BMC Microbiology

RESEARCH Open Access

Genomic characteristics and phylogenetic relationships of *Cutibacterium acnes* breast milk isolates

Jiaqi Sun[1,2,3], Guoxuan Hang[1,2,3], Huimin Lv[1,2,3], Yu Li[1,2,3], Qiujie Song[1,2,3], Zhi Zhong[1,2,4], Zhihong Sun[1,2,4] and Wenjun Liu[1,2,4,5*]

Abstract

Background *Cutibacterium acnes* is one of the most commonly found microbes in breast milk. However, little is known about the genomic characteristics of *C. acnes* isolated from breast milk. In this study, the sequencing and assembly results of 10 *C. acnes* isolates from breast milk were compared with the genomic data of 454 strains downloaded from NCBI, and the characteristics of breast milk isolates from various perspectives, including phylogeny, genomic characteristics, virulence genes, drug resistance genes, and carbohydrate utilization, were elucidated.

Results The findings of this study revealed no differences between the breast milk isolates and other isolates in terms of genomic features, phylogenetic relationships, virulence, and resistance-related genes. However, breast milk-derived isolates exhibited significantly lower copies of the carbohydrate metabolic enzyme genes GT5 and GT51 ($P<0.05$) and a higher copy number of the GH31 gene ($P<0.05$) than others. *C. acnes* primarily consists of three genetic branches (A, B, and C), which correspond to the three subspecies of *C. acnes* (*C. acnes* subsp. *elongatum*, *C. acnes* subsp. *defendens*, *C. acnes* subsp. *acnes*). The genetic differences between branches B and C were smaller than that between branch A. Branches A and B carry a higher number of copies of carbohydrate enzymes, including CE1, CE10, GH3, and CBM32 than branch C. Branches B and C possess the carbohydrate enzymes PL8 and GH23, which are absent in branch A. Core genes, core intergenic regions, and concatenated sequences of core genes and core intergenic regions were compared to construct a phylogenetic tree, and it was found that core intergenic regions could be used to describe phylogenetic relationships.

Conclusions It is therefore speculated that the *C. acnes* in breast milk originates from the nipple or breast surface. This study provides a novel genetic basis for genetic differentiation of *C. acnes* isolates from breast milk.

Highlights

- Breast milk isolates had lower copies of GT5 and GT51, and higher GH31.
- *C. acnes* consists of three genetic branches corresponding to the three subspecies.
- Branches A and B had more copies of CE1, CE10, GH3, and CBM32 than C.
- Core intergenic regions can be used to describe the phylogenetic relationships of *C. acnes*.

*Correspondence:
Wenjun Liu
wjliu168@163.com

Full list of author information is available at the end of the article

Keywords *Cutibacterium acnes*, Genomics, Breast milk, Carbohydrate enzymes

Introduction

Cutibacterium acnes (*C. acnes*) is a common gram-positive facultative anaerobic bacillus found on human skin, belonging to the phylum Actinobacteria [1]. It plays a crucial role in maintaining skin homeostasis and preventing the colonization of harmful pathogens [2]. However, it is also an opportunistic causative agent of common acne [1, 3, 4]. The genomics of *C. acnes* have been a subject of increasing interest because of recent advancements in high-throughput sequencing and bioinformatics. Scholz et al. revised its taxonomy through high-resolution phylogenetic analysis and reclassified *Propionibacterium acnes*, previously known as a skin species, into the genus *Cutibacterium* [5]. *C. acnes* is usually categorized into three types based on gene sequences or biological features [6]. Studies have shown that different types *C. acnes* have different genomic features and pathogenic tendencies [7]. Further genomic analyses have provided valuable insights into the genetic differentiation of *C. acnes*, contributing to a better understanding of this microorganism. Urbaniak et al. [8] summarized the identification of several bacterial types in breast milk, including *C. acnes*, using both culture-dependent and culture-independent techniques. The bacteria in breast milk are believed to originate from the mother's gastrointestinal tract, with bacterial exposure to the breast during breastfeeding, as well as retrograde flow of milk from the infant's mouth into the mother's milk ducts [9]. Notably, *C. acnes* is found in breast milk [10], although little is known about breast milk-derived isolates.

Advancements in genomics have revealed the genetic diversity and differentiation of *C. acnes* [11]. Oh et al. reported the diversity of *C. acnes* on human skin in 2014 [12] and a subsequent study found that *C. acnes* is one of the dominant cutaneous commensals, exhibiting relatively low levels of variability compared to *Staphylococcus epidermidis* [13]. Through comparative genomics analysis, Cobian et al. found that adaptive evolutionary outcomes in *C. acnes* manifested in different genetic branches with distinct CRISPR-Cas systems of types I-E [14]. Furthermore, related studies have identified a possible association between different genetic branches of *C. acnes* and pathogenic phenotypes [15]. Zhao et al. [15] identified the association of *C. acnes* type II strains with acne in Chinese patients using MLST. Studies have shown that *C. acnes* phylogenetic clade I-1 is significantly associated with acne, while clade I-2, II, and III strains are associated with health carriers and some opportunistic infections [16]. Genomic analysis facilitates further understanding of the pathogenicity of *C. acnes* [10], a microorganism commonly found in breast milk. However, knowledge of the genomic characterization of breast milk-derived isolates is limited.

We conducted a comparative genomic analysis of 464 strains of *C. acnes* to gain a deeper understanding of the genetic background and functional genomic characteristics of *C. acnes* isolated from breast milk. This study elucidated the characteristics of breast milk-derived isolates in terms of their genetic background, virulence genes, drug resistance genes, and carbohydrate utilization genes. Additionally, we compared the population genetic background and differences in functional genes among the different subspecies and genetic branches of *C. acnes*. The findings of this study provide a genetic basis for breast milk-derived *C. acnes* isolates.

Methods

Strain information

The strains used in this study were isolated from breast milk by our research team [17]. Ten *C. acnes* isolates were obtained from breast milk samples of healthy puerperia women in Hohhot, Inner Mongolia Autonomous Region, China, and the specific strain information is presented in Table S1.

Furthermore, the genome data of 454 *C. acnes* strains obtained from the NCBI Refseq database as of August 6, 2023, were included in this study. These strains were selected based on their threshold of >95% ANI with the type strains and by excluding strains with apparent assembly issues, as listed in Table S2.

Strain activation

In this study, the isolates retrieved from the ampoule and preserved via vacuum freeze-drying were inoculated into 5 mL fresh de Mann, Rogosa, and Sharpe (MRS) agar plates (Oxoid, Thermo Fisher Scientific, Inc., Basingstoke, UK) for activation. The plates were incubated at a stable temperature of 37 °C for 24±2 h. The colonies were streaked onto de MRS agar plates for observation. The morphology and characteristics of colonies were observed after 24±2 h of incubation. Single colonies were selected and inoculated in MRS broth for further cultivation. The culture results were examined by microscopic analysis and strains with consistent morphological characteristics were selected for subsequent experiments.

