

基于比较基因组学对新疆不同地域人源长双歧杆菌长亚种的遗传差异分析

张慧敏^{1,2}, 张雪玲², 孙海龙^{1,2*}, 倪永清^{1,2*}

1 石河子大学 食品学院, 新疆 石河子

2 新疆特色益生菌与乳品技术八师重点实验室, 新疆 石河子

张慧敏, 张雪玲, 孙海龙, 倪永清. 基于比较基因组学对新疆不同地域人源长双歧杆菌长亚种的遗传差异分析[J]. 微生物学报, 2026, 66(1): 301-321.

ZHANG Huimin, ZHANG Xueling, SUN Hailong, NI Yongqing. Comparative genomics reveals the genetic differences of human-derived *Bifidobacterium longum* subsp. *longum* in different regions of Xinjiang[J]. Acta Microbiologica Sinica, 2026, 66(1): 301-321.

摘要: 【目的】以新疆不同地域、不同人群来源的长双歧杆菌长亚种(*Bifidobacterium longum* subsp. *longum*)为研究对象, 解析肠道专性共栖细菌在宿主人群中的系统发育和基因组多样性, 为针对不同人群开发个性化功能益生菌提供理论依据。【方法】从地理跨度较大的新疆喀什、伊犁的维吾尔族和哈萨克族母婴人群中分离出 136 株 *B. longum* subsp. *longum*, 并通过公共数据库整合来自国内其他地域人群的菌株进行比较基因组分析。【结果】*B. longum* subsp. *longum* 基因组的平均大小、G+C 含量、编码序列数量分别为 2.38 Mb、59.91%、2 160 个。基于核心基因构建的系统发育树显示, 来自新疆的所有菌株在系统发育树上属于 4 个进化支(clade), 来自同一民族、不同地域人群的菌株其系统发育呈现不同的进化支, 地理上更相近、不同人群来源的菌株存在一定程度的重叠。对更大地理范围(全中国)的分析显示, *B. longum* subsp. *longum* 菌株呈现明显的东西部地理来源聚集特征和协同进化现象。基于附属基因和碳水化合物代谢功能基因的分析显示, 来源于同一系统发育进化支的同一民族、不同地域人群菌株的功能基因谱丰度也差异较大; 碳水化合物水解酶家族 GH13 (α -淀粉酶)和 GH43 (β -木糖苷酶)等在喀什地区菌株基因组中更为丰富; 相反, 地理上相近、即使来自不同民族人群的菌株, 其基因组中蛋白质直系同源簇(cluster of orthologous groups of proteins, COG)功能基因谱、碳水化合物水解酶关联基因家族谱更相似。【结论】新疆不同地域人群来源的 *B. longum* subsp. *longum* 菌株系统发育呈现明显的地理和民族分布特征以及功能基因分布特征, 地理尺度范围越大, 菌株的地理分布特征越明显。需要基于更大范

资助项目: 新疆生产建设兵团重点领域科技攻关项目(2023AB050, 2024AB050); 石河子市财政科技计划(2024GY02); 国家自然科学基金联合基金(U1903205)

This work was supported by the Special Funds for Science and Technology Development of Xinjiang Production and Construction Corps (2023AB050, 2024AB050), the Financially Science and Technology Project of Shihezi (2024GY02), and the Joint Funds of the National Natural Science Foundation of China (U1903205).

*Corresponding authors. E-mail: SUN Hailong, sunhl610@shzu.edu.cn; NI Yongqing, niyqlzu@sina.com

Received: 2025-07-05; Accepted: 2025-07-23; Published online: 2025-08-22

围菌株基因组的大数据进一步验证人群的地理分布尺度与菌株协同进化和特异性的关系。

关键词：长双歧杆菌长亚种；比较基因组学；碳水化合物活性酶

Comparative genomics reveals the genetic differences of human-derived *Bifidobacterium longum* subsp. *longum* in different regions of Xinjiang

ZHANG Huimin^{1,2}, ZHANG Xueling², SUN Hailong^{1,2*}, NI Yongqing^{1,2*}

1 School of Food Science and Technology, Shihezi University, Shihezi, Xinjiang, China

2 Key Laboratory for Xinjiang Special Probiotics and Dairy Technology of the Eighth Division of Xinjiang Production and Construction Corps, Shihezi, Xinjiang, China

Abstract: [Objective] To study the phylogenetic relationship and genomic diversity of intestinal obligate commensal bacteria in different populations from various regions of Xinjiang and provide a theoretical basis for developing personalized functional probiotics for different populations. **[Methods]** A total of 136 strains of *Bifidobacterium longum* subsp. *longum* were isolated from mother-infant populations of Uygur and Kazak ethnic groups in Kashgar and Yili regions of Xinjiang. Comparative genomic analysis was conducted with data of the strains from other regions in China that were available in public databases. **[Results]** The average genome size, G+C content, and the number of coding sequences of *B. longum* subsp. *longum* were 2.38 Mb, 59.91%, and 2 160, respectively. The phylogenetic tree constructed based on core genes showed that all strains from Xinjiang belonged to four clades in the phylogenetic tree. Strains from the same ethnic group but from different geographical regions were in different clades, and there was a certain degree of overlap between geographically closer and different population-derived strains. The analysis of a larger geographical range (China) showed that *B. longum* subsp. *longum* strains and their functional genes presented obvious geographical and ethnic distribution characteristics. The analysis of COG functional genes and carbohydrate hydrolyase-related genes showed that the functional gene spectra varied greatly among strains from the same ethnic group but in different regions. The carbohydrate hydrolyase-related gene families GH13 (α -amylases) and GH43 (β -amylases) were more abundant in the strains from Kashgar region. Conversely, even strains from different ethnic groups but from geographically close regions had similar spectra of COG functional genes and carbohydrate hydrolyase-related gene families. **[Conclusion]** The *B. longum* subsp. *longum* strains and their functional genes from different geographical regions and ethnic groups in Xinjiang showed obvious geographical and ethnic distribution characteristics. As the geographical scale becomes large, the geographical distribution characteristics of the strains become more obvious. The relationship between the geographical distribution scale of populations and the co-evolution and specificity of strains should be verified based on larger-scale genomic data of strains.

Keywords: *Bifidobacterium longum* subsp. *longum*; comparative genomics; carbohydrate-active enzymes

人体肠道是由数以万计的微生物构成的复杂动态生态体系, 成百上千种微生物共同维持着肠道微生态平衡。与人类肠道特异性共栖的物种被称为人类关联微生物(human-associated mutualists and commensals), 在人类专性共生、相互适应过程中形成协同进化关系的肠道细菌被称为共生体^[1-2]。宿主为肠道共生体提供了营养和稳定的栖息环境, 这对宿主的健康和发育具有重要意义^[3-5]。双歧杆菌(*Bifidobacterium*)是肠道共生体的典型代表, 作为人体肠道核心菌群的重要组成部分, 其定殖始于新生儿出生阶段, 并伴随宿主终生^[6]。尽管这种分类方式缺乏严格的科学依据, 但根据人类生活史以及双歧杆菌在肠道中的优势度和流行率, 人源双歧杆菌(human-residential bifidobacteria, HRB)可分为婴儿型(infant-type HRB)和成人型(adult-type HRB), 前者包括两歧双歧杆菌(*Bifidobacterium bifidum*)、长双歧杆菌长亚种(*Bifidobacterium longum* subsp. *longum*)、长双歧杆菌婴儿亚种(*Bifidobacterium longum* subsp. *infantis*)和短双歧杆菌(*Bifidobacterium breve*), 其特点是能够代谢各种人源聚糖(包括母乳寡糖和肠道黏膜多糖); 后者包括 *Bifidobacterium longum* subsp. *longum*、青春双歧杆菌(*Bifidobacterium adolescentis*)、假小链双歧杆菌(*Bifidobacterium pseudocatenulatum*)和链状双歧杆菌(*Bifidobacterium catenulatum*)等, 这些菌种普遍以代谢宿主饮食来源的各種植物源聚糖来获取能源^[7]。其中, *B. longum* subsp. *longum* 具有最广泛的生态位分布, 在婴幼儿、成人以及百岁老人肠道中均普遍存在, 这与其丰富的遗传多样性和多变的聚糖代谢能力密切相关, 少部分菌株甚至可以代谢人源寡糖, 大部分菌株代谢植物源聚糖, 有的甚至兼具代谢两类聚糖的功能, 这使得它们在不同人群肠道中具有很强的适应性和竞争力^[8-10]。

研究发现动物宿主肠道的严格共栖细菌与宿主存在协同进化关系, 即种群的遗传分化呈现明显的宿主特异性^[11-12]。近十几年, 作为研

究哺乳动物肠道共栖细菌共生体进化的范式物种, 罗伊氏乳杆菌(*Limosilactobacillus reuteri*)呈现出明显的宿主特异的系统发育谱系结构^[13-14]。对于更加严格厌氧的肠道共生菌, 目前针对类人猿的肠道微生物群研究证实, 部分拟杆菌和双歧杆菌谱系与宿主存在协同进化, 菌群的系统进化与来源宿主具有明显的对应关系^[4,15-16]。针对不同人群的宏基因组组装基因组(metagenome assembled genomes, MAGs)研究发现, 一些双歧杆菌种群也体现出这种协同进化的趋势, 但尚未开展基于分离菌株基因组学的深入研究^[17-18]。*B. longum* subsp. *longum* 作为专性共生细菌的典型代表, 受宿主遗传因素(如基因型、免疫特征)和环境因素(如饮食结构、地理区域)的共同影响, 其生理和遗传均呈现出最丰富的多样性^[19-21]。母乳喂养婴儿体内的 *B. longum* subsp. *longum* 菌株携带更多母乳低聚糖(human milk oligosaccharides, HMOs)利用基因, 而配方奶喂养婴儿的菌株则显著富集代谢乳糖和淀粉类基因, 老年人肠道中的菌株富集降解膳食纤维类基因, 如阿拉伯木聚糖酶系统^[9-10,20,22]。此外, 宿主的免疫系统通过调控肠道黏膜环境影响 *B. longum* subsp. *longum* 的定殖^[23-24]。*B. longum* subsp. *longum* 在维持人类肠道健康、预防和治理肠道相关疾病方面具有巨大的应用潜力^[25-26]。因此通过开展比较基因组学研究和功能基因组学分析, 揭示人类肠道微生物组中 *B. longum* subsp. *longum* 菌株的遗传多样性至关重要。

我国新疆地域广袤、多民族聚集, 但各民族聚集地空间距离较远、相对隔离, 民族群体之间通婚率低, 饮食多样。本研究从新疆喀什、伊宁的维吾尔族和尼勒克县的哈萨克族母婴人群中获取了 136 株 *B. longum* subsp. *longum* 菌株, 采用比较基因组的方法从全基因组水平揭示不同来源菌株的基因组特征和关键功能基因的差异。同时, 从公共数据库下载 114 株来源于中国其他地区人群菌株的基因组数据, 进行了基于核心基因的系统进化关系分析, 以期为

以长双歧杆菌为代表的人体肠道共生菌的协同进化理论提供证据,并为开发针对特定地域人群的个性化益生菌菌株干预策略奠定理论基础。

1 材料与方法

1.1 材料

1.1.1 参与者和样本采集

本研究采集长期居住于新疆喀什市、伊宁市和尼勒克县的人源粪便样品。采样前已向志愿者告知采样目的,志愿者同意并签署家属告知页及知情同意书。采样时取样者佩戴无菌手套,用取样勺挖取 2–5 g 粪便放入取样管中,注意避免粪尿混合。采集到的样品置于车载冰箱并立即运输至实验室,处理前保存于 -80°C 。本研究通过石河子大学医学院第一附属医院医学伦理委员会批准,编号为 KJ2021-088-01。

1.1.2 培养基

改良 MRS 固体培养基(g/L): 蛋白胨 10.00, 牛肉膏 10.00, 酵母提取物 5.00, 葡萄糖 20.00, 吐温-80 1.00 mL, K_2HPO_4 2.00, 乙酸钠 5.00, 柠檬酸二铵 2.00, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 0.58, $\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$ 0.25, 琼脂 20.00, L-半胱氨酸盐酸盐 0.50, 莫匹罗星 0.05, pH 6.0, 115°C 灭菌 20 min。

改良 Wilkins-chalgren 培养基(g/L): 改良 Wilkins-chalgren 琼脂 48.00, 大豆蛋白胨 5.00, 吐温-80 1.00 mL, 冰醋酸 1.00 mL, L-半胱氨酸盐酸盐 0.50, 莫匹罗星 0.05, 诺氟沙星 0.20, 115°C 灭菌 20 min。

