07	North America	USA-South Carolina	Derby
5	KA9658AA	SRR3402176	FSIS1606193	Livestock	Swine	Animal-Swine-Sow	1905/07	North America	USA-Tennessee	Derby
5	LA0025AA	SRR3457645	HIY0271	Environment	Water/River	Environmental-Water; Sewage Sand Island Wastewater	2010/11	North America	USA-Hawaii	Takoradi
5	LA2258AA	SRR3572401	PNUSAS002182	Human	Human	Age: 30-39; Stool	2010/03	North America	USA	Ago
5	MA0339AA	SRR3223887	OSF046202	Environment	Plant	Macadamia Nuts	2015/08	North America	USA-Texas	NA
5	NA8253AA	SRR5017151	PNUSAS005004	NA	NA	NA	NA	North America	USA	Ealing (Predicted)
5	OA0567AA	SRR5045020	FSIS1608444	Livestock	Bovine	Comminuted Beef	1905/07	North America	USA-Washington	Lexington
5	PA7341AA	NA	NCTC5722	NA	NA	NA	NA	NA	NA	Derby
5	PA7373AA	NA	NCTC8256	NA	NA	NA	NA	NA	NA	Meleagridis
5	PA7389AA	NA	NCTC13175	NA	NA	NA	NA	NA	NA	Goldcoast
5	QA5280AA	ERR1817508	MS160389	Human	Human	<i>Homo sapiens</i> ; Human	1905/07	Europe	Ireland	Apapa

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
5	SA8719AA	SRR5576114	FSIS1701130	Livestock	Swine	Product-Raw-Intact-Pork	1905/07	North America	USA-North Carolina	Rissen
5	SA9671AA	SRR5584547	PNUSAS011882	Human	Human	Age: 30-39; Stool	2015/03	North America	USA	Corvallis (Predicted)
5	TA2653AA	SRR5647067	FSIS1701312	Poultry	Poultry	Raw Intact Chicken	1905/07	North America	USA-Georgia	Kentucky
5	TA6377AA	SRR5761324	PNUSAS016179	Livestock	Swine	Beef	2017/05	North America	USA-Wisconsin	Soerenga (Predicted)
5	TA8487AA	SRR5817827	FDA209009	Food	Bovine	Beef Bone: Foodon_0330535, Meat and Bone Meal: Foodon_00002738	2003/08	North America	USA	Havana
5	TA8944AA	SRR5820381	PNUSAS016799	NA	NA	NA	NA	North America	USA	NA
5	UA5320AA	SRR5924434	2015AM-0163	NA	NA	NA	NA	North America	USA	Lille (Predicted)
5	UA5376AA	SRR5926911	2015AM-1018	NA	NA	NA	NA	North America	USA	Kentucky (Predicted)
5	VA5581AA	SRR6038051	PNUSAS021676	NA	NA	NA	NA	North America	USA	I 45:b:- (Predicted)

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
5	WA4571AA	ERR2177148	DTU2016_2029_ PRJ1084_ <i>Salmonella</i> _	Livestock	Swine	<i>Sus scrofa</i> <i>domesticus</i> ; Pig	1905/07	Europe	Denmark	NA
			livingstone_2014_ 60_15_1							
			21754	Human	Human	<i>Homo sapiens</i> ; Human	2014/05	Europe	UK-England	Manchester
			78745	Human	Human	<i>Homo sapiens</i> ; Human	2014/12	Europe	UK-England	Adelaide
			FDA165725	Environment	Plant	Oregano	2002/03	Asia	Turkey	Telaviv
			FAR0170	Aquatic Animal	Fish	Frozen Yellow Fin Tuna	2010/11	Asia	Indonesia	Lexington
			FSW0113	Food	Plant	Chili Powder	2013/03	Asia	India	Cerro
			Ke355-sc-2012- 07-27T11:52:33Z- 1458217	NA	NA	NA	NA	NA	NA	Typhimurium
			CFSAN001588	Livestock	Bovine	<i>Bos taurus</i> ; Feces, Dairy Cow	2004/09	North America	USA- Pennsylvania	Cerro
			PNUSAS002174	Human	Human	Age: 40-49; Urine	2016/04	North America	USA	Serotype Pending
6a	MA2182AA	SRR3711017	FSIS1606822	Livestock	Swine	Rte Product-Pork	1905/07	North America	USA-Missouri	Liverpool
			C629	Poultry	Poultry	Chicken Carcass	2014/12	Asia	China	Indiana
			ADRD-L787	Food	Swine	Meat Meal	1905/06	South America	Brazil	[6],[14],18:Z4, Z23:[1,5]
6a	PA8554AA	SRR5173682								