DNA extraction and sequencing

The *C. acnes* genome was extracted using the CTAB method [18, 19]. The concentration and purity of the DNA of the strain were assessed using a NanoDrop (ND-1000, Thermo Fisher Scientific Inc., Wilmington, DE, United States) micro-UV spectrophotometer and 0.8%

agarose gel electrophoresis. The genomic DNA is randomly fragmented into pieces of approximately 350 bp in length. After the end modifications are completed, a single adenine ("A") is added to the 3' ends of the double-stranded DNA. Illumina sequencing adapters are then ligated to both ends of the library DNA. After library fragment selection, the DNA library is amplified by PCR. In the constructed small DNA fragment library, each inserted fragment is subjected to paired-end sequencing, with each end sequenced for 150 bp. An Illumina NovaSeq 6000 platform (Illumina, Inc., San Diego, CA, United States), a high-throughput sequencing platform, was used to accomplish second-generation sequencing. The sequencing data obtained from this high-quality process yielded an average coverage of approximately 500×.

Genome assembly

The data were subsequently filtered using readfq.v5.0 software to retain approximately 100× high-quality data [20]. The genomes were assembled using SOAPdenove2 software (v2.0) [21] and the assembly results were corrected for single bases based on filtered high-quality reads using SOAP software (v2.2) [22]. GapCloser software (http://sourceforge.net/proiects/soapdenovo2/files/GapCloser/) was used to fill the vacancies. The completeness and contamination of all strain assembly results were assessed using the lineage_wf parameter of CheckM software (v1.1) [23], and genomes with less than 95% completeness or more than 5% contamination were eliminated. Using self-made script, the basic information such as genome size, GC content, N50, N90 and so on were analyzed.

Calculation of average nucleotide identity (ANI), dissimilarity (genes and intergenic regions)

In this study, fastANI software [24] was used to calculate the average nucleotide identity (ANI) of 464 *C. acnes* samples, with the comparison fragment length set to 1000.

Using the methodologies described by Wielgoss et al. [25] and Luo et al. [26], Jaccard distance was applied to identify the presence or absence of genes within accessory genomes across different strains as well as to assess the dissimilarity within intergenic regions and between intergenic regions between genes. Subsequently, cluster analysis was performed using the complete linkage method on the matrix to generate a heat map.

Pan -core gene and intergenic region sets

Strains were annotated using Prokka software [27] and the resulting pan-core gene set was analyzed using Roary software [28]. The composition of the intergenic regions of the *C. acnes* population was analyzed using piggy software (v1.5) [29] to construct the *C. acnes* pan-core intergenic region set using the default parameters of the software. All strains were analyzed separately by software with the same parameters.

Tandem sequences based on core genes and inter-core gene regions, phylogenetic tree construction

The core gene sequences and core intergenic sequences were combined using the "ape" package of the R software (https://github.com/gjuggler/ape). To construct the phyl ogenetic tree, TreeBeST software [30] was used, based on the neighbor-joining (NJ) method for core genes, inter-core gene regions, and tandem sequences of core genes and inter-core gene regions. The Bootstrap value for this process was set to 1000. The cophenetic distance of the phylogenetic tree was calculated using the "ape" package in R, and the similarity between distance matrices was assessed using the Mantel test on the distance matrices with the "vegan" package.

Construction of SNP phylogenetic tree and principal component analysis

In this study, the complete genome sequence of *C. acnes* ATCC 6919^{T} (GCF_008728435.1) was used as the reference genome, as recommended in previous studies [31]. The core SNPs of the *C. acnes* population were computed using the snippy-multi component of Snippy software (v4.6) [32]. Phylogenetic trees were constructed using Rxmal software (v8.2) [33] by employing the maximum likelihood (ML) approach on SNPs after removal of recombination.

Principal component analysis (PCA) was conducted on the Core-SNPs of the *C. acnes* population using the-pca parameter in plink software (v1.9) [34].

Resistance gene, virulence gene annotation

The genome sequences of 464 *C. acnes* strains were functionally annotated using two databases, the Comprehensive Antibiotic Resistance Database (CARD) [35] and the Virulence Factors Database (VFDB) [36] (sequence similarity > 90%, sequence coverage > 60%) for resistance genes and virulence genes, respectively. Annotation was performed using a previously research method [37].

Carbohydrate metabolizing enzymes annotation

Based on a previous research methodology [38], functional gene annotation was performed through a BLAST search against the CAZy databases [39] using the hmmscan tool [40]. The parameters set for the search were E-value < $1 \times e^{-10}$, identity > 70%, and coverage percentage > 70%.

Statistical analysis and data visualization

Statistical analysis of the data and data visualization were performed using R software (v4.2). The phylogenetic tree

results were visualized using the iTol website [41]. Data visualization was performed using the ggplot2 package and box plots were plotted in R using the ggpubr package. Heat maps were plotted using the pheatmap package. Ternary phase diagrams were plotted using the ggtern package. The Wilcoxon rank-sum test was applied using R software to calculate *P* values, and a *P* value < 0.05 was considered statistically significant.

Data Availability

Data is deposited in the National Center for Biotechnology Information (NCBI) with accession numbers PRJNA1160923.

Results

Genomic information for *C. Acnes*

The genomes of 10 *C. acnes* isolates which obtained from breast milk were sequenced and analyzed, revealing a genome size of 2.39±0.03 Mb, a GC content of 60.06% ± 0.02%, and a coding sequence (CDS) number of 2,327.20±20.29 (Fig. 1, Table S3). The results of these analyses were compared with the basic information of the NCBI genome sequences, and it was determined that there was no statistically significant difference between the above results ($P>0.05$, Fig. 1). Furthermore, the scaffold number of these 10 *C. acnes* strains was 25.60±10.59, and they had a high level of integrity (>97%) and low levels of contamination (<3%). These findings suggest that the genomes of these 10 *C. acnes* isolates were of high quality and could be used for subsequent bioinformatic analysis.

Based on the genomic information analysis from 464 *C. acnes* strains, the findings indicate that the genome size is 2.39±0.03 Mb, with a GC content of 60.06% ± 0.07% and CDS of 2329.31±32.15. These results demonstrated that *C. acnes* have a high GC content. The results indicate that *C. acnes* has a high GC content, consistent with the findings of Cobian N et al. [14].