1.2 主要试剂和仪器

TE 缓冲液 [10 mmol/L Tris (pH 8.0), 10 mmol/L 乙二胺四乙酸 (ethylene diamine tetraacetic acid, EDTA)]、*Taq* Plus Master Mix (2 \times)和 100 bp Plus Marker, 南京诺唯赞生物科技股份有限公司; 琼脂糖、GoldView, 北京全式金生物技术股份有限公司。

PCR 扩增仪, SensoQuest 公司; Power Pac

Universal 水平电泳仪, Bio-Rad 公司; BioNumerics v8.0, Applied Maths 公司; 引物合成和菌株 Sanger 测序均由苏州金唯智生物科技有限公司完成。

1.3 双歧杆菌菌株的分离和鉴定

将 1 g 新鲜粪便样品用 9 mL MRS 液体培养基进行梯度稀释, 吸取 100 μL 稀释度为 10^{-3} 、 10^{-4} 、 10^{-5} 的稀释液分别涂布于改良 MRS 培养基(MRS、0.5% L-半胱氨酸盐酸盐、莫匹罗星 25 mg/L)和改良 Wilkins-chalgren 培养基, 每个稀释度设置 2 个平行, 37°C 厌氧培养 48–72 h 后, 从每个平板选取 10–20 个凸起、光滑、边缘水润的白色或乳白色菌落, 经富集后进行镜检, 丢弃球菌, 将呈 V 型、Y 型、棍棒型的菌株在上述培养基上连续划线 2–3 次, 直至镜检出现纯菌株为止。挑选纯化后的单菌落于含有 1 mL 液体 MRS 培养基(补充添加 0.5% L-半胱氨酸盐酸盐)的 2 mL 菌种保藏管中, 厌氧富集 48 h 后, 向纯培养物中加入 0.5 mL 已灭菌的 75% 甘油, 将菌液与甘油混合均匀后置于 -80°C 保藏。

菌株 DNA 的提取采用苯酚-氯仿-玻璃珠击打的方法。采用引物 Bif-*groEL*-F (5'-TCCG ATTACGAYCGYGAGAAGCT-3')和 Bif-*groEL*-R (5'-CSGICYTCGGTSGTCAGGAACAG-3')对分离菌株的 DNA 进行 PCR 扩增。PCR 反应体系 (25 μL): 2 \times *Taq* Plus Master Mix (p213) 12.5 μL , 上、下游引物(20 $\mu\text{mol/L}$)各 0.5 μL , DNA 模板 3 μL , ddH₂O 8.5 μL 。PCR 反应条件: 95°C 预变性 8 min; 95°C 变性 40 s, 60°C 退火 40 s, 72°C 延伸 40 s, 共 30 个循环; 72°C 终延伸 8 min。PCR 产物经 1.2% 琼脂糖凝胶电泳检测后, 在 500 bp 左右的位置得到清晰单一条带, 将扩增产物送往苏州金唯智生物科技有限公司进行测序。

1.4 *B. longum* subsp. *longum* 基因组 *de novo* 测序

将保藏在 $-80\text{ }^{\circ}\text{C}$ 超低温冰箱中鉴定为 *B. longum* subsp. *longum* 的菌株取出, 接种于改良 MRS 肉汤培养基中, $37\text{ }^{\circ}\text{C}$ 厌氧培养 48 h, 传代培养至菌株恢复活力。在无菌条件下 $4\text{ }^{\circ}\text{C}$ 、8 000 r/min 离心 5 min 收集菌体。使用 FastPure[®] Bacteria DNA Isolation Mini Kit 提取菌株总 DNA, 采用革兰氏阳性菌的提取方法并在试剂盒说明书的基础上对部分提取方法稍作修改, 包括将溶菌酶的孵育时间延长至 30 min, 蛋白酶 K 孵育时间延长至 40 min。采用 NanoDrop 2000 检测提取 DNA 的纯度和浓度, 使用琼脂糖凝胶电泳法检测 DNA 的完整度。将浓度 $\geq 20\text{ ng}/\mu\text{L}$, 总量 $\geq 4\text{ }\mu\text{g}$, $OD_{260}/OD_{280}=1.8\text{--}2.0$ 的菌株 DNA 寄送至上海派森诺生物科技股份有限公司。采用 Illumina NovaSeq 测序平台, 每个样品提供不低于基因组 100 \times 的 PE150 测序数据量。

1.5 *B. longum* subsp. *longum* 比较基因组分析

本研究使用了 250 株 *B. longum* subsp. *longum* 菌株, 包括 136 株新测序菌株和 114 株从 NCBI 数据库中获取的公开 *B. longum* subsp. *longum* 菌株基因组数据。此外, NCBI 数据库中有关 *B. longum* 4 个亚种标准菌株的参考基因组序列, 即 *B. longum* subsp. *longum* JCM 1217、*B. longum* subsp. *infantis* ATCC 15697、*B. longum* subsp. *suis* LMG 21814 和 *B. longum* subsp. *suillum* JCM 19995 也纳入后续分析。所有菌株序列经过低质量 reads 过滤、去除接头和引物后, 获得高质量的 reads 序列。利用短序列组装软件 SOAPdenovo2 (<http://soap.genomics.org.cn/>) 对二代测序后的优化序列进行多个 Kmer 参数的拼接, 得到最优的 contigs 组装结果, 然后把 reads 比对到 contig 上, 根据 reads 的 paired-end 和 overlap 关系对组装结果进行局部组装和优

化, 形成 scaffolds。另外, NCBI 数据库中 *B. longum* subsp. *longum* JCM 1217 参考基因组序列也纳入后续分析。

采用 Roary 软件对其核心基因(core gene)进行统计, 以氨基酸序列一致性 $>90\%$ 为标准划分基因家族, 构建 *B. longum* subsp. *longum* 的核心基因集和附属基因集(accessory gene)。利用 Roary 得到的核心基因序列使用 fasttree v2.1.10 软件中的 GTR 模型, 基于最大似然(maximum likelihood, ML)法构建 *B. longum* subsp. *longum* 的核心基因系统发育树。使用在线工具 iTOL (<https://itol.embl.de/>)美化系统发育树图。平均核苷酸一致性(average nucleotide identity, ANI)是通过比对基因组的同源序列以鉴定菌株亲缘关系。利用 FastANI 计算菌株两两之间的 ANI 值, 将整理后的 ANI 值使用 R 语言中 pheatmap 包绘制聚类热图。将预测得到的基因组相应的氨基酸序列文件, 通过碳水化合物活性酶数据库(carbohydrate active enzymes, CAZy)对所有菌株基因组序列进行碳水化合物活性酶的注释。使用 eggno-mapper v2.1.10 工具将编码区序列(coding sequences, CDS)与蛋白质直系同源(cluster of orthologous groups of proteins, COG)数据库进行比对, 得到基因功能注释信息。

2 结果与分析

2.1 双歧杆菌的分离和鉴定

本研究的志愿者主要来自新疆喀什市、伊宁市和尼勒克县, 其年龄主要分布在 0–2 岁和 25–32 岁之间。经 *groEL* 基因测序, 从 72 份新疆喀什市、伊犁地区、伊宁市和尼勒克县人群粪便样本中共分离得到隶属于 9 个种的 749 株双歧杆菌, 其中 *B. longum* subsp. *longum* 271 株(占比 36.18%)、*B. adolescentis* 149 株(占比 19.89%)、*B. breve* 145 株(占比 19.36%)、*B. bifidum* 65 株(占比 8.68%)、*B. animalis* 35 株(占比 4.67%)、*B. catenulatum* 35 株(占比 4.67%)、

B. pseudocatenulatum 30 株 (占比 4.01%)、*B. angulatum* 18 株 (占比 2.40%)、*B. merycicum* 1 株 (占比 0.13%) (图 1)。菌株分离鉴定结果表明, *B. longum* 在本研究的志愿者队列中普遍存在且为优势物种, 其次是 *B. adolescentis* 和 *B. breve*。

2.2 *B. longum* subsp. *longum* 基因组特征分析

本研究基于 Illumina NovaSeq 测序平台, 对分离自新疆 3 个不同地区的 136 株人源 *B. longum* subsp. *longum* 进行了全基因组测序, 包括伊宁市维吾尔族 31 株、喀什市维吾尔族 51 株和尼勒克县哈萨克族 54 株。对所有测序的 *B. longum* subsp. *longum* 基因组序列进行 *de novo* 组装, 并系统评估了菌株基因组的拼接质量和基因组组装信息。136 株 *B. longum* subsp. *longum* 组装结果显示, 基因组大小平均为 (2.38 ± 0.12) Mb, 平均 G+C 含量为 $(59.91 \pm 0.31)\%$, Scaffolds 数量平均为 80 个, N50 的平均长度为 298 557 bp, 编码序列

(coding sequence, CDS) 数量为 1 755–2 565 个。CheckM 质控结果表明, 所有纳入分析的基因组序列完整度 (completeness) 均在 99% 以上, 且污染度 (contamination) 在 5% 以下。

2.3 *B. longum* subsp. *longum* 核心基因和泛基因集的构建

在微生物领域, 一个物种的全部基因被定义为泛基因组 (pan-genome), 包括核心基因组和非核心基因组, 非核心基因组可进一步分为 3 类, 即软核基因组、云基因组和壳基因组。核心基因组是指存在于所有基因组序列中的基因家族, 具有高度保守性和稳定性。软核基因组指存在于 95% 的基因组的基因家族; 云基因组被定义为只存在于少数基因组序列中的基因家族 ($0 < \text{基因在菌株中的出现率} < 15\%$); 壳基因组包括除软核基因组和云基因组之外, 在大多数基因组序列中存在的基因家族 ($15\% < \text{基因在菌株中的出现率} < 95\%$)。以氨基酸序列一致性 $> 90\%$ 的标准划分基因家族的结果表明, 136 株

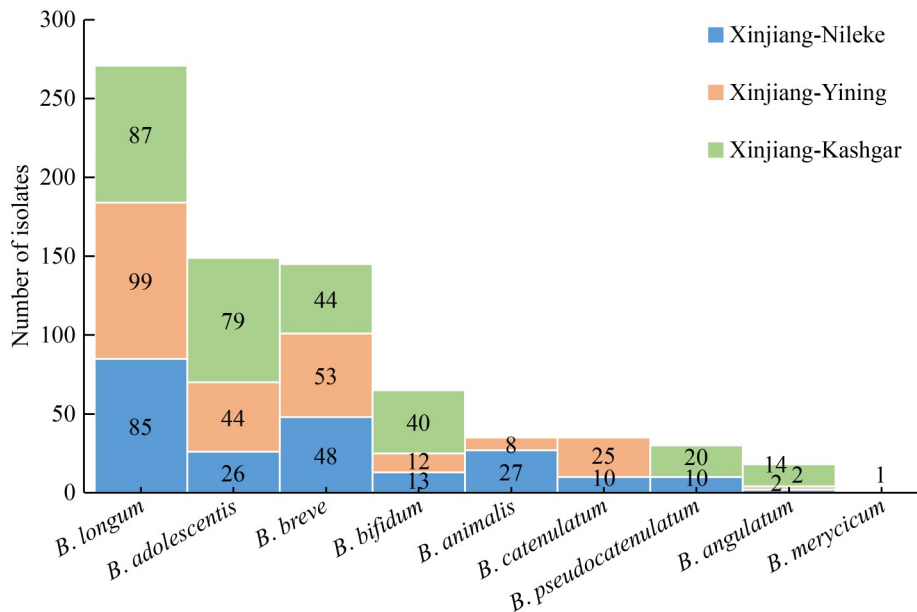


图1 新疆尼勒克县、伊宁市和喀什市人源双歧杆菌种水平的分布

Figure 1 Distribution of human-derived *Bifidobacterium* species in Nileke County, Yining City, and Kashgar City, Xinjiang.