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
6a	RA8952AA	SRR5429819	PNUSAS009827	Human	Human	Age: 70-79; Stool	2017/02	North America	USA	Kiambu (Predicted)
6a	WA4868AA	SRR6203663	PNUSAS020535	Human	Human	NA	1905/07	North America	USA-Wyoming	Adelaide (Predicted)
6	BA4248AA	SRR1411132	FMA0191	food	Plant	Kaju Mix (Cashew Mix)	2011/10	Asia	India	Mgulani
			<i>Salmonella</i> _ser_ Paratyphi_B_05_							
6	BA7799AA	ERR226484	5227-sc-2012-09- 04T12:40:57Z- 1476173	NA	NA	NA	NA	NA	NA	Paratyphi B
6	BA8969AA	ERR573933	PO579	NA	NA	NA	NA	Africa	Nigeria	Typhi
6	EA3440AA	SRR2107132	FDA324225 2-1	Food	Plant	Tahina	2005/04	Asia	Lebanon	Senftenberg
6	EA4121AA	SRR2057790	2015K-0409	Human	Human	Age: 60-69; Stool	2015/04	North America	USA	Derby
6	EA7330AA	SRR1967676	65391	Human	Human	<i>Homo sapiens</i> ; Human	2014/09	Europe	UK-England	NA
6	FA1025AA	SRR1646569	FDA631614 C2 5-1	Food	Plant	Garlic Powder	2010/06	Asia	India	47:Z4Z23:-
6	FA3119AA	SRR1548394	MOD1_Sal_881	Human	Human	<i>Homo sapiens</i> ; Stool	2010/01	North America	USA-Alabama	Stanleyville
6	HC4842AA	GCF_009648795	41578	Human	Human	<i>Homo sapiens</i> ; Human	1905/07	North America	USA-Ohio	Heidelberg
6	IA7353AA	ERR1082775	BS_00165	NA	NA	NA	NA	Africa	Gambia	NA
6	KA0046AA	SRR3038207	FDA192654A	Wild Animal	Reptile	<i>Gecko gecko</i>	2002/09	Asia	China	Virchow

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血型
6	KA1778AA	SRR3081685	CFSAN031809	Human	Human	<i>Homo sapiens</i> ; Feces	2013/08	Africa	Ethiopia-Addis Ababa	Virchow
6	KA3425AA	GCF_000953495	1326/28	Poultry	Poultry	Chicken; Central Veterinary Lab, Weybridge	1905/05	Europe	UK-England	Infantis
6	KA9800AA	ERR1363512	53_3	Livestock	Bovine	Cattle	1905/07	Africa	Nigeria	Eko
6	LA4036AA	ERR1082911	BS_14918-sc-2322486	NA	NA	NA	NA	Africa	Gambia	Hull (Predicted)
6	MA7660AA	SRR4241279	FDA144381-1	Environment	Plant	Fennel	2001/08	Africa	Egypt	Emek
6	MA7706AA	SRR4243036	CFSAN031406	Livestock	Swine	Swine; Feces	2013/04	Africa	Tanzania	Colindale (Predicted)
6	NA1463AA	SRR4421062	PNUSAS004463	Human	Human	Age: 70-79; Urine	2016/08	North America	USA	I 9,46: Nonmotile
6	OA2311AA	SRR5060586	PNUSAS005740	Human	Human	Age: 5-9; Stool	2016/07	North America	USA	Bareilly (Predicted)
6	PA7348AA	NA	NCTC7102	NA	NA	NA	NA	NA	NA	NA
6	PA7396AA	NA	NCTC5717	NA	NA	NA	NA	NA	NA	Heidelberg
6	PA9341AA	SRR5189045	PNUSAS007189	Human	Human	Age: 70-79; Urine	2016/12	North America	USA	Kintambo (Predicted)
6	QA4517AA	SRR5230181	OSF058121	Livestock	Bovine	Buffalo Horn	1905/07	North America	USA- Washington	Isangi (Predicted)
6	QA8791AA	SRR5335967	PNUSAS008822	Human	Human	Age: 0-4; Stool	2016/08	North America	USA	NA