Pan-core gene set and pan-core intergenic region set

Construction of the pan-core gene set and pan-core intergenic region set for the 464 *C. acnes* strains was performed. The pan-gene set for the 464 *C. acnes* samples comprised 10,072 genes, including 193 core genes, 3,130 shell genes, and 6,249 cloud genes. Notably, the percentage of cloud genes in the pan-gene set was 62.04%, whereas the percentage of core genes was only 1.92%. These findings suggested that *C. acnes* exhibit genetic heterogeneity. The pan-gene curve set demonstrated a rapid increase in the number of genomes (Fig. S1A), whereas the core gene curve leveled off with an increasing number of genes (Fig. S1B).

The pan-intergenic region set of the 464 *C. acnes* strains consisted of 6,616 intergenic regions, of which 174 were core intergenic regions. Consistent with the results obtained from the pan-core gene set, the core intergenic region of *C. acnes* represented a relatively small proportion of the pan intergenic region (174/6,616), further indicating that *C. acnes* exhibit genetic heterogeneity.

Phylogenetic analysis

In this study, core genes, intergenic regions, and concatenated sequences of core genes and core intergenic regions were used to construct a phylogenetic tree. Phylogenetic analysis revealed a consistent phylogenetic structure for the 464 *C. acnes* strains across the core gene phylogenetic tree, core intergenic region phylogenetic tree, concatenated core gene and core intergenic region phylogenetic tree (Fig. 2A-C). The 464 *C. acnes* strains were grouped into three main genetic branches (A, B, and C), corresponding to the three subspecies of *C. acnes* (*C. acnes* subsp. *elongatum*, *C. acnes* subsp. *defendens*, and *C. acnes* subsp. *acnes*). Branch C was further divided into three lineages (CI, CII, and CIII). Additionally, there were significant genetic distance between different genetic branches, whereas genetic distance within the same branch were relatively minor.

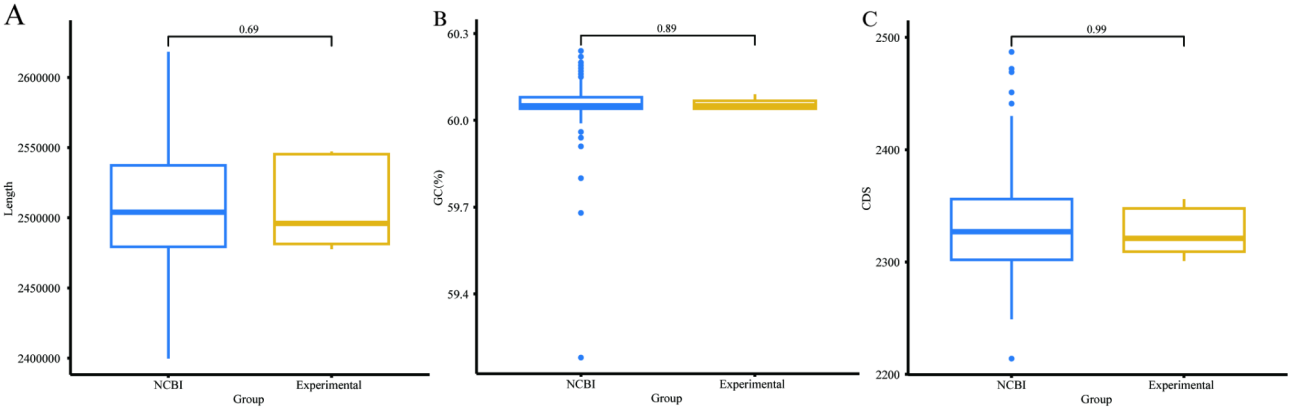

Fig. 1 Box plots were used to compare the basic information of *Cutibacterium acnes* strains sourced from breast milk and non-breast milk, and significance was calculated for (**A**) genome size, (**B**) GC content (**C**) CDS