B. longum subsp. *longum* 的泛基因组共包含 7 064 个基因, 其中包括 1 286 个核心基因、91 个软核基因、1 089 个云基因和 4 598 个壳基因(图 2A)。本研究通过绘制核心基因-泛基因组曲线发现, 随着菌株数目的增加, 泛基因组数量呈现上升趋势; 相比之下, 核心基因数量逐渐趋于稳定(图 2B)。

2.4 *B. longum* subsp. *longum* 核心基因系统发育树分析

基于 250 株 *B. longum* subsp. *longum*, 以及 *B. longum* subsp. *longum* JCM 1217、*B. longum* subsp. *infantis* ATCC 15697、*B. longum* subsp. *suis* LMG 21814 及 *B. longum* subsp. *suillum* JCM 19995, 利用核心基因通过 1 000 次引导迭代构建 *B. longum* 核心基因的最大似然树。核心基因系统发育树显示, 250 株 *B. longum* subsp. *longum* 和 *B. longum* subsp. *longum* JCM 1217 聚集在同一个进化分支上, 与 *B. longum* subsp. *infantis* ATCC 15697、*B. longum* subsp. *suis* LMG 21814 和 *B. longum* subsp. *suillum* JCM 19995 明

显分离(图 3A)。根据菌株来源信息注释的结果表明, 来自我国不同地理区域人源的 *B. longum* subsp. *longum* 呈现地域分布特征, 按照进化支可分为 10 个 clades。其中, clade1 和 clade5 主要以中国西部(新疆、西藏和广西等)人群分离株为主, clade2-4 以中国东部(包括江苏、浙江、四川、广东和北京等地区)的分离株为主; clade6-8 以中国中部地区(河南、山东和安徽等)的人源分离株为主; clade9 和 clade10 的分离株来源于中国西北地区和中部地区人群。

系统发育树可直观反映菌株间的群体结构和遗传进化关系。为研究 *B. longum* subsp. *longum* 种内遗传进化关系, 本研究基于 136 株 *B. longum* subsp. *longum*, 采用最大似然法, 利用 1 286 个核心基因通过 1 000 次引导迭代构建核心基因系统发育树。核心基因系统发育树显示, 本研究中分离的 136 株新疆地区人源 *B. longum* subsp. *longum* 分为 4 个不同的进化分支(图 3B)。根据菌株来源的民族信息注释结果表明, 进化支 A、B3、C1 及 D1 中维吾尔族人

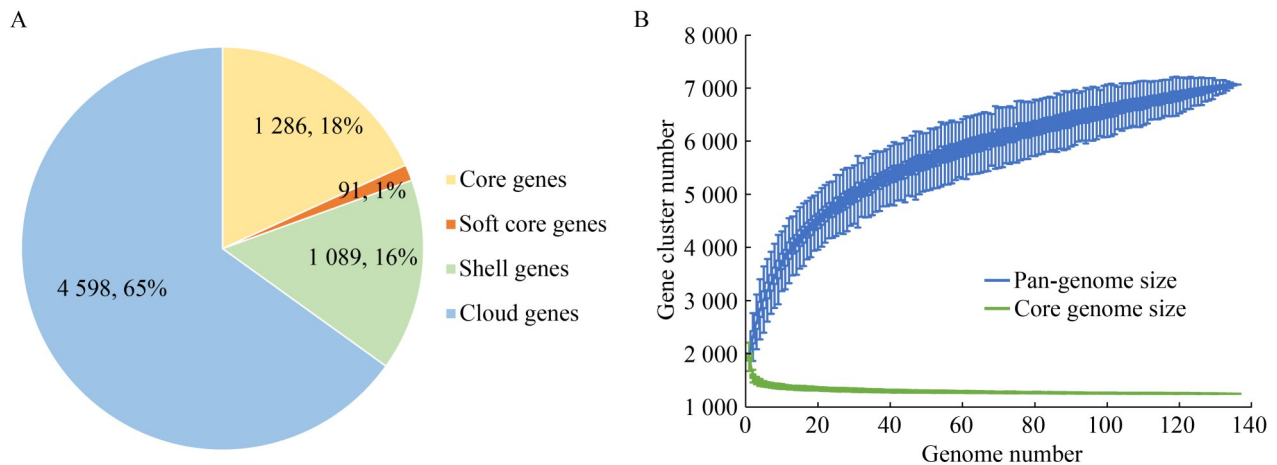


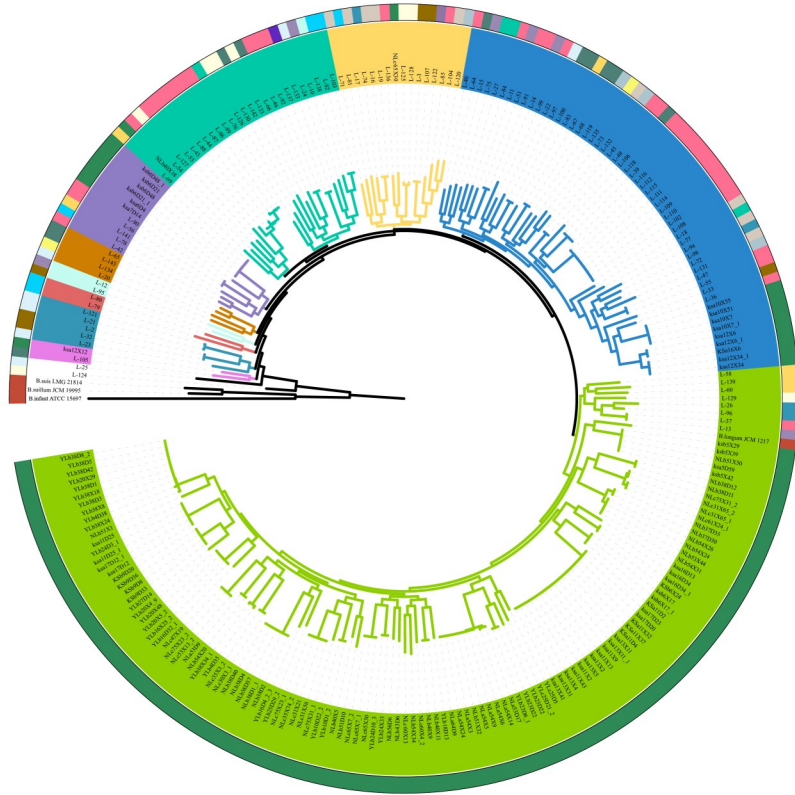
图2 *Bifidobacterium longum* subsp. *longum* 菌株的核-泛基因组。A: *Bifidobacterium longum* subsp. *longum* 菌株泛基因组[包括核心基因(core genes)、软核基因(soft core genes)、云基因(cloud genes)和壳基因(shell genes)]; B: 核心基因-泛基因组曲线(绿色代表核心基因, 蓝色代表泛基因组)。

Figure 2 Core genes and pan-genome of *Bifidobacterium longum* subsp. *longum* strains. A: *B. longum* subsp. *longum* strain pan-genome [Including core genes, soft core genes, cloud genes, and shell genes]; B: Core gene-pan-genome curves (Green represents core genes, and blue represents pan-genes).

A

Tree scale: 0.01

Geography	Clade
Reference	Clade1
Xinjiang	Clade2
Beijing	Clade3
Shandong	Clade4
Jiangsu	Clade5
Hubei	Clade6
Guangdong	Clade7
Henan	Clade8
Sichuan	Clade9
Gansu	Clade10
Yunnan	
Shanghai	
Anhui	
Fujian	
Zhejiang	
Xizang	
Guangxi	
Hainan	



B

Tree scale: 0.01

Geography
Xinjiang-Kashgar-Uyghur
Xinjiang-Yining-Uyghur
Xinjiang-Nileke-Uyghur

Ethnic
Uyghur
Kazakh

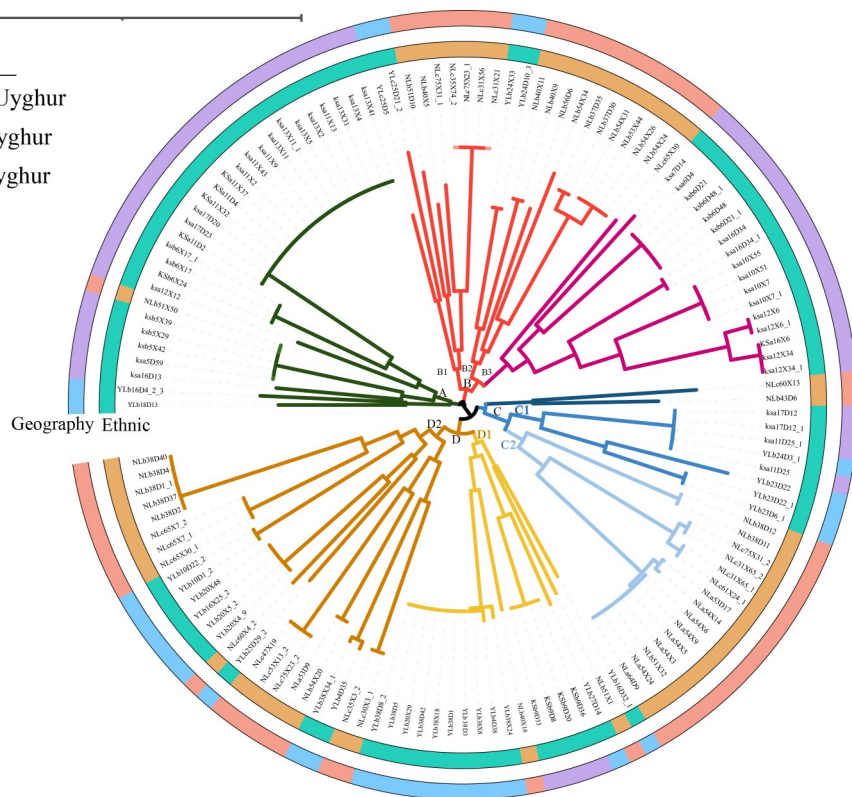


图3 *Bifidobacterium longum* subsp. *longum*核心基因系统发育树。A: 基于250株*B. longum* subsp. *longum*的核心基因构建的系统发育树(最外圈为样本来源的地理信息; 内圈为进化分支谱系, 不同的颜色代表不同的谱系); B: 基于分离的136株菌的核心基因序列使用ML法构建系统发育树(系统发育树最外圈为样本来源的地理信息, 内圈为民族信息)。

Figure 3 Phylogenetic tree of core genes of *Bifidobacterium longum* subsp. *longum*. A: A phylogenetic tree constructed based on the core genes of 250 strains of *B. longum* subsp. *longum* (The outermost circle represents the geographical information of the sample sources; The inner circle represents the evolutionary branching lineages, with different colors representing different lineages); B: Phylogenetic tree constructed using the core gene sequences of 136 isolated strains by the ML method (The outermost circle of the phylogenetic tree represents the geographical information of the sample sources, and the inner circle represents the ethnic information).

群分离株的占比分别为 96.55%、94.44%、100.00% 和 88.89%, 其中进化支 A 和 B3 菌株几乎全部来自喀什的维吾尔族母婴人群; B1、B2、C2 及 D2 进化分支中哈萨克族人群分离株占比分别为 77.78%、83.33%、100.00% 和 64.00%。值得注意的是, 根据地理来源进行分类时, D 进化分支表现为同一地域不同民族的分离株呈分簇聚集的特征, 即来自新疆伊犁哈萨克自治州 2 个地区的维吾尔族和哈萨克族来源分离株聚集。

2.5 *B. longum* subsp. *longum* 附属基因集分析

研究发现细菌附属基因的分布模式在一定程度上反映了菌株基因组的特异性。本研究将泛基因集中的核心基因(存在频率 $\geq 95\%$)剔除后, 基于 5 778 个附属基因的存在与缺失分布结果, 根据 Jaccard 距离矩阵, 基于菌株附属基因存在与否的“1/0”矩阵进行主成分分析。当分离株按照民族来源进行分类时, 维吾尔族和哈萨克族菌株的重叠性较高(图 4A)。进化支 A 和 B3 均为喀什地区维吾尔族人源分离株, B1、B2 和 C2 均为尼勒克哈萨克族人源分离株。NMDS 分析图的结果表明, 进化支根据地理和民族来源聚集。相较于新疆尼勒克县, 来自新疆喀什市和伊宁市 2 个不同地区的分离株分簇聚集, 而

来自相近地理区域的新疆伊宁市和尼勒克县的分离株呈现聚集趋势。本研究中的哈萨克族菌株均分离自新疆尼勒克县, 维吾尔族源菌株分离自新疆喀什和伊宁两地。综上所述, 附属基因的聚类模式表明, 虽然维吾尔族和哈萨克族的遗传背景不同, 但由于新疆伊宁市和尼勒克县的地理位置较为接近, 相近地理区域的不同族群可能因相似的生活习惯和饮食结构导致菌株进化模式趋同。新疆喀什和伊宁不同地区的维吾尔族之间由于生活环境和饮食习惯的差异, 导致它们具有分离源特异性。