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
6	RA9678AA	SRR5387490	ADRD-1002	Wild Animal	Avian	Veterinary Clinical/Research; Other Animal; Intestine Psittacine	1995/10	North America	USA-Oklahoma	Enteritidis
	RA9809AA	SRR5382730	PNUSAS010324	Human	Human		2017/02	North America	USA-Wyoming	Heidelberg (Predicted)
	TA2474AA	SRR5645417	PNUSAS011270	Human	Human		2017/03	North America	USA	NA
	TA2913AA	SRR5655461	PNUSAS013958	Human	Human		2017/05	North America	USA	Ohio
	TA9704AA	SRR5832650	PNUSAS016800	NA	NA		NA	North America	USA	NA
	UA6856AA	SRR5947770	PNUSAS019001	NA	NA		NA	North America	USA	Infantis (Predicted)
6	UA7785AA	SRR5962555	FDA1018254-001-001	Food	Plant	Seeded Vegetables (Solanaceous); Chili Whole Pepper	2017/07	Asia	India	Livingstone (Predicted)
6	ABFH02	NA	NA	NA	NA	NA	NA	NA	NA	NA
7	BA3330AA	SRR1257310	FSW0112	Aquatic Animal	Fish	Striped Bass Whole, Frz	2013/03	Asia	Taiwan, China	Lansing
7	BA4548AA	SRR1027786	FL_FLDACS-96405	Environment	Plant	Cilantro	2009/12	North America	USA-Florida	Hvittingfoss (predicted)
7	BA4755AA	SRR1525572	2013K-0640	Human	Human	Age: 20-29; Stool	2013/04	North America	USA	Saintpaul (Predicted)
7	CA5643AA	SRR1645995	H124580295	Human	Human	<i>Homo sapiens</i> ; Human	2012/11	Europe	UK-England	Newport