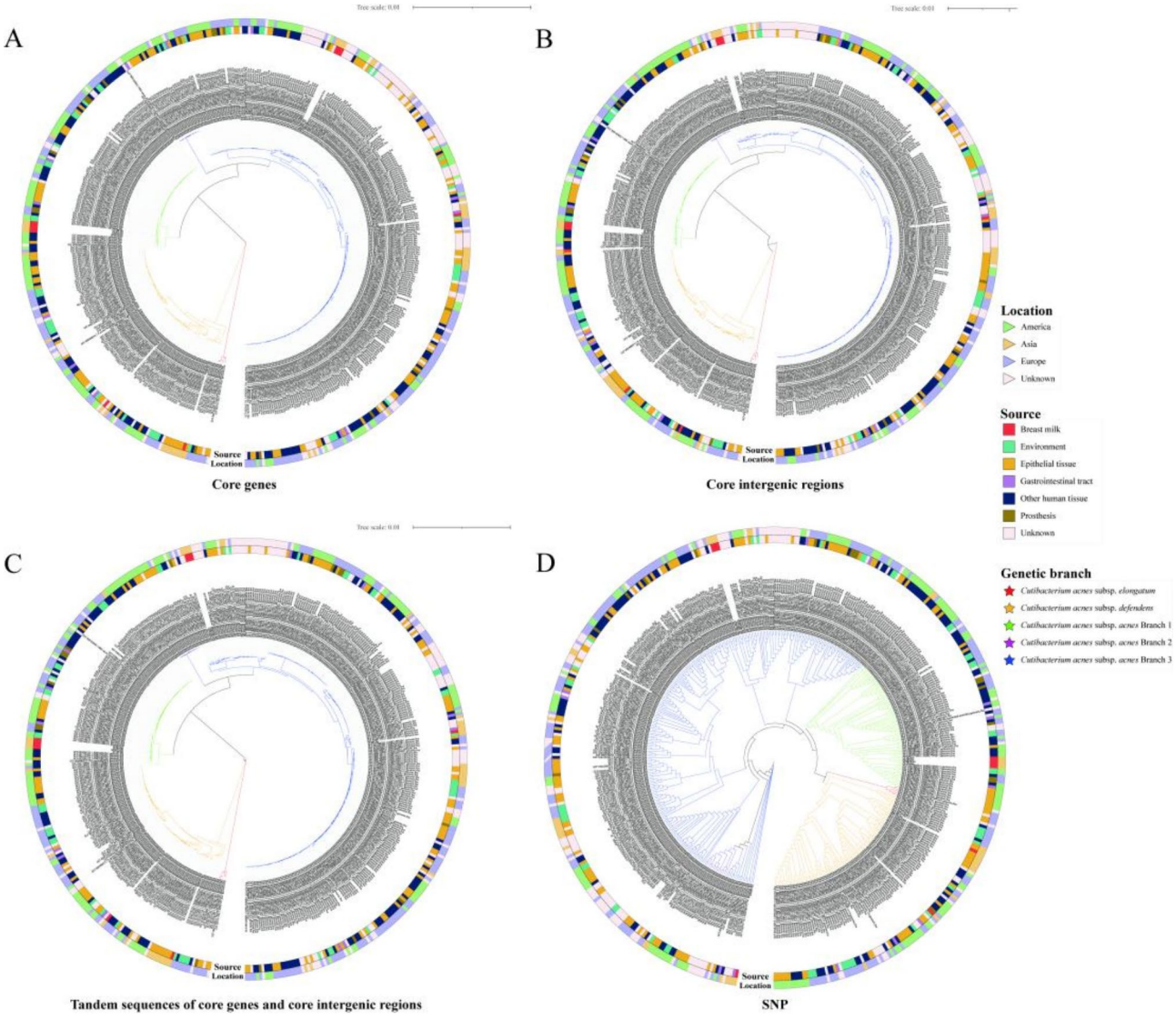

Fig. 2 A phylogenetic tree of *Cutibacterium acnes* was constructed using 464 strains (with average nucleotide identity values greater than 0.95% compared with the reference strain ATCC6919^{T}). The outer circle represents the isolation location characteristics and the inner circle represents the isolation source characteristics. (**A**) Core genes, (**B**) core intergenic regions, (**C**) concatenated sequences of core genes and core intergenic regions, and (**D**) single nucleotide polymorphism (SNP)

Compared to phylogenetic trees based solely on core genes or core intergenic regions, the concatenated core genes and core intergenic regions provided a clearer genetic distance among strains to construct a phylogenetic tree (Fig. 2A-C). Therefore, this concatenated system reflects the genetic relationships between different subspecies more accurately.

Combining the information on strain isolation locations and sources, no distinct phylogenetic clustering of *C. acnes* was observed based on isolation sources or locations in the phylogenetic trees. Two strains of breast milk isolates, YL12-10 and SRN1R1, were found in branch B, whereas the remaining eight breast milk-derived isolates were distributed in branch C, with four strains each in lineages CI and CIII. The ratio of the 10 breast milk-derived isolates on branches B and C was 1:4, which was similar to the ratio of all *C. acnes* strains on branches B and C (97:362). The cophenetic distance were calculated through the core gene phylogenetic tree (Table S5), the core gene region phylogenetic tree (Table S6), and concatenated sequences of core genes and core intergenic regions phylogenetic tree (Table S7). The Mantel test value between the core gene phylogenetic tree and the core intergenic region phylogenetic tree was 0.9575, and the Mantel test value between the core gene phylogenetic tree and concatenated sequences of core genes and core intergenic regions phylogenetic tree was 0.9572. The Mantel test value between the core intergenic region

phylogenetic tree and concatenated sequences of core genes and core intergenic regions phylogenetic tree was 0.96. Therefore, the phylogenetic tree constructed using the core intergenic region is more similar to the phylogenetic tree constructed using concatenated sequences of core genes and core intergenic regions.

SNP calling

SNP calling revealed 123,522 SNP sites, among which 63,695 were synonymous mutations, 47,418 were non-synonymous mutations, and 12,581 were located in intergenic regions. The transition to transversion ratio (Ts/Tv) was calculated to be 5.0039.

A phylogenetic tree was constructed based on core gene SNP analysis, and the results demonstrated a structure consistent with the phylogenetic trees constructed using core genes, concatenated sequences of core genes, and core intergenic regions (Fig. 2D). Further analysis revealed that the *C. acnes* population is primarily composed of three major genetic branches, providing additional support for their differentiation. Compared to branch A, the genetic differences between branches B and C were relatively small. Principal component analysis based on SNPs further confirmed this hypothesis. In PC2, branch A was distant from branches B and C, indicating significant genetic distinction ($P<0.05$) (Fig. S2A). Within branch C, the distance between CIII and CII on PC1 was smaller than that of CI, emphasizing the differences among these three branches (Fig. S2B).

Result of ANI and Dissimilarity (genes and intergenic regions)

All strains exhibited ANI values greater than 96.5% when compared to the type strain ATCC6919^{T} (Fig. 3A), indicating that all strains belonged to *C. acnes*. Analysis of ANI among 464 *C. acnes* strains revealed three major branches (A, B, and C), with branch C further subdivided into three lineages: CI, CII, and CIII. This phylogenetic relationship was consistent with the results of ANI. The ANI values within the branches were significantly higher than those between branches ($P<0.05$, Fig. S3). Furthermore, the ANI value between branches B and C was significantly higher than that between branches A and B as well as between branches A and C ($P<0.05$, Fig. S3). This suggests that the differences between branches B and C were smaller than those between branch A, which aligns with the phylogenetic relationship. In branch C, the ANI between lineages CI and CII was significantly lower than that between CI and CIII ($P<0.05$, Fig. S4), as well as between CII and CIII ($P<0.05$, Fig. S4). Moreover, the ANI between CI and CIII was significantly higher than that between CII and CIII ($P<0.05$, Fig. S4).

The heatmap depicting the dissimilarity between genes (Fig. 3B) and dissimilarity between intergenic regions (Fig. 3C) in *C. acnes* revealed that the strains from evolutionary branches A, B, and C clustered together based on the results of the clustering analysis. Additionally, within the C evolutionary branch, lineage CI, CII, and CIII were clustered together, indicating a low degree of dissimilarity among strains within each evolutionary branch. This clustering pattern aligns with the phylogenetic tree and ANI clustering findings.

Comparison of differences in different genetic branches

To gain a deeper understanding of the branch characteristics, we compared the differences between the branches. Branch A exhibited a larger genome size of 2.44 ± 0.04 Mb, which was greater than that of branches B and C, and there was a statistically significant difference in genome size among the three branches ($P<0.05$, Fig. S5). On the other hand, Branch C demonstrated a higher GC% of 60.07% ± 0.07%, which was significantly higher than that of branch A ($P<0.05$), but there was no statistical difference when compared to branch B ($P>0.05$). The average number of CDS in branch A was 2427.33 ± 64.06, which was significantly greater than that in branches B and C ($P<0.05$). However, there was no statistically significant difference in CDS between branches B and C ($P>0.05$). Within branch C, the genome size and CDS of lineage CI were significantly larger than those of CII and CIII ($P<0.05$), whereas the GC% of CII was significantly greater than that of CI and CIII ($P<0.05$).