2.6 ANI 分析

ANI 是通过比对基因组的同源序列来鉴定菌株亲缘关系。在比较基因组学分析过程中通常认为 ANI 值 $>95\%$ 为同一物种, ANI 可用于判断基因组间的相似性及评估基因组间多态性程度^[27]。本研究对 136 株 *B. longum* subsp. *longum* 菌株进行 ANI 计算, 聚类热图的结果表明, ANI 值均在 0.985 以上, 均为同一物种(图 5)。维吾尔族菌株组间的 ANI 值显著高于哈萨克族源菌株($P_{\text{adj}}=8\times 10^{-8}$), 组内不同地理区域组间 ANI 值的统计结果表明, 新疆尼勒克县哈萨克族菌株间的 ANI 值为 0.989 ± 0.003 , 低于新疆伊宁市维吾尔族(0.991 ± 0.003)和新疆喀什市维吾尔族(0.991 ± 0.004)菌株间的 ANI 值。显著

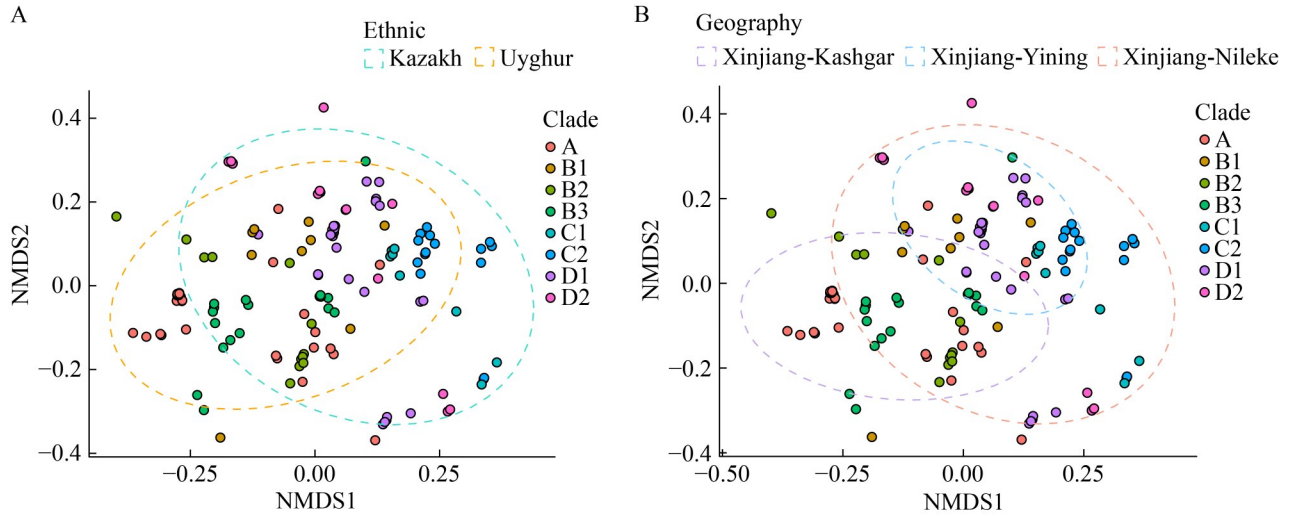


图4 *Bifidobacterium longum* subsp. *longum* 菌株附属基因聚类分析。A: 不同民族来源长双歧杆菌长亚种附属基因的NMDS分析。不同颜色的点代表不同进化支谱系, 椭圆表示按照分离菌株的民族信息圈出的90%的置信区间; B: 不同地域来源长双歧杆菌长亚种附属基因的NMDS分析。不同颜色的椭圆代表根据分离菌株的分离地信息圈出的90%的置信区间。

Figure 4 Cluster analysis of accessory genes of *Bifidobacterium longum* subsp. *longum* strains. A: Analysis of accessory genes of *B. longum* subsp. *longum* strains from different ethnic origins using NMDS method. Different colored dots represent different evolutionary lineage branches, and the ellipses indicate the 90% confidence intervals circled based on the ethnic information of isolated strains; B: Analysis of accessory genes of *B. longum* subsp. *longum* strains from different geographical origins using NMDS method. Different colored ellipses represent the 90% confidence intervals circled based on the information of strain isolation locations.

性检验结果表明, 新疆喀什市和伊宁市不同地理区域来源的维吾尔族组间 ANI 值无显著差异 ($P_{\text{adj}}=0.054$), 新疆尼勒克县哈萨克族和 2 个不同地理区域来源的维吾尔族菌株间 ANI 值均差异显著 ($P_{\text{adj}}<0.05$) (图 6)。

2.7 *B. longum* subsp. *longum* 的碳水化合物活性酶表征

136 株 *B. longum* subsp. *longum* 共注释到 55 种碳水化合物活性酶, 其中包括 32 种糖基水解酶家族 (glycoside hydrolases, GHs)、12 种糖基转移酶家族 (glycosyl transferases, GTs)、7 种碳水化合物结合模块 (carbohydrate-binding modules, CBMs)、3 种碳水化合物酯酶家族 (carbohydrate esterases, CEs) 和 1 种多糖裂解酶 (polysaccharide lyases, PLs), 体现了其丰富的碳水化合物代谢能

力 (图 7)。在所有碳水化合物活性酶家族中存在 23 个检出率为 100.00% 的碳水化合物活性酶, 包括 14 个 GHs、6 个 GTs、CBM48 及 CE9。在本研究的 *B. longum* subsp. *longum* 基因集中, GT2 家族、GH13 家族和 GH43 家族占据主导地位, 分别占全部碳水化合物酶的 13.80%、13.06% 和 8.86%。*B. longum* subsp. *longum* 丰富的碳水化合物活性酶有助于其在多种碳源环境下生长。

Kruskal-Wallis 显著性检验结果表明, 新疆 3 个不同地区分离株携带的碳水化合物活性酶在涉及 18 个 CAZyme 家族基因数量上存在显著差异 ($P_{\text{adj}}<0.05$, Benjamini-Hochberg 方法校正 P 值)。全基因组比较分析显示, 新疆喀什市维吾尔族分离株显著富集 GH13 (α -淀粉酶)、GH43

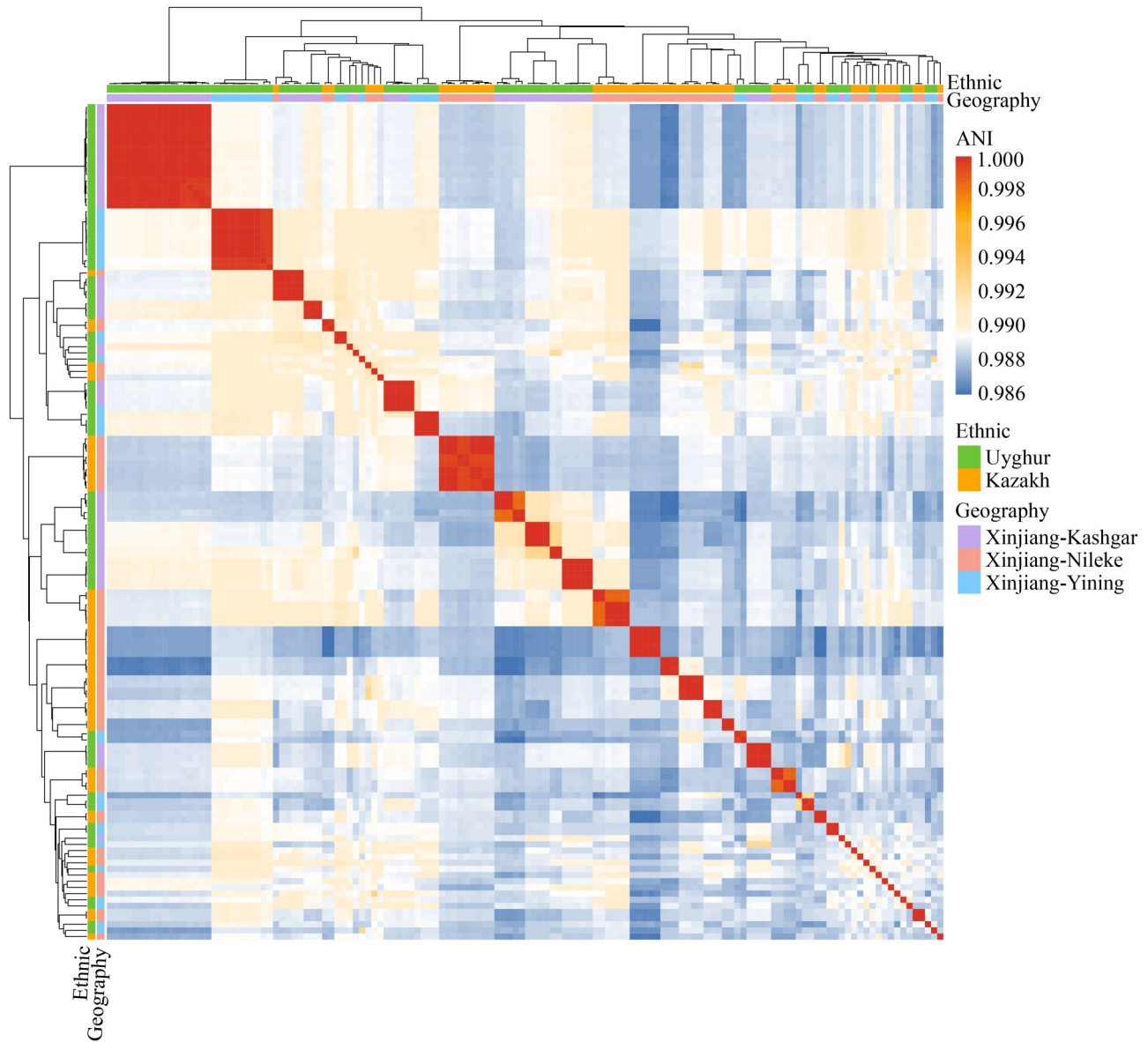


图5 *Bifidobacterium longum* subsp. *longum*的平均核苷酸一致性。热图左侧和顶部的条状注释分别表示菌株的民族和分离地信息。

Figure 5 Average nucleotide identity of *Bifidobacterium longum* subsp. *longum* strains. The horizontal and vertical bars on the left and top of the heatmap respectively represent the strain's origin and the location of isolation.

(木聚糖酶)、GH51 (L-阿拉伯呋喃糖苷酶)、GH127、GH27、GH136、GT111、GT32 和 CE8 家族, 这些酶类主要参与膳食纤维和谷物细胞壁多糖的降解; 其中, 新疆喀什市维吾尔族分离株携带 GH13 家族基因 11 个和 GH51 家族基

因 4 个, 均显著高于新疆伊宁市维吾尔族分离株。新疆伊宁市维吾尔族分离株携带 GH120 (木糖苷酶)、GH121 (β -1,2-葡聚糖酶)、GH129、GT2、CE9 和 CBM13 (β -葡聚糖结合模块)的数量较多, 其中伊宁市维吾尔族分离株 GT2 的基

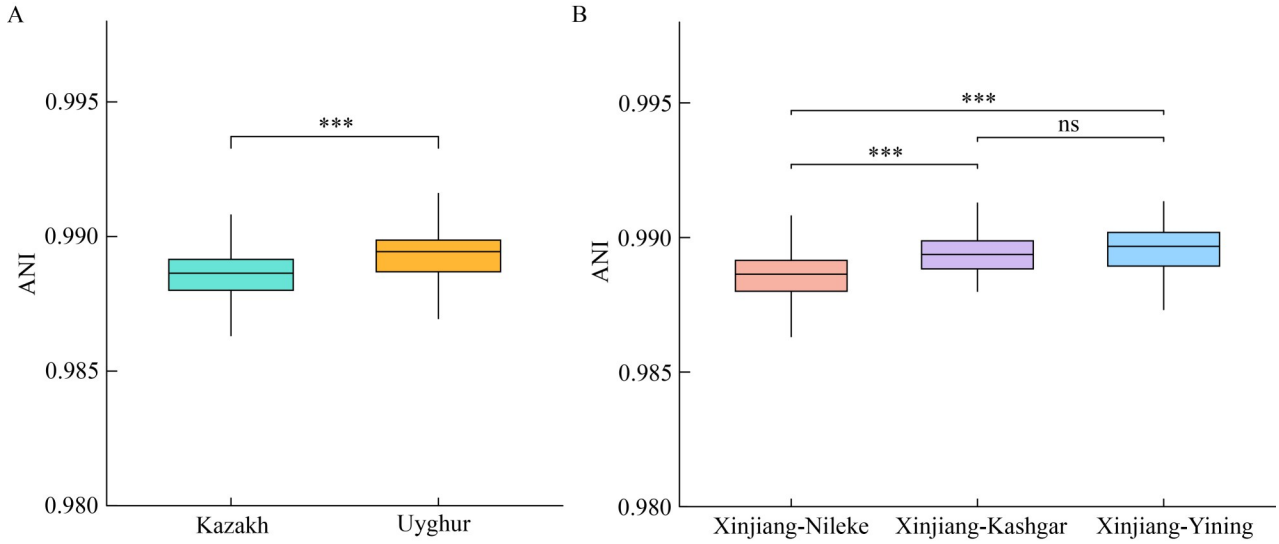


图6 不同人群*Bifidobacterium longum subsp. longum*分离株的ANI相似性比较。A: 不同民族组间ANI比较; B: 不同地理区域组间ANI相似性比较。显著性分析结果由Benjamini-Hochberg方法校正 [$P_{\text{adj}} > 0.05$ 为不显著(ns); ***: $P_{\text{adj}} < 0.001$]。

Figure 6 Comparison of ANI similarity among *Bifidobacterium longum subsp. longum* isolates from different populations. A: Comparison of ANI among different ethnic groups; B: Comparison of ANI similarity among different geographical regions. The significance analysis results were corrected using the Benjamini-Hochberg method [$P_{\text{adj}} > 0.05$ indicates non-significant (ns); ***: $P_{\text{adj}} < 0.001$].