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
7	EA2449AA	ERR433181	88_500	NA	NA	NA	NA	NA	NA	Dublin
7	EA4852AA	SRR2015693	FDA359549	Wild Animal	Amphibian	Frog, Cleaned	2005/12	Asia	Thailand	Wandsworth
7	EA6432AA	ERR317005	M554/ NCTC20376	NA	NA	NA	1935/12	Europe	Germany-Lower Saxony	Typhisuis
7	EA8847AA	SRR1952758	FDA901347-1-A	Aquatic Animal	Fish	Frz. Raw Whole Round Large Mouth Bass, Bulk Pack	2015/03	Asia	Taiwan, China	Virchow (Predicted)
7	EA9164AA	SRR1915426	PNUSAS000165	Human	Human	Age: 20-29; Stool	2015/01	North America	USA	Baildon
7	FA0974AA	SRR1648175	VA-WGS-00353	Wild Animal	Avian	Laridae	2012/06	North America	USA-Virginia	Singapore
7	FA1304AA	SRR1643164	H123180685	Human	Human	<i>Homo sapiens</i> ; Human	2012/08	Europe	UK-England	Umbilo
7	FA3553AA	SRR1535705	FSF0117	Environment	Plant	Fenugreek Leaves	2008/04	Asia	Pakistan	Telhashomer
7	FA3917AA	SRR1501628	FDA851691	Environment	Plant	Oregano Ground	2014/03	Europe	Greece	Abony (Predicted)
7	FA3920AA	SRR1501624	FDA853189 1-B	Aquatic Animal	Fish	Herring Hilsa	2014/03	Asia	Myanma	Brunei (Predicted)
7	FA3979AA	SRR1481651	MOD1_Sal_850	Human	Human	<i>Homo sapiens</i> ; Stool	2012/05	North America	USA - Pennsylvania	Coeln
7	FA4334AA	SRR1300674	FMA0170	Environment	Plant	Cumin Seeds	2011/08	Asia	Syria	Vitkin
7	FA4449AA	SRR1292306	FNW18r72	Food	Plant	Pepper - Cayenne Chili, Hot Peppers	2008/02	Africa	Ethiopia	Kingabwa
7	HA9442AA	SRR2533410	FDA321962 1-1 C2	Food	Plant	Pepper Powder	2005/03	Asia	China	Kotbus (Predicted)

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
7	HC4846AA	SRR2544674	CFSAN000189	Food	NA	Food	2003/05	Asia	India	Bareilly
7	IA5597AA	ERR369397	29_61	NA	NA	NA	NA	NA	NA	Enteritidis
7	IA9120AA	SRR1966118	27066	Human	Human	<i>Homo sapiens</i> ; Human	2014/07	Europe	UK-England	Fulica
7	IC1230AA	ERR424913	SARB19_FB	NA	NA	NA	1905/06	Europe	Switzerland	Enteritidis
7	JA4191AA	SRR1646228	H124660719	Human	Human	<i>Homo sapiens</i> ; Human	2012/11	Europe	UK-England	Kenya
7	JA6278AA	SRR1964469	FDA457694 2-1	Food	Plant	Dhana Jeera Powder	2008/02	Asia	India	Java monophasic
7	JA8292AA	SRR2532676	FDA218712-1	Environment	Plant	Huiled Sesame	2003/02	Asia	India	Freiburg
7	JA8984AA	SRR2960582	CFSAN041501	Environment	Plant	Italian Pine Nuts	2015/10	North America	USA- Washington	Baldon
7	JA9452AA	FDA645338	SRR2979529	Environment	Plant	Macadamia Nuts Clusters	2010/07	North America	USA- New York	Aberdeen (Predicted)
7	JA9653AA	SRR2973723	FSIS1501522	Livestock	Bovine	Animal-Cattle- Beef Cow	1905/07	North America	USA-Florida	Reading
7	JA9825AA	SRR3002514	PNUSAS001232	Human	Human	Age: 30-39; Stool	2008/07	North America	USA	Isangi
7	KA1801AA	SRR3081662	CFSAN031786	Human	Human	<i>Homo sapiens</i> ; Feces	2005/10	Africa	Ethiopia-Addis Ababa	Kotbus
7	KA2057AA	SRR3088928	CFSAN031675	Livestock	Swine	Swine; Feces	2005/09	Africa	Ethiopia-Addis Ababa	NA
7	KA2709AA	SRR3115977	CFSAN031753	Livestock	Bovine	Bovine; Feces	2013/04	Africa	Ethiopia-Addis Ababa	Saintpaul
7	KA3377AA	GCF_001454965	FORC_020	Human	Human	Veinous Blood	2014/08	Asia	South Korea	Newport