To investigate the differences in functional genes among the three branches of *C. acnes*, we annotated 464 strains of *C. acnes* using the VFDB, CARD, and CAZy databases. The results of VFDB annotation indicated that only a small number of strains were annotated with virulence genes (Table S4). For instance, in the *C. acnes* HL083PA1 (GCF_000144225.1) genome, *sig*A/*rpo*V genes encode RNA polymerase sigma factor, whereas in the *C. acnes* K124 (GCF_003131265.1) genome, *sig*A/*rpo*V is related to flagellar biosynthesis proteins. In the *C. acnes* 07T299 (GCF_024329285.1) genome, *nar*H is related to nitrate reductase. Although the above strains carry these genes, whether they express the related virulence factors needs to be confirmed through phenotypic experiments.

Furthermore, the CARD database annotation revealed that all 464 strains of *C. acnes* contained rifamycin antibiotic-related genes (*rpo*B2) and peptide antibiotic-related genes (*ugd*), whereas only two strains of *C. acnes* were annotated to contain the tetracycline antibiotic-related gene (*tet*(W)), and six strains of *C. acnes* carried genes related to lincosamide antibiotics (*Erm*X). In addition, two strains of *C. acnes* contained genes related to macrolide antibiotics (*Erm*(50)). Finally, one strain of *C. acnes* carried genes related to fluoroquinolone antibiotics (*Pmp*M).

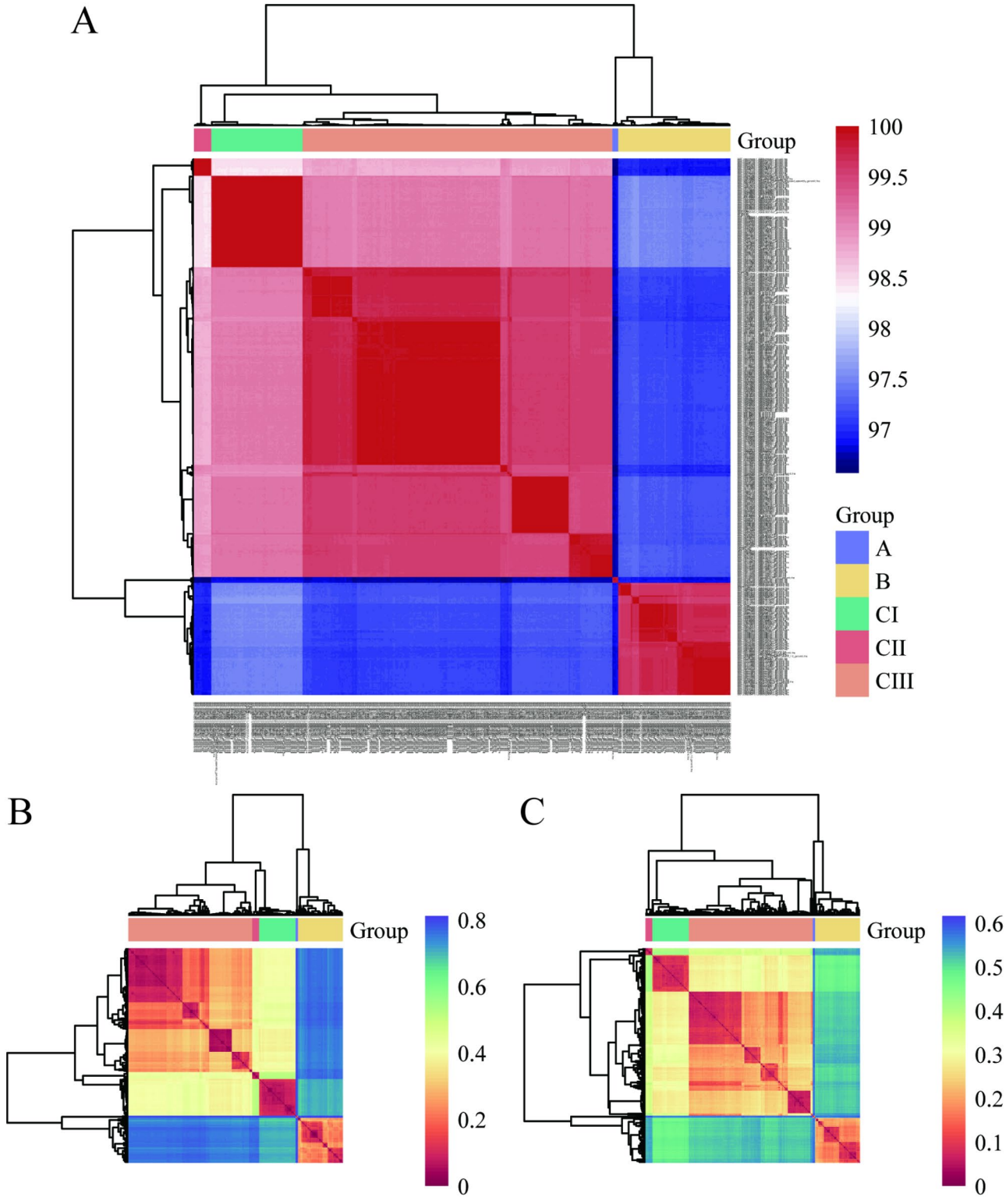

Fig. 3 Based on the phylogenetic tree, the genetic relationships among 464 strains of *Cutibacterium acnes* were divided into three genetic branches (A, B, C), with branch C further divided into (CI, CII, and CIII). Different colors represent the distinctive features of the separation of the five genetic branches (branches A, B, CI, CII, and CIII). (**A**) Average nucleotide identity (ANI) clustering heat map. (**B**) Gene dissimilarity heatmap based on the presence or absence of genes. (**C**) Intergenic region dissimilarity heatmap based on the presence or absence of intergenic regions

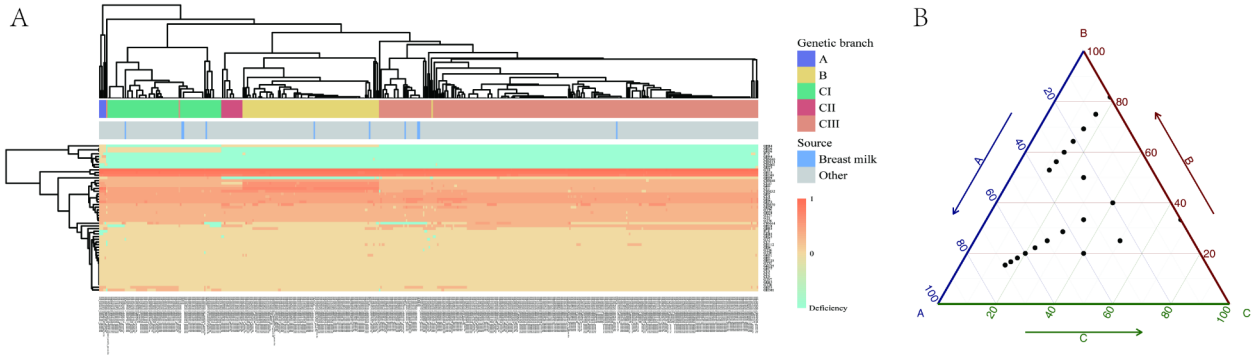

Fig. 4 (**A**) Cluster heatmap of 464 strains of *Cutibacterium acnes* based on carbohydrate-active enZYmes (CAZy) clustering ("Deficiency" means deficiency of the CAZy gene. The lg value of CAZY enzyme gene copy number was homogenized for all strains.). (**B**) A ternary map was drawn based on the frequency of occurrence of each CAZy gene in the 464 strains of *Cutibacterium acnes* genome in three major genetic branches