因数量显著高于喀什市维吾尔族分离株, 分别为 12 个和 10 个。新疆尼勒克县哈萨克族分离株在参与分解乳制品来源多糖的 GH42 (β -半乳糖苷酶)、GH53 (β -1,4-半乳聚糖酶)和 CBM61 (昆布多糖结合模块)的基因数量显著高于新疆喀什维吾尔族分离株, 但与新疆伊宁县维吾尔族分离株间无显著差异。除此之外, GH27、GH51、GH129、GH136、GT2、GT32、GT111 和 CE9 家族基因数量在来自相似地理环境和饮食环境的新疆伊宁市维吾尔族和尼勒克县哈萨克族分离株间均无显著差异(图 8)。

2.8 *B. longum subsp. longum* 菌株基于 COG 数据库功能基因分析

为探究不同地理区域分离源菌株在功能基因上的差异, 基于 COG 数据库对 136 个菌株基因组进行注释。被注释到最多的 6 个功能单元为碳水化合物转运和代谢(20 384, 9.38%)、翻译

(19 247, 8.86%)、复制和修复(18 616, 8.57%)、氨基酸代谢(18 292, 8.42%)、转录(17 906, 8.24%)和细胞壁/膜/生物合成(10 747, 4.95%)(图 9A)。COG 数据库注释结果表明, 3 个不同地区分离株在参与染色质结构(B)、碳水化合物代谢(G)、辅酶代谢(H)、次级代谢产物的合成和转运(Q)、信号转导机制(T)、细胞内运输(U)和细胞骨架(Z)的基因数量间无显著差异($P > 0.05$)。新疆尼勒克县哈萨克族源分离株无机离子转运(P)和防御机制(V)相关的基因数量高于其他分离源($P < 0.05$); 新疆伊宁市维吾尔族分离株能量产生与转化(C)、氨基酸代谢(E)、脂质代谢(I)、复制和修复(L)相关的基因数量高于其他分离源; 新疆喀什维吾尔族分离株携带的 RNA 加工(A)、细胞周期调控(D)、核苷酸代谢(F)、翻译(J)、转录(K)、细胞壁/膜生物合成(M)、细胞运动(N)、翻译后修饰(O)的基因数量高于其他分离源(图 9B)。

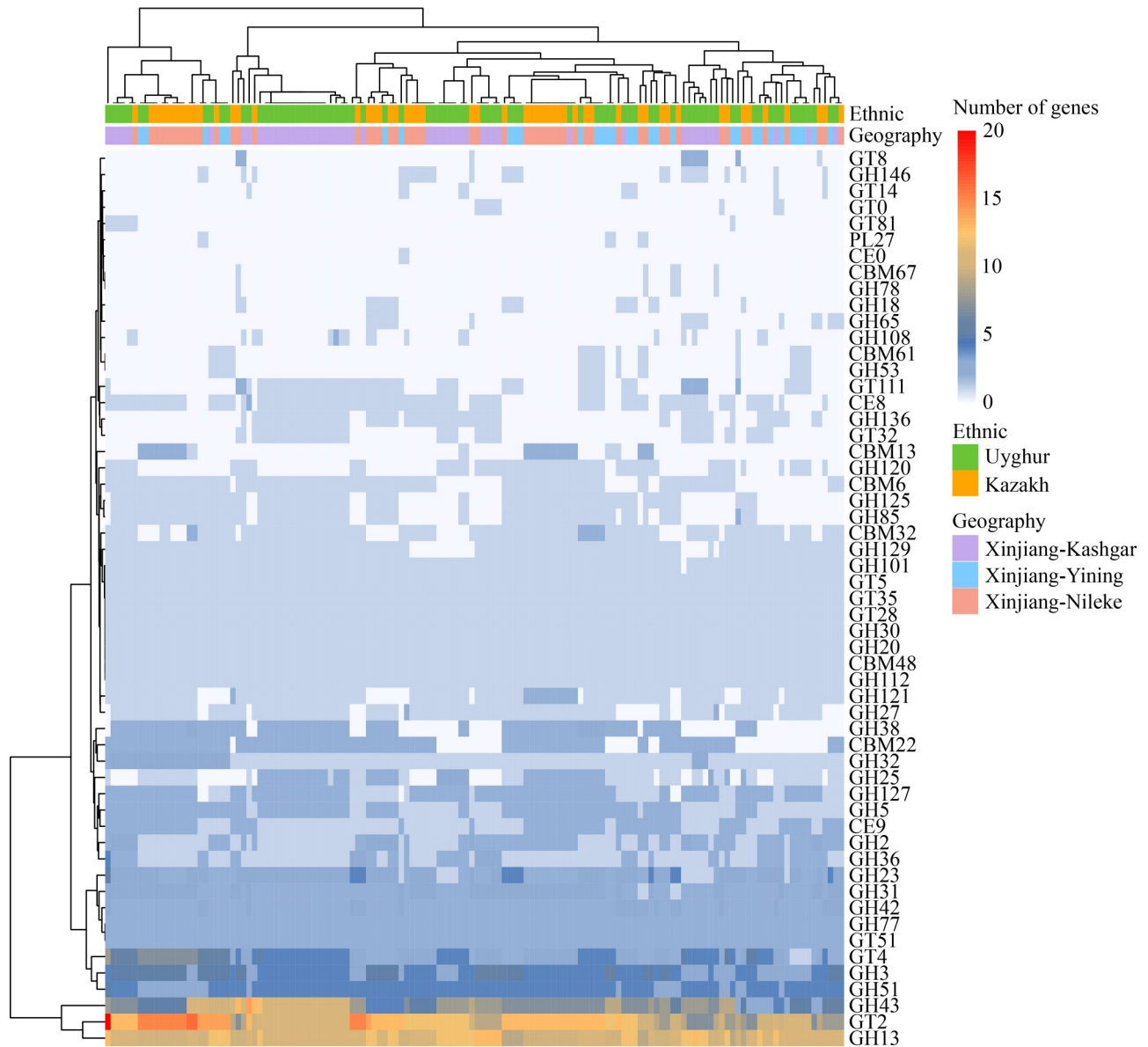


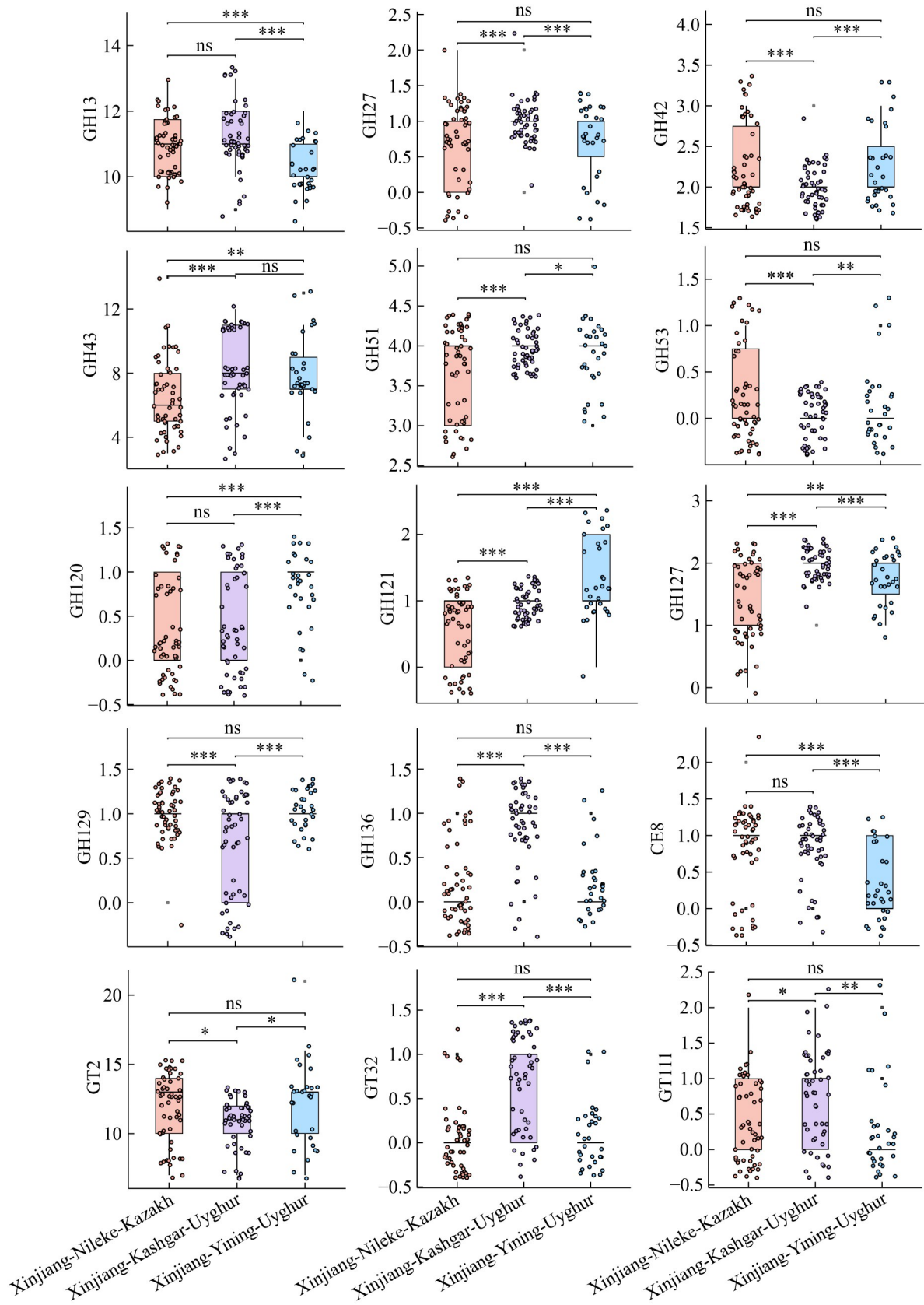
图7 不同人群*Bifidobacterium longum subsp. longum*分离株携带碳水化合物活性酶数量差异

Figure 7 The differences in the number of carbohydrate active enzymes carried by *Bifidobacterium longum subsp. longum* isolates from different populations.