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
7	KA3396AA	SRR955200	ATCC 10708	NA	NA	NA	NA	NA	NA	Braenderup
7	KA3414AA	SRR2533537	ATCC 8391	NA	NA	Clinical/Research	NA	NA	NA	Thompson
7	KA3418AA	GCF_000973685	CDC 2010K-2159	Human	Human	<i>Homo sapiens</i> ; Stool	2010/08	North America	USA	Newport
7	KA3428AA	GCF_000940935	USDA-ARS- USMARC-1927	Environment	Animal- Related	Livestock- Associated Habitat	2011/01	North America	USA	Newport
7	KA3451AA	GCF_000742815	C500	Poultry	Poultry	Chicken	2010/01	Asia	China-Jiangsu	Choleraesuis
7	KA3454AA	SRR921435	CFSAN001921	Poultry	Poultry	Food	1905/07	North America	USA-New York	Typhimurium
7	KA3481AA	SRR1183972	EC20110357	Human	Human	<i>Homo sapiens</i> ; Human-Sporadic	1905/06	North America	Canada-Alberta	Enteritidis
7	KA3497AA	SRR955196	ATCC BAA-1592	Environment	Plant	Tomatoes	NA	North America	USA- Pennsylvania	Anatum
7	KA3510AA	GCF_000020925	CT_02021853	NA	NA	NA	NA	NA	NA	Dublin
7	KA3517AA	GCF_000018705	SPB7	Human	Human	Stool	1905/06	Asia	Malaysia- Penang	Paratyphi B var. Java
7	KA3519AA	GCF_000018385	SARB49/ RKS4594/IP2/88	Human	Human	NA	1905/06	Europe	France	Paratyphi C
7	KA3520AA	GCF_000016045	SL254	NA	NA	NA	NA	NA	NA	Newport
7	KA3522AA	GCF_000009525	287/91/NCTC 13346	NA	NA	Chicken	NA	South America	Brazil	Gallinarum

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
7	KA3526AA	GCA_000006945	LT2; SGSC1412;	Laboratory	Laboratory	NA	1905/05	NA	NA	Typhimurium
			ATCC700720;							
			SARA2;							
			ATCC19585;							
			NCTC12416							
7	KA3643AA	SRR3156751	FDA340434-4	Wild Animal	Rodent	Flying Squirrel Excreta	2005/07	Asia	China	Thompson
7	KA4907AA	SRR3185219	PNUSAS001581	Human	Human	Age: 20-29; Stool	2014/08	North America	USA	Hvittingfoss (Predicted)
7	KA5720AA	SRR3210365	FDA939385	Environment	Plant	Cucumbers	2016/02	North America	USA-Florida	Aberdeen (Predicted)
7	KA7037AA	SRR3295863	CVM N57680F	Livestock	Swine	Animal-Swine-Sow	1905/07	North America	USA-Georgia	Chailey
7	KA8875AA	SRR3331085	FSIS1606117	Livestock	Swine	Animal-Swine-Sow	1905/07	North America	USA-Illinois	Saintpaul
7	LA4856AA	SRR3572992	PNUSAS002082	Human	Human	Age: 40-49; Stool	NA	North America	USA	Holcomb
7	LA4921AA	SRR3547964	FDA957243-2	Environment	Plant	Zaatar	2016/04	Asia	Isreal-Bnei	Richmond (Predicted)
									Shimon Regional Council	
7	MA1291AA	SRR3934137	CFSAN042365	Environment	Water/River	Water	2015/07	North America	USA-Ohio	Amager (Predicted)
7	MA2957AA	SRR3999889	PNUSAS003114	ND	ND/Others	Stool	2016/07	North America	USA	Irumu