The CAZy database annotation revealed that *C. acnes* possess a diverse array of carbohydrate-active enzymes (Fig. 4A), including multiple copies of glycosyltransferases (GT4) and glycoside hydrolases (GH13 and GH109). The ternary phase diagram analysis indicated that different genetic branches exhibited varying carbohydrate metabolism capabilities (Fig. 4B). Branches A and B contained a higher number of copies of esterases (CE1 and CE10), glycoside hydrolases (GH3), and carbohydrate-binding modules (CBM32) than branch C. Additionally, branches B and C possess carbohydrate enzymes such as polysaccharide lyases (PL8) and glycoside hydrolases (GH23), which are absent in branch A. Furthermore, we compared the differences in carbohydrate-related genes between breast milk isolates and other source isolates. Our findings revealed that the breast milk isolates did not exhibit source-specific carbohydrate metabolism enzymes. However, compared to others strains, there were significantly fewer glycosyltransferases GT5 ($P<0.05$) and GT51 ($P<0.05$) but more GH31 genes in breast milk isolates ($P<0.05$).

Discussion

As a microorganism commonly found on human skin, *C. acnes* is known for its ability to adapt to its habitat [2]. Breast milk is a crucial source of nutrition for infants and has a significant impact on their health [42]. Breast milk contains a rich diversity of microorganisms, including *Staphylococcus epidermidis*, *Streptococcus agalactiae*, and *Clostridium acnes* [10]. However, there is limited knowledge regarding the population genetic background and functional genetic characteristics of *C. acnes* isolated from breast milk. To address this knowledge gap, we sequenced and assembled the genomes of ten strains of *C. acnes* isolated from breast milk and compared them with those of 454 strains of *C. acnes* available in the NCBI database. Fernández et al. [43] reported that microorganisms inhabiting maternal skin, such as *C. acnes* and *Corynebacterium* spp., may be transferred during breastfeeding. The phylogenetic tree revealed that 10 strains isolated from breast milk predominantly clustered within branches B and C at a ratio of 1:4. This ratio was similar to the overall ratio of *C. acnes* strains in branches B and C (97:362, respectively). Notably, no distinct clustering pattern was observed among the breast milk isolates. The absence of breast milk isolates in branch A may be due to the limited number of breast milk isolates. Therefore, we hypothesize that the strains isolated from breast milk are nearly identical to those from other isolated sources, probably because they came from other isolated sources, especially the nipple or breast surface. This is consistent with Dombrowska-Pali A et al. suggestion that the human milk microbiome comes from the mother's breast or the infant's oral cavity [9].

Genomic characteristics and phylogenetic relationships also supported this hypothesis. Furthermore, there were no significant differences between breast milk and other isolated sources strains in terms of virulence and antibiotic resistance genes. Although there was a significant increase in copy number of the carbohydrate metabolism enzyme gene GH31, we speculate this may be due to adaptive evolution to the breast milk environment. One of the most prominent features of breast milk is its abundance of free human milk oligosaccharides (HMOs) [44]. Glycoside hydrolase (GH) family 31 can hydrolyze α-Gal groups at the ends of oligosaccharides and polysaccharides [45]. Therefore, it is hypothesized that *C. acnes* strains sourced from breast milk have acquired more genes from the GH31 family through mechanisms like horizontal gene transfer, thereby adapting to the low oligosaccharide-rich environment of breast milk.

Although intergenic regions are composed of non-coding DNA sequences, they are essential components of bacterial genomes, harboring regulatory regions with extensive distributions that play a vital role in bacterial phenotypic variation [46]. In this study, we constructed phylogenetic trees for *C. acnes* using core genes, core intergenic regions, concatenated sequences of

core genes and core intergenic regions, and SNPs. The results showed that the structures of these phylogenetic trees were generally consistent and all displayed distinct clustering of different subspecies. The construction of a phylogenetic tree based on core intergenic regions was compared with the construction of a phylogenetic tree based on core genes, and a slight difference in the sub-branch within branch C, where strain HOL1 was located, was observed. This lineage was closer to the root of branch C in the phylogenetic tree constructed based on core intergenic regions rather than core genes, and strain SK187 was separated from this lineage. Phylogenetic trees constructed using concatenated sequences of core genes and core intergenic regions, which were more similar to the tree constructed based on core intergenic regions alone, indicated genetic differences between the core intergenic regions in branch C, where strain HOL1 is located. Suresh et al. [47] utilized core intergenic regions sequences to construct a phylogenetic tree of *Escherichia coli* core intergenic regions, which provided a clearer pattern of carrying polyketide synthase (*pks*) islands compared to the core gene phylogenetic system, thereby showing separate evolutionary branches of *pks*-negative clusters from *pks*-positive and mixed clusters. Nielsen et al. [48] conducted a population genomics analysis of *Streptococcus pneumoniae* and *Staphylococcus aureus*, and found a high degree of correlation between core intergenic regions and core genes. While most phylogenetic relationships currently focus on core genes, this study suggests that core intergenic regions can be used to describe phylogenetic relationships, thereby elucidating the evolutionary history of a species.