3 讨论与结论

双歧杆菌(*Bifidobacterium*)作为人类肠道关键专性共栖菌, 其定殖过程起始于新生儿分娩时期, 并在宿主的整个生命周期中持续存在, 但不同种在宿主不同年龄阶段的流行率和相对丰度差异很大^[28-32]。本研究基于培养法从新

疆喀什和伊犁地区的维吾尔族和哈萨克族母婴人群中分离到隶属于双歧杆菌的9个种, 包括 *B. longum subsp. longum*、*B. adolescentis*、*B. breve*、*B. bifidum*、*B. animalis*、*B. catenulatum*、*B. pseudocatenulatum*、*B. angulatum* 和 *B. merycicum*。其中, *B. longum subsp. longum* 是分离频率最高的种, 在母亲和婴幼儿队列中的



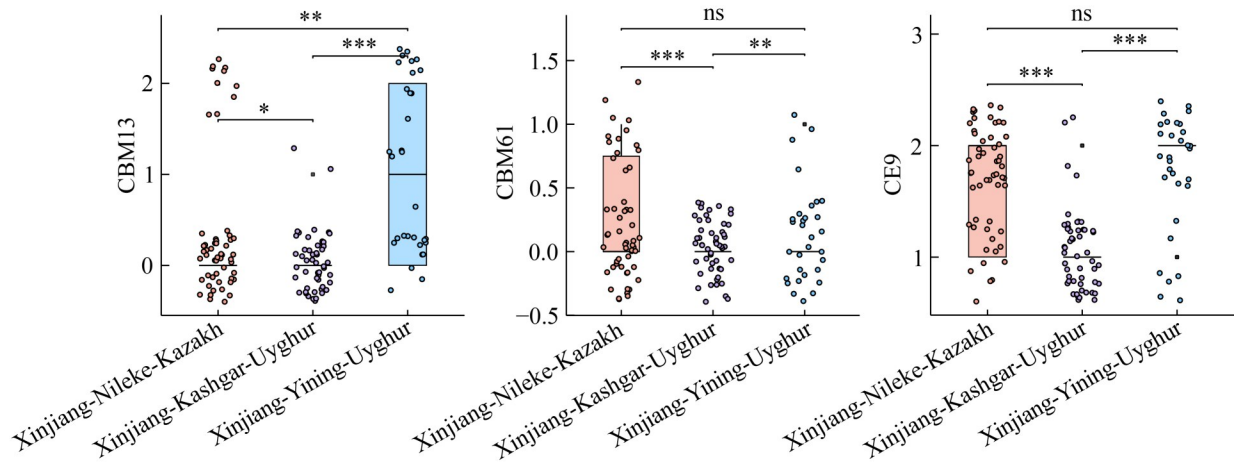
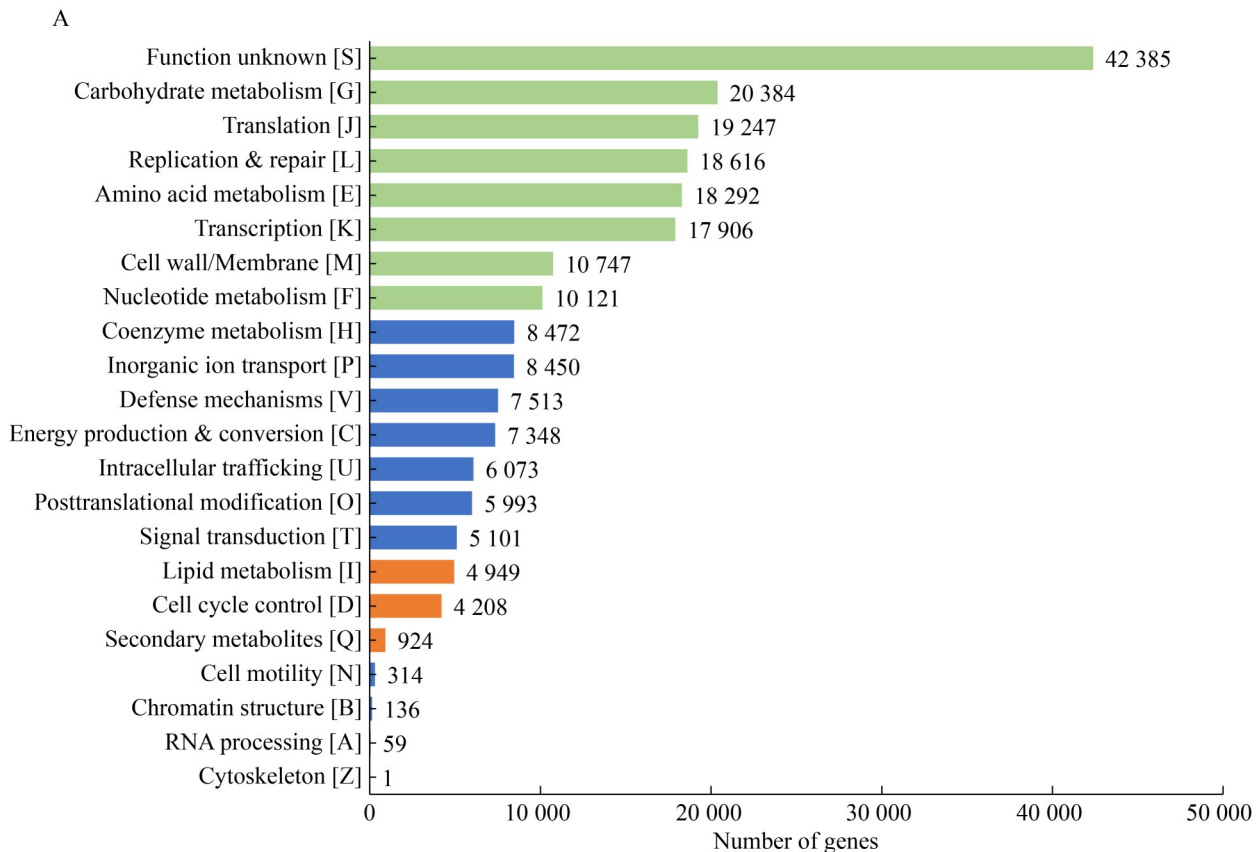


图8 不同人群*Bifidobacterium longum subsp. longum*分离株携带碳水化合物活性酶数量差异箱线图。显著性分析结果由Benjamini-Hochberg方法校正[$P_{adj}>0.05$ 为不显著(ns); *: $0.01<P_{adj}<0.05$; **: $0.001<P_{adj}<0.01$; ***: $P_{adj}<0.001$]。

Figure 8 Box plots of the number of carbohydrate active enzymes carried by *Bifidobacterium longum subsp. longum* isolates from different populations. The significance analysis results were corrected using the Benjamini-Hochberg method [$P_{adj}>0.05$ indicates non-significant (ns); *: $0.01<P_{adj}<0.05$; **: $0.001<P_{adj}<0.01$; ***: $P_{adj}<0.001$].



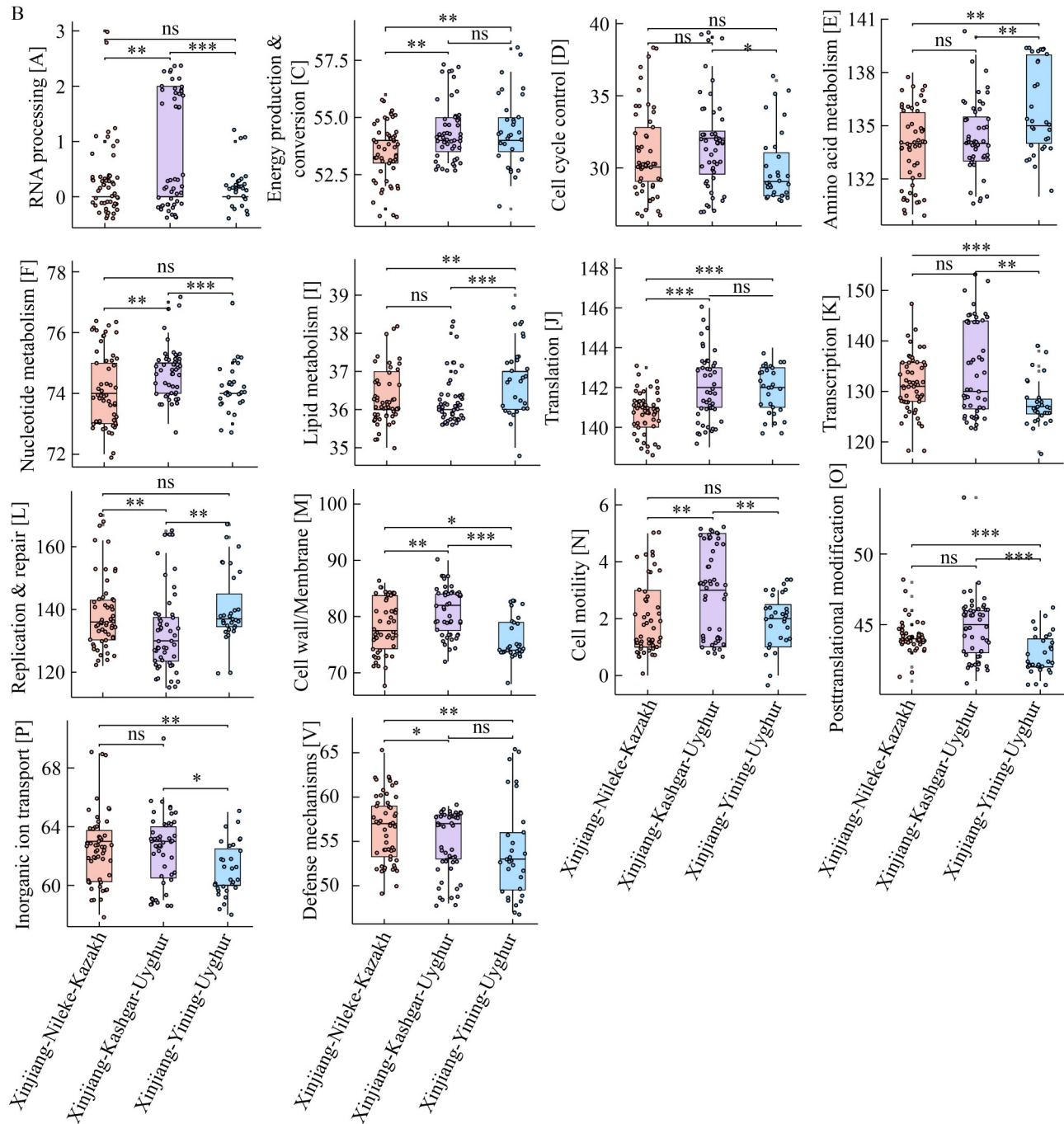


图9 不同人群*Bifidobacterium longum subsp. longum*分离株COG数据库注释结果。A: COG注释结果统计; B: 人群间差异COG箱线图。显著性分析结果由Benjamini-Hochberg方法校正[$P_{adj}>0.05$ 为不显著(ns); *: $0.01<P_{adj}<0.05$; **: $0.001<P_{adj}<0.01$; ***: $P_{adj}<0.001$]。

Figure 9 COG database annotation results of *Bifidobacterium longum subsp. longum* isolates from different populations. A: Statistical results of COG annotations; B: Box plot of inter-group differences in the population. The significance analysis results were corrected using the Benjamini-Hochberg method [$P_{adj}>0.05$ indicates non-significant (ns); *: $0.01<P_{adj}<0.05$; **: $0.001<P_{adj}<0.01$; ***: $P_{adj}<0.001$].

流行率分别为 80.65% 和 63.64%，其次是 *B. adolescentis* 和 *B. breve*。*B. adolescentis* 是成人型双歧杆菌的代表，在成人组中的分离率为 35.48%，但在 12.12% 的婴幼儿中也分离到；相反，婴儿型双歧杆菌 *B. breve* 在婴幼儿中的分离率高达 48.48%，在成人中也有 19.35% 的个体分离到。这些结果与其他一些人群的研究结果类似，例如荷兰 2 岁前婴幼儿队列和爱尔兰母婴队列的研究均报道了 *B. longum* subsp. *longum* 很高的流行率，其次是 *B. adolescentis* 和 *B. breve*^[30,33]。因此，根据人类年龄划分的人源双歧杆菌不同种的生态位尚无严格的界限，但一个普遍共识是 *B. longum* subsp. *longum* 是流行率最高、遗传多样性最复杂的物种^[34]。

研究发现，动物宿主肠道共栖细菌与宿主是否存在协同进化关系，与细菌的生态位、生理特征有关^[35-36]。例如，罗伊氏乳杆菌 (*Limosilactobacillus reuteri*) 作为典型的宿主关联型细菌，主要分布于哺乳动物的肠道，多个研究发现不同哺乳动物来源菌株群的系统发育谱系结构呈现出明显的宿主特异性^[13,37-39]；相反，*Lactiplantibacillus plantarum* 具有更广泛的生态位，在发酵蔬菜、肉乳制品、动物肠道、污水和植物表面等均有分布，其菌株的系统发育结构与来源并不严格对应^[40-41]。普遍认为，以 *Bifidobacterium*、拟杆菌 (*Bacteroides*) 和嗜黏蛋白阿克曼菌 (*Akkermansia muciniphila*) 等为代表的宿主关联型共栖细菌，其种群会逐渐形成具有地域或者宿主特异的遗传多样性格局^[41-42]。本研究基于平均核苷酸一致性 (ANI) 和核心基因构建的系统发育树结果显示，在更大的地理范围 (全中国)，以西部、中部及东部沿海的地理跨度，*B. longum* subsp. *longum* 菌株呈现明显的地理来源聚集特征。最近，有人针对横跨欧洲、非洲及亚洲更大地理尺度人群的研究显示 (跨洲尺度)，基于肠道宏基因组 MAG 数据分析显示解纤维素拟杆菌 (*Bacteroides cellulosilyticus*)、*B. longum* subsp. *longum* 和 *B. bifidum* 等专性共