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
7	MA5543AA	SRR4044440	2014AM-2256	NA	NA	NA	NA	North America	USA	Potsdam (Predicted)
7	NA5739AA	SRR4841727	PNUSAS004693	Human	Human	Age: 20-29; Stool	2016/08	North America	USA	Saintpaul (Predicted)
7	OA1338AA	ERR1737316	L01465-14	NA	NA	NA	NA	NA	NA	NA
7	PA2828AA	SRR5116290	FDA824513- C001-001	ND	ND/Others	Millet	2013/08	Europe	Ukraine	NA
7	PA3226AA	SRR5081401	PNUSAS004335	Human	Human	Age: 30-39; Stool	2013/06	North America	USA	NA
7	PA7344AA	traces-0yvJMKI	NCTC6245	NA	NA	NA	NA	NA	NA	Manhattan
7	PA7361AA	traces-0BiQQrR	NCTC4444	NA	NA	NA	NA	NA	NA	Enteritidis
7	PA7395AA	traces-0ovkgVJ	NCTC6802	NA	NA	NA	NA	NA	NA	Hartford
7	PA7407AA	traces-0PbVeix	NCTC9684	NA	NA	NA	NA	NA	NA	Paratyphi B var. Java
7	PA7408AA	traces-0usBeLC	NCTC92	NA	NA	NA	NA	NA	NA	Stanley
7	PA7409AA	traces-0APxmDg	NCTC5706	NA	NA	NA	NA	NA	NA	Java monophasic
7	PA7416AA	GCF_001890445	CFSAN003887	Environment	Plant	Nuts; Raw Almond Kernel	2006/09	North America	USA-California	Muenchen (Predicted)
7	PA7630AA	GCA_000499465	3114	NA	NA	Nonpareil NA	NA	NA	NA	Bovismorbificans
7	QA4197AA	ERR1816694	MS160254	Human	Human	<i>Homo sapiens</i> ; Human	1905/07	Europe	Ireland	NA
7	QA6097AA	SRR5251245	NC_P04-01GX	Poultry	Poultry	Poultry	1905/07	Africa	Uganda	NA
7	QA8981AA	SRR5336279	PNUSAS008470	NA	NA	NA	NA	North America	USA	NA

群组	序列编号	检索号	菌株名	来源类型	具体来源	详细信息	采集时间	来源大洲	采集地点	血清型
7	UA7425AA	SRR5951691	FDA1014609-C001-002	Food	Composite Food	Multi-Ingredient; Corn Flakes; Snack Food	2017/06	Asia	India	NA
7	UA8964AA	SRR5993054	PNUSAS021357	NA	NA	NA	NA	North America	USA	Saintpaul (Predicted)
7	VA8064AA	SRR6107610	PNUSAS024056	Human	Human	Human	NA	North America	USA	NA
7	WA5309AA	SRR6217794	PNUSAS025241	NA	Plant	NA	NA	North America	USA	NA
7	ABFG01	NA	NA	NA	NA	NA	NA	NA	NA	NA

致 谢

首先，感谢中科院上海免疫与感染所的 Daniel Falush 教授为我提供生物信息学科研训练的机会，并对我的毕业论文给予指导。感谢 Daniel 实验室的刘思奇师姐对文中 SNP loading 部分的计算提供指导和帮助，并且提供肺炎克雷伯菌的对应计算结果用于比较分析，感谢 Roberto C. Torres 博士针对 fineSTRUCTURE 软件的使用给予我耐心的帮助，感谢实验室其他老师、同学一年以来对我的关心与照顾。

四年的本科学习即将结束，此处我还想感谢在本科前三年为我提供的宝贵学习机会和科研指导的郑丙莲老师，在郑老师实验室我受到了基础且系统的训练，认识了科学研究最真实的一面，也体会到了科学的魅力。感谢郑老师实验室的李磊博士和胡倩倩师姐在实验方面的帮助和在生活上的照顾，与郑老师实验室同学们一起度过的三年时光是我难忘且宝贵的记忆。

此外我要感谢在本科二年级时激发起我好奇心和探索欲的化学系雷杰老师；感谢悉心帮助我，对于我未来规划产生深刻影响的蔡亮老师；感谢一年多以来一直鼓励我，介绍我和 Daniel 相识，并由此改变了我的学习轨迹的徐书华老师。

最后感谢始终陪伴我的家人和朋友。

行至此处，何其幸运；不一而足，唯有感恩。