It is important to recognize that *C. acnes* possess three primary genetic branches, labeled A, B, and C, which are evident in both the phylogenetic tree structures and SNP-based principal component analysis. Furthermore, branch C can be subdivided into three lineages, which align with the three subspecies of *C. acnes* reported by Andrew McDowell et al. [49] as type III, II, and I, corresponding to *C. acnes* subsp. *elongatum*, *C. acnes* subsp. *defendens*, *and C. acnes* subsp. *acnes*, respectively. This observation supports the idea that *C. acnes* subsp. *acnes* can be further divided into three lineages. Analysis of CAZy revealed differences in the copy numbers of esterases (CE1 and CE10), glycoside hydrolases (GH3), and carbohydrate-binding modules (CBM32) between branches A, B, and C. Additionally, the carbohydrate enzyme genes polysaccharide lyases (PL8) and glycoside hydrolases (GH23) were present in branches B and C, but not in branch A. An annotation using the CARD database revealed that all 464 strains of *C. acnes* contained genes related to rifamycin antibiotics (*rpo*B2) and peptide antibiotics (*ugd*). Ulrika Furustrand Tafin et al. [50] discovered the emergence of clindamycin resistance in *C. acnes*, which was associated with mutations in the *rpo*B gene. Recently, Abriouel et al. [51] found that the *ugd* gene is widely present in S3 samples, with *C. acnes* being the dominant bacterial species in these samples. Therefore, we concluded that *rpo*B2 and *ugd* are inherent resistance-related genes present in the genome of *C. acnes.*

Conclusion

Comparative genomic analysis was conducted on *Cutibacterium acnes* strains obtained from breast milk and others. The results indicated that there were no differences in the genomic characteristics, phylogenetic relationships, virulence genes, and antibiotic resistance genes between breast milk isolates and others isolates. It is therefore speculated that the *C. acnes* in breast milk originates from the nipple or breast surface. However, significant differences were observed in the carbohydrate metabolism enzyme functional genes (GT5, GT51, and GH31), which suggests an adaptive evolution to the breast milk environment. This study identified three main genetic branches (A, B, and C) in *C. acnes*, which correspond to three subspecies (*C. acnes* subsp. *elongatum*, *C. acnes* subsp. *defendens*, *and C. acnes* subsp. *acnes*). Branches B and C exhibited smaller genetic differences than branch A. Branches A and B carried a higher number of copies of carbohydrate enzymes, such as CE1, CE10, GH3, and CBM32. Additionally, branches B and C possess carbohydrate enzymes PL8 and GH23, which are absent in branch A. Comparative phylogenetic tree analysis using concatenated sequences of core genes and core intergenic regions indicated that core intergenic regions could be used alongside core genes to describe phylogenetic relationships. This study provides a genetic basis for genetic differentiation of *C. acnes* strains from breast milk sources.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-024-03717-3.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7

Acknowledgements
All the authors thank Weicheng Li for advice.

Author contributions
Jiaqi Sun: Validation, visualization, writing original draft. Guoxuan Hang: Writing- Reviewing and Editing. Huimin Lv, Yu Li and Zhi Zhong: Data

curation. Qiujie Song: Visualization. Zhihong Sun: Supervision. Wenjun Liu: Conceptualization, Project administration. All authors read and approved the final manuscript.

Funding
This research supported by the National Natural Science Foundation of China (U23A20260), and the Science and Technology Development Fund Project for Central Guiding Local development (2022ZY0209), and the earmarked fund for CARS36.

Declarations

Ethics approval and consent to participate
This study was approved by the Special Committee on Scientific Research and Academic Ethics of Inner Mongolia Agricultural University (No. 2018-059), and written informed consent was received from all participants. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Key Laboratory of Dairy Biotechnology and Engineering (IMAU), Ministry of Education, Inner Mongolia Agricultural University, Hohhot, P.R. China
[2]Key Laboratory of Dairy Products Processing, Ministry of Agriculture and Rural Affairs, Inner Mongolia Agricultural University, Hohhot, P.R. China
[3]Inner Mongolia Key Laboratory of Dairy Biotechnology and Engineering, Inner Mongolia Agricultural University, Hohhot, P.R. China
[4]Collaborative Innovative Center for Lactic Acid Bacteria and Fermented Dairy Products, Ministry of Education, Inner Mongolia Agricultural University, Hohhot, P. R. China
[5]College of Food Science and Engineering, Inner Mongolia Agricultural University, 306 Zhaowuda Road, Hohhot, Inner Mongolia Autonomous Region, China

Received: 19 August 2024 / Accepted: 14 December 2024
Published online: 06 January 2025