栖细菌存在明显的地理聚集特征，但该研究尚无来自分离菌株的证据^[18]。在本研究中，新疆喀什和伊宁 2 个相距 800 km、来自同一民族 (维吾尔族) 不同人群的分离菌株，其系统发育呈现各自的进化支 (图 3)，进化支 A 和 B3 菌株几乎全部来自喀什的维吾尔族母婴人群，而来自伊犁维吾尔族人群的菌株绝大部分隶属于进化支 D，来自伊犁尼勒克的哈萨克族菌株隶属于进化支 B1 和 B2。特别值得注意的是，来自地理上更相近、不同人群来源的菌株在进化支内存在一定程度的重叠，例如进化支 D 中绝大部分菌株来自伊犁的哈萨克族和维吾尔族母婴人群 (除了 D1 中 3 株菌来自喀什地区)。此外，也存在喀什、伊犁 2 个不同地域人群菌株的重叠 (进化支 C)。总体上，上述结果倾向于支持一个假设，即地理上更近的人群其肠道宿主关联型细菌 (human-associated) 在遗传上更相近。这个结果可通过微生物种群遗传结构形成过程来解释，即同一地理区域的人群 (相当于局域尺度) 具有更相近的饮食习惯，饮食来源的植物性聚糖对 *B. longum* subsp. *longum* 菌株的多样性具有选择作用^[31,43]。尽管 *B. longum* subsp. *longum* 作为专性共栖细菌只在人体肠道、母乳、口腔有检测报告，但目前也不能排除另一种机制，即共栖细菌可以通过人群的亲密接触发生随机的横向传递，驱动种群内菌株的基因重组或者更替 (turnover)，在一定地理范围的人群中形成更相近的谱系 (lineage)，从而隶属于同一个进化支^[44]；反之，则存在地理上的隔离，造成共栖细菌的地理人群特化趋势，但目前并不清楚整个过程，需要更大范围、更多菌株的基因组大数据验证。由于公共数据库中来源于中国其他地区的菌株数量有限且缺乏分离源的详细信息，因此本研究未能将宿主年龄等因素纳入分析考量。然而，大量研究已证实，双歧杆菌的分布格局和功能特性与宿主年龄、饮食习惯以及健康状况之间存在着紧密且复杂的关联^[5-6]。双歧杆菌在健康母乳喂养的婴儿肠道中占主导地位，

而在成年时期其含量较低但相对稳定^[7]。不同的食物成分会为肠道内的微生物提供各异的营养物质，进而塑造出不同的肠道菌群结构^[45]。研究表明与仅食用大米的人群相比，食用小麦和小麦/米混合饮食的个体中双歧杆菌的含量显著增加^[31,43]。因此，人体肠道共栖细菌的系统发育在当地(local)还是区域(region)的精细地理尺度上呈现明显的人群地理特征需要更多的研究，包括针对不同生活史的细菌、不同生活习俗的人群^[46-47]。

双歧杆菌凭借其独特的糖分解特性，尤其是对人乳低聚糖及植物源聚糖的高效代谢能力，在婴儿肠道菌群定殖及宿主健康调控方面产生重要影响^[48]。肠道定殖是益生菌在人体内发挥有益作用的重要基础，*B. longum* subsp. *longum* 作为肠道内常见的微生物，在体内和体外试验中已被证明具有较强的肠道定殖能力^[49]。本研究中，新疆不同地理区域、不同民族人源 *B. longum* subsp. *longum* 菌株的基因组共注释到 55 种碳水化合物活性酶基因家族，其中包括与乳糖和 HMO 分解相关的 GH2、GH20、GH42、GH112 和 GH129 家族^[50-51]。差异分析的结果表明，新疆喀什市维吾尔族分离株显著富集 GH13、GH43 和 GH51 等家族，新疆伊宁市维吾尔族分离株携带 GH120、GH121 和 GH129 等的数量较多。值得注意的是，*B. longum* subsp. *longum* 作为专性共栖细菌只生存于肠道，与肠道黏膜层紧密互作，这种严格的生态位依赖性使其与宿主形成协同进化关系^[52-53]。与自由生活型、游牧生活型和严格共生式生活的乳酸菌一样，*B. longum* subsp. *longum* 菌株表现出不同的基因组特征，反映了它们在生态适应性方面的差异，是长期协同进化的结果^[54-55]。诸多研究已证实，*B. longum* subsp. *longum* 菌株具备编码 GH43、GH120 等糖苷水解酶家族成员的能力，可有效降解植物细胞壁多糖，如阿拉伯木聚糖、半纤维素等，从而促进全谷物、坚果等食物中膳食纤维的分解^[10]。此外，GH42 (β -半

乳糖苷酶)和 GH53 (半乳聚糖酶)通过特异性切割 β -1,4-半乳糖苷键，不仅能高效分解乳制品中的乳糖，还可靶向水解 HMOs 核心结构中的乳糖-N-二糖单元，从而在母婴菌群及婴儿肠道生态构建中发挥双重作用^[56]。地理环境由遗传、人文环境等多种因素共同构成，在宿主和微生物长期协同进化的过程中相同、相近人群来源的 *B. longum* subsp. *longum* 谱系与宿主形成营养共生关系，环境和饮食等选择压促使共生细菌种群逐渐形成具有地域人群特异性的聚集^[45,57-58]。研究表明代谢不同碳水化合物和抵抗宿主施加的选择压的能力，有助于 *B. longum* subsp. *longum*、脆弱拟杆菌(*Bacteroides fragilis*)等专性共栖细菌在结肠中的定殖和持续存在^[59]。因此，在 *B. longum* subsp. *longum* 益生菌株的开发中应根据不同地区的宿主人群进行菌株的筛选与功能优化以实现精准营养的开发趋势。

作者贡献声明

张慧敏：实验方法设计，数据收集与处理，实验结果可视化，论文撰写；张雪玲：研究资源采集，数据收集与处理；孙海龙：研究概念生成，研究资源采集，实验设计验证与核实；倪永清：研究概念生成，研究课题监管与指导，论文审阅与修订。

作者利益冲突公开声明

作者声明不存在任何可能会影响本文所报告工作的已知经济利益或个人关系。

参考文献

- [1] HIRANO R, SAKANAKA M, YOSHIMI K, SUGIMOTO N, EGUCHI S, YAMAUCHI Y, NARA M, MAEDA S, AMI YT, GOTOH A, KATAYAMA T, IIDA N, KATO T, OHNO H, FUKIYA S, YOKOTA A, NISHIMOTO M, KITAOKA M, NAKAI H, KURIHARA S. Next-generation prebiotic promotes selective growth of bifidobacteria, suppressing *Clostridioides difficile*[J]. *Gut Microbes*, 2021, 13(1): 1973835.
- [2] HENRICK BM, RODRIGUEZ L, LAKSHMIKANTH T, POU C, HENCKEL E, ARZOOMAND A, OLIN A, WANG J, MIKES J, TAN ZY, CHEN Y, EHRlich AM, BERNHARDSSON AK, MUGABO CH, AMBROSIANI

- Y, GUSTAFSSON A, CHEW S, BROWN HK, PRAMBS J, BOHLIN K, et al. Bifidobacteria-mediated immune system imprinting early in life[J]. *Cell*, 2021, 184(15): 3884-3898.e11.
- [3] RODRIGUEZ CI, MARTINY JBH. Evolutionary relationships among bifidobacteria and their hosts and environments[J]. *BMC Genomics*, 2020, 21(1): 26.
- [4] LUGLI GA, ALESSANDRI G, MILANI C, MANCABELLI L, RUIZ L, FONTANA F, BORRAGÁN S, GONZÁLEZ A, TURRONI F, OSSIPRANDI MC, MARGOLLES A, van SINDEREN D, VENTURA M. Evolutionary development and co-phylogeny of primate-associated bifidobacteria[J]. *Environmental Microbiology*, 2020, 22(8): 3375-3393.
- [5] TURRONI F, MILANI C, van SINDEREN D, VENTURA M. Bifidobacteria: ecology and coevolution with the host[M]//*The Bifidobacteria and Related Organisms*. Amsterdam: Elsevier, 2018: 213-220.
- [6] ARBOLEYA S, WATKINS C, STANTON C, ROSS RP. Gut bifidobacteria populations in human health and aging[J]. *Frontiers in Microbiology*, 2016, 7: 1204.
- [7] WONG CB, ODAMAKI T, XIAO JZ. Insights into the reason of human-residential bifidobacteria (HRB) being the natural inhabitants of the human gut and their potential health-promoting benefits[J]. *FEMS Microbiology Reviews*, 2020, 44(3): 369-385.
- [8] TURRONI F, MILANI C, VENTURA M, van SINDEREN D. The human gut microbiota during the initial stages of life: insights from bifidobacteria[J]. *Current Opinion in Biotechnology*, 2022, 73: 81-87.
- [9] CORDEIRO RL, SANTOS CR, DOMINGUES MN, LIMA TB, PIROLA RAS, MORAIS MAB, COLOMBARI FM, MIYAMOTO RY, PERSINOTI GF, BORGES AC, de FARIAS MA, STOFFEL F, LI C, GOZZO FC, van HEEL M, GUERIN ME, SUNDBERG EJ, WANG LX, PORTUGAL RV, GIUSEPPE PO, et al. Mechanism of high-mannose N-glycan breakdown and metabolism by *Bifidobacterium longum*[J]. *Nature Chemical Biology*, 2022, 19(2): 218-229.
- [10] KELLY SM, MUNOZ-MUNOZ J, van SINDEREN D. Plant glycan metabolism by bifidobacteria[J]. *Frontiers in Microbiology*, 2021, 12: 609418.
- [11] KIJNER S, CHER A, YASSOUR M. The infant gut commensal *Bacteroides dorei* presents a generalized transcriptional response to various human milk oligosaccharides[J]. *Frontiers in Cellular and Infection Microbiology*, 2022, 12: 854122.
- [12] NISHIYAMA K, YOKOI T, SUGIYAMA M, OSAWA R, MUKAI TK, OKADA N. Roles of the cell surface architecture of *Bacteroides* and *Bifidobacterium* in the gut colonization[J]. *Frontiers in Microbiology*, 2021, 12: 754819.
- [13] LI FY, LI XD, CHENG CC, BUJDOŠ D, TOLLENAAR S, SIMPSON DJ, TASSEVA G, PEREZ-MUÑOZ ME, FRESE S, GÄNZLE MG, WALTER J, ZHENG JS. A phylogenomic analysis of *Limosilactobacillus reuteri* reveals ancient and stable evolutionary relationships with rodents and birds and zoonotic transmission to humans[J]. *BMC Biology*, 2023, 21(1): 53.
- [14] WANG S, YANG B, ROSS RP, STANTON C, ZHAO JX, ZHANG H, CHEN W. Comparative genomics analysis of *Lactobacillus ruminis* from different niches[J]. *Genes*, 2020, 11(1): 70.
- [15] MOELLER AH, CARO-QUINTERO A, MJUNGU D, GEORGIEV AV, LONSDORF EV, MULLER MN, PUSEY AE, PEETERS M, HAHN BH, OCHMAN H. Cospeciation of gut microbiota with hominids[J]. *Science*, 2016, 353(6297): 380-382.
- [16] NISHIDA AH, OCHMAN H. A great-ape view of the gut microbiome[J]. *Nature Reviews Genetics*, 2019, 20(4): 195-206.
- [17] DÍAZ R, TORRES-MIRANDA A, ORELLANA G, GARRIDO D. Comparative genomic analysis of novel *Bifidobacterium longum* subsp. *longum* strains reveals functional divergence in the human gut microbiota[J]. *Microorganisms*, 2021, 9(9): 1906.
- [18] SUZUKI TA, FITZSTEVENS JL, SCHMIDT VT, ENAV H, HUUS KE, NGWESE MM, GRIEBHAMMER A, PFLEIDERER A, ADEGBITE BR, ZINSOU JF, ESEN M, VELAVAN TP, ADEGNIKA AA, SONG LH, SPECTOR TD, MUEHLBAUER AL, MARCHI N, KANG H, MAIER LS, BLEKHMAN R, et al. Codiversification of gut microbiota with humans[J]. *Science*, 2022, 377(6612): 1328-1332.
- [19] YANG SH, XIE XQ, MA J, HE XX, LI Y, DU MZ, LI LY, YANG LS, WU QP, CHEN W, ZHANG JM. Selective isolation of *Bifidobacterium* from human faeces using pangenomics, metagenomics, and enzymology[J]. *Frontiers in Microbiology*, 2021, 12: 649698.
- [20] LI WX, LIANG HW, HE WX, GAO XW, WU ZN, HU TY, LIN XQ, WANG MM, ZHONG YY, ZHANG HF, GE L, JIN X, XIAO L, ZOU YQ. Genomic and functional diversity of cultivated *Bifidobacterium* from human gut microbiota[J]. *Heliyon*, 2024, 10(5): e27270.
- [21] NAGPAL R, KURAKAWA T, TSUJI H, TAKAHASHI T, KAWASHIMA K, NAGATA S, NOMOTO K, YAMASHIRO Y. Evolution of gut *Bifidobacterium* population in healthy Japanese infants over the first three years of life: a quantitative assessment[J]. *Scientific Reports*, 2017, 7: 10097.
- [22] MAO S, ZHAO A, JIANG H, YAN JY, ZHONG WX, XUN YP, ZHANG YM. Patterns of human milk oligosaccharides in mature milk are associated with certain gut microbiota in infants[J]. *Nutrients*, 2024, 16(9): 1287.
- [23] XIAO MF, ZHANG C, DUAN H, NARBAD A, ZHAO JX, CHEN W, ZHAI QX, YU LL, TIAN FW. Cross-feeding of bifidobacteria promotes intestinal homeostasis: a lifelong perspective on the host health[J]. *npj Biofilms and Microbiomes*, 2024, 10: 47.
- [24] ZHANG SQ, CHEN L, HU M, ZHU JJ. 2'-Fucosyllactose (2'-FL) changes infants gut microbiota composition and their metabolism in a host-free human colonic model[J]. *Food Research International*, 2023, 173: 113293.
- [25] YAO Y, CAI XY, YE YQ, WANG FM, CHEN FY, ZHENG CH. The role of microbiota in infant health: from early life to adulthood[J]. *Frontiers in Immunology*,