References

1. Christensen GJM, Brüggemann H. Bacterial skin commensals and their role as host guardians. Beneficial Microbes. 2014;5(2):201–15.
2. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Program NCS, Bouffard GG, Blakesley RW, Murray PR. Topographical and temporal diversity of the human skin microbiome. Science. 2009;324(5931):1190–1192.
3. Rosenthal M, Goldberg D, Aiello A, Larson E, Foxman B. Skin microbiota: microbial community structure and its potential association with health and disease. Infect Genet Evol. 2011;11(5):839–48.
4. Szabó K, Erdei L, Bolla BS, Tax G, Bíró T, Kemény L. Factors shaping the composition of the cutaneous microbiota. Br J Dermatol. 2017;176(2):344–51.
5. Scholz CF, Kilian M. The natural history of cutaneous propionibacteria, and reclassification of selected species within the genus Propionibacterium to the proposed novel genera Acidipropionibacterium gen. nov., Cutibacterium gen. nov. and Pseudopropionibacterium gen. Nov. Int J Syst Evol MicroBiol. 2016;66(11):4422–32.
6. Yu Y, Champer J, Agak GW, Kao S, Modlin RL, Kim J. Different Propionibacterium acnes phylotypes induce distinct immune responses and express unique surface and secreted proteomes. J Invest Dermatology. 2016;136(11):2221–8.
7. McLaughlin J, Watterson S, Layton AM, Bjourson AJ, Barnard E, McDowell A. Propionibacterium acnes and acne vulgaris: new insights from the integration of population genetic, multi-omic, biochemical and host-microbe studies. Microorganisms. 2019;7(5):128.
8. Urbaniak C, Burton JP, Reid G. Breast, milk and microbes: a complex relationship that does not end with lactation. Women's Health. 2012;8(4):385–98.
9. Dombrowska-Pali A, Wiktorczyk-Kapischke N, Chrustek A, Olszewska-Słonina D, Gospodarek-Komkowska E, Socha MW. Human milk Microbiome—A review of scientific reports. Nutrients. 2024;16(10):1420.
10. Togo A, Dufour J-C, Lagier J-C, Dubourg G, Raoult D, Million M. Repertoire of human breast and milk microbiota: a systematic review. Future Microbiol. 2019;14(7):623–41.
11. Dréno B, Pécastaings S, Corvec S, Veraldi S, Khammari A, Roques C. Cutibacterium acnes (Propionibacterium acnes) and acne vulgaris: a brief look at the latest updates. J Eur Acad Dermatol Venereol. 2018;32:5–14.
12. Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA. Biogeography and individuality shape function in the human skin metagenome. Nature. 2014;514(7520):59–64.
13. Oh J, Byrd AL, Park M, Kong HH, Segre JA. Temporal stability of the human skin microbiome. Cell. 2016;165(4):854–66.
14. Cobian N, Garlet A, Hidalgo-Cantabrana C, Barrangou R. Comparative genomic analyses and CRISPR-Cas characterization of Cutibacterium acnes provide insights into genetic diversity and typing applications. Front Microbiol. 2021;12:758749.
15. Zhao S, Ci J, Xue J, Wang Y, Li X, Hao L, Tian L, Guo H, Xin C, Zhao Y. Cutibacterium acnes type II strains are associated with acne in Chinese patients. Antonie Van Leeuwenhoek. 2020;113(3):377–88.
16. Lomholt HB, Kilian M. Population genetic analysis of Propionibacterium acnes identifies a subpopulation and epidemic clones associated with acne. PLoS ONE. 2010;5(8):e12277.
17. Zhong Z, Tang H, Shen T, Ma X, Zhao F, Kwok L-Y, Sun Z, Bilige M, Zhang H. Bifidobacterium animalis subsp. lactis Probio-M8 undergoes host adaptive evolution by glc U mutation and translocates to the infant's gut via oral-/entero-mammary routes through lactation. Microbiome. 2022;10(1):197.
18. Zhu H, Qu F, Zhu L-H. Isolation of genomic DNAs from plants, fungi and bacteria using benzyl chloride. Nucleic Acids Res. 1993;21(22):5279.
19. Mo L, Yu J, Jin H, Hou Q, Yao C, Ren D, An X, Tsogtgerel T, Zhang H. Investigating the bacterial microbiota of traditional fermented dairy products using propidium monoazide with single-molecule real-time sequencing. J Dairy Sci. 2019;102(5):3912–23.
20. Liu W, Li W, Zheng H, Kwok L-Y, Sun Z. Genomics divergence of Lactococcus lactis subsp. lactis isolated from naturally fermented dairy products. Food Res Int. 2022;155:111108.
21. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):2047–217X-2041-2018.
22. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. Bioinformatics. 2008;24(5):713–4.
23. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
24. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114.
25. Wielgoss S, Didelot X, Chaudhuri RR, Liu X, Weedall GD, Velicer GJ, Vos M. A barrier to homologous recombination between sympatric strains of the cooperative soil bacterium Myxococcus xanthus. ISME J. 2016;10(10):2468–77.
26. Luo D, Wang X, Feng X, Tian M, Wang S, Tang S-L, Ang P Jr, Yan A, Luo H. Population differentiation of Rhodobacteraceae along with coral compartments. ISME J. 2021;15(11):3286–302.
27. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
28. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3.
29. Yu J, Song Y, Ren Y, Qing Y, Liu W, Sun Z. Genome-level comparisons provide insight into the phylogeny and metabolic diversity of species within the genus Lactococcus. BMC Microbiol. 2017;17:1–10.
30. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 2009;19(2):327–35.
31. Yang J, Tsukimi T, Yoshikawa M, Suzuki K, Takeda T, Tomita M, Fukuda S. Cutibacterium acnes (Propionibacterium acnes) 16S rRNA genotyping of microbial samples from possessions contributes to owner identification. mSystems. 2019;4:e00594–19.
32. Seemann T. Snippy: Rapid haploid variant calling and core genome alignment. GitHub. In.; 2020.

33. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. Genome Biol. 2016;17:1–14.
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
35. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, Edalatmand A, Petkau A, Syed SA, Tsang KK. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. Nucleic Acids Res. 2023;51(D1):D690–9.
36. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res. 2022;50(D1):D912–7.
37. Lu J, Shen T, Zhang Y, Ma X, Xu S, Awad S, Du M, Zhong Z. Safety assessment of Enterococcus lactis based on comparative genomics and phenotypic analysis. Front Microbiol. 2023;14:1196558.
38. Li W, Sun J, Jing Y, Zhao J, Wu Q, Liu J, Kwok L-Y, Zhang W, Sun Z, Zhong Z. Comparative genomics revealed wide intra-species genetic heterogeneity and lineage-specific genes of Akkermansia muciniphila. Microbiol Spectr. 2022;10(3):e02439–02421.
39. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2009;37(suppl1):D233–8.
40. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 2013;41(12):e121–121.
41. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.
42. Sánchez C, Franco L, Regal P, Lamas A, Cepeda A, Fente C. Breast milk: a source of functional compounds with potential application in nutrition and therapy. Nutrients. 2021;13(3):1026.
43. Fernández L, Rodríguez JM. Human milk microbiota: origin and potential uses. Milk Mucosal Immun Microbiome: Impact Neonate. 2020;94:75–85.
44. Coppa G, Pierani P, Zampini L, Bruni S, Carloni I, Gabrielli O. Characterization of oligosaccharides in milk and feces of breast-fed infants by high-performance anion-exchange chromatography. Bioactive Compon Hum milk. 2001:307–14.
45. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50(D1):D571–7.
46. Oren Y, Smith MB, Johns NI, Kaplan Zeevi M, Biran D, Ron EZ, Corander J, Wang HH, Alm EJ, Pupko T. Transfer of noncoding DNA drives regulatory rewiring in bacteria. Proc Natl Acad Sci. 2014;111(45):16112–16117.
47. Suresh A, Shaik S, Baddam R, Ranjan A, Qumar S, Jadhav S, Semmler T, Ghazi IA, Wieler LH, Ahmed N. Evolutionary dynamics based on comparative genomics of pathogenic Escherichia coli lineages harboring polyketide synthase (Pks) island. MBio. 2021;12(1):03634–03620. https://doi.org/10.1128/mbio
48. Nielsen FD, Møller-Jensen J, Jørgensen MG. Adding context to the pneumococcal core genes–a bioinformatic analysis of the intergenic pangenome of Streptococcus pneumoniae. bioRxiv. 2021;2021(2008):2029–458057.
49. McDowell A, Barnard E, Liu J, Li H, Patrick S. Proposal to reclassify Propionibacterium acnes type I as Propionibacterium acnes subsp. acnes subsp. nov. and Propionibacterium acnes type II as Propionibacterium acnes subsp. defendens subsp. Nov. Int J Syst Evol MicroBiol. 2016;66(12):5358–65.
50. Furustrand Tafin U, Trampuz A, Corvec S. In vitro emergence of rifampicin resistance in Propionibacterium acnes and molecular characterization of mutations in the rpoB gene. J Antimicrob Chemother. 2013;68(3):523–8.
51. Abriouel H, Manetsberger J, Lerma LL, Blanco MDP, Nogueras RM, Gómez NC, Benomar N. Metagenomic insights into microbial contamination in critical healthcare environments and the efficacy of a novel HLE disinfectant. Infect Disease Health. 2023;28(4):282–9.

Publisher's note