- 2021, 12: 708472.
- [26] KRISHNAMURTHY HK, PEREIRA M, BOSCO J, GEORGE J, JAYARAMAN V, KRISHNA K, WANG TH, BEI K, RAJASEKARAN JJ. Gut commensals and their metabolites in health and disease[J]. *Frontiers in Microbiology*, 2023, 14: 1244293.
- [27] JAIN C, RODRIGUEZ-R LM, PHILLIPPY AM, KONSTANTINIDIS KT, ALURU S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries[J]. *Nature Communications*, 2018, 9: 5114.
- [28] ARBOLEYA S, BOTTACINI F, O'CONNELL-MOTHERWAY M, RYAN CA, ROSS RP, van SINDEREN D, STANTON C. Gene-trait matching across the *Bifidobacterium longum* pan-genome reveals considerable diversity in carbohydrate catabolism among human infant strains[J]. *BMC Genomics*, 2018, 19(1): 33.
- [29] LAWLEY B, MUNRO K, HUGHES A, HODGKINSON AJ, PROSSER CG, LOWRY D, ZHOU SJ, MAKRIDES M, GIBSON RA, LAY C, CHEW C, LEE PS, WONG KH, TANNOCK GW. Differentiation of *Bifidobacterium longum* subspecies *longum* and *infantis* by quantitative PCR using functional gene targets[J]. *PeerJ*, 2017, 5: e3375.
- [30] AN R, FONTANA F, van DAELE E, VENTURA M, VLIJGER A, van ELBURG RM, KNOL J, MILANI C, BELZER C. Longitudinal changes in bifidobacterial population during the first two years of life[J]. *Beneficial Microbes*, 2024, 15(3): 227-240.
- [31] LU J, ZHANG L, ZHANG H, CHEN YT, ZHAO JX, CHEN W, LU WW, LI MK. Population-level variation in gut bifidobacterial composition and association with geography, age, ethnicity, and staple food[J]. *npj Biofilms and Microbiomes*, 2023, 9: 98.
- [32] SOLÍS G, GUEIMONDE M. The gut microbiota in infants: focus on *Bifidobacterium*[J]. *Microorganisms*, 2023, 11(2): 537.
- [33] FEEHILY C, O' NEILL IJ, WALSH CJ, MOORE RL, KILLEEN SL, GERAGHTY AA, LAWTON EM, BYRNE D, SANCHEZ-GALLARDO R, NORI SRC, NIELSEN IB, WORTMANN E, MATTHEWS E, O'FLAHERTY R, RUDD PM, GROEGER D, SHANAHAN F, SALDOVA R, McAULIFFE FM, van SINDEREN D, et al. Detailed mapping of *Bifidobacterium* strain transmission from mother to infant via a dual culture-based and metagenomic approach[J]. *Nature Communications*, 2023, 14: 3015.
- [34] DERRIEN M, TURRONI F, VENTURA M, van SINDEREN D. Insights into endogenous *Bifidobacterium* species in the human gut microbiota during adulthood[J]. *Trends in Microbiology*, 2022, 30(10): 940-947.
- [35] GARUD NR, POLLARD KS. Population genetics in the human microbiome[J]. *Trends in Genetics*, 2020, 36(1): 53-67.
- [36] SHOEMAKER WR, CHEN D, GARUD NR. Comparative population genetics in the human gut microbiome[J]. *Genome Biology and Evolution*, 2022, 14(1): evab116.
- [37] PARK S, STEINEGGER M, CHO HS, CHUN J. Metagenomic association analysis of gut symbiont *Limosilactobacillus reuteri* without host-specific genome isolation[J]. *Frontiers in Microbiology*, 2020, 11: 585622.
- [38] FRESE SA, BENSON AK, TANNOCK GW, LOACH DM, KIM J, ZHANG M, OH PL, HENG NCK, PATIL PB, JUGE N, MACKENZIE DA, PEARSON BM, LAPIDUS A, DALIN E, TICE H, GOLTSMAN E, LAND M, HAUSER L, IVANOVA N, KYRPIDES NC, et al. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*[J]. *PLoS Genetics*, 2011, 7(2): e1001314.
- [39] OH PL, BENSON AK, PETERSON DA, PATIL PB, MORIYAMA EN, ROOS S, WALTER J. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution[J]. *The ISME Journal*, 2009, 4(3): 377-387.
- [40] MAO BY, YIN RM, LI XS, CUI SM, ZHANG H, ZHAO JX, CHEN W. Comparative genomic analysis of *Lactiplantibacillus plantarum* isolated from different niches[J]. *Genes*, 2021, 12(2): 241.
- [41] DUAR RM, LIN XB, ZHENG JS, MARTINO ME, GRENIER T, PÉREZ-MUÑOZ ME, LEULIER F, GÄNZLE M, WALTER J. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*[J]. *FEMS Microbiology Reviews*, 2017, 41(Supp_1): S27-S48.
- [42] MARTINO ME, BAYJANOV JR, CAFFREY BE, WELS M, JONCOUR P, HUGHES S, GILLET B, KLEEREBEZEM M, van HIJUM SAFT, LEULIER F. Nomadic lifestyle of *Lactiplantibacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats[J]. *Environmental Microbiology*, 2016, 18(12): 4974-4989.
- [43] LU J, ZHANG L, ZHAI QX, ZHAO JX, ZHANG H, LEE YK, LU WW, LI MK, CHEN W. Chinese gut microbiota and its associations with staple food type, ethnicity, and urbanization[J]. *npj Biofilms and Microbiomes*, 2021, 7: 71.
- [44] ZHANG JC, KNIGHT R. Genomic mutations within the host microbiome: adaptive evolution or purifying selection[J]. *Engineering*, 2023, 20: 96-102.
- [45] YOUNGBLUT ND, REISCHER GH, WALTERS W, SCHUSTER N, WALZER C, STALDER G, LEY RE, FARNLEITNER AH. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades[J]. *Nature Communications*, 2019, 10: 2200.
- [46] LINDSTRÖM ES, LANGENHEDER S. Local and regional factors influencing bacterial community assembly[J]. *Environmental Microbiology Reports*, 2012, 4(1): 1-9.
- [47] FEESER KL, van HORN DJ, BUELOW HN, COLMAN DR, McHUGH TA, OKIE JG, SCHWARTZ E, TAKACS-VESBACH CD. Local and regional scale heterogeneity drive bacterial community diversity and composition in a polar desert[J]. *Frontiers in Microbiology*, 2018, 9: 1928.
- [48] MILANI C, TURRONI F, DURANTI S, LUGLI GA, MANCABELLI L, FERRARIO C, van SINDEREN D, VENTURA M. Genomics of the genus *Bifidobacterium*

- reveals species-specific adaptation to the glycan-rich gut environment[J]. *Applied and Environmental Microbiology*, 2015, 82(4): 980-991.
- [49] ZHANG CC, YU ZM, ZHAO JX, ZHANG H, ZHAI QX, CHEN W. Colonization and probiotic function of *Bifidobacterium longum*[J]. *Journal of Functional Foods*, 2019, 53: 157-165.
- [50] SÁNCHEZ C, FENTE C, REGAL P, LAMAS A, LORENZO MP. Human milk oligosaccharides (HMOs) and infant microbiota: a scoping review[J]. *Foods*, 2021, 10(6): 1429.
- [51] MILANI C, LUGLI GA, DURANTI S, TURRONI F, MANCABELLI L, FERRARIO C, MANGIFESTA M, HEVIA A, VIAPPANI A, SCHOLZ M, ARIOLI S, SANCHEZ B, LANE J, WARD DV, HICKEY R, MORA D, SEGATA N, MARGOLLES A, van SINDEREN D, VENTURA M. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut[J]. *Scientific Reports*, 2015, 5: 15782.
- [52] ZHANG M, HANG XM, TAN J, YANG H. The host genotype and environment affect strain types of *Bifidobacterium longum* subsp. *longum* inhabiting the intestinal tracts of twins[J]. *Applied and Environmental Microbiology*, 2015, 81(14): 4774-4781.
- [53] GEERLINGS SY, OUWERKERK JP, KOEHORST JJ, RITARI J, AALVINK S, STECHER B, SCHAAP PJ, PAULIN L, de VOS WM, BELZER C. Genomic convergence between *Akkermansia muciniphila* in different mammalian hosts[J]. *BMC Microbiology*, 2021, 21(1): 298.
- [54] XIAO Y, ZHAO JX, ZHANG H, ZHAI QX, CHEN W. Mining genome traits that determine the different gut colonization potential of *Lactobacillus* and *Bifidobacterium* species[J]. *Microbial Genomics*, 2021, 7(6): 000581.
- [55] XIAO Y, ZHAI QX, ZHANG H, CHEN W, HILL C. Gut colonization mechanisms of *Lactobacillus* and *Bifidobacterium*: an argument for personalized designs[J]. *Annual Review of Food Science and Technology*, 2021, 12: 213-233.
- [56] LUO YH, XIAO Y, ZHAO JX, ZHANG H, CHEN W, ZHAI QX. The role of mucin and oligosaccharides via cross-feeding activities by *Bifidobacterium*: a review[J]. *International Journal of Biological Macromolecules*, 2021, 167: 1329-1337.
- [57] ROTHSCHILD D, WEISSBROD O, BARKAN E, KURILSHIKOV A, KOREM T, ZEEVI D, COSTEA PI, GODNEVA A, KALKA IN, BAR N, SHILO S, LADOR D, VILA AV, ZMORA N, PEVSNER-FISCHER M, ISRAELI D, KOSOWER N, MALKA G, WOLF BC, AVNIT-SAGI T, et al. Environment dominates over host genetics in shaping human gut microbiota[J]. *Nature*, 2018, 555(7695): 210-215.
- [58] AIHETANMU S, LIANG ZX, ZHANG XL, LUO BL, ZHANG HM, HUANG J, TIAN FW, SUN HL, NI YQ. Genetic specialization of key bifidobacterial phylotypes in multiple mother-infant dyad cohorts from geographically isolated populations[J]. *Frontiers in Microbiology*, 2024, 15: 1399743.
- [59] ALMEIDA A, NAYFACH S, BOLAND M, STROZZI F, BERACOCHEA M, SHI ZJ, POLLARD KS, SAKHAROVA E, PARKS DH, HUGENHOLTZ P, SEGATA N, KYRPIDES NC, FINN RD. A unified catalog of 204, 938 reference genomes from the human gut microbiome[J]. *Nature Biotechnology*, 2021, 39(1): 105-114.