Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Biofilm Formation and Genomic Signatures of Circulating Bordetella pertussis Strains Isolated in Beijing, China

Dongshan Yan
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Jie Che
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Hanying Dai
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Luyao Guo
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Jie Li
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Bohan Chen
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Yuan Gao
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Hairui Wang
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Xueping Liu
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Li Xu
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Fuyi Han

National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Yunchuan Qiao

National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Maojun Zhang

National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

Zhujun Shao

shaozhujun@icdc.cn

National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

---

Research Article

Keywords: Bordetella pertussis, Biofilm, Autoagglutination, Pan-GWAS, SNP-GWAS

Posted Date: July 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10154686/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Microbiology on September 2nd, 2026. See the published version at https://doi.org/10.1186/s12866-026-05503-9.

Abstract

Background

Pertussis, an acute respiratory infection caused by Bordetella pertussis (*B. pertussis*), has resurged in China. Biofilm formation has recently been associated with the pathogenesis of *B. pertussis*. This study aims to analyze the biofilm formation capacity of circulating isolates and investigate the phenotype–genotype associations between biofilm formation and genomic characteristics.

Results

Among 84 *B. pertussis* isolates, 69.05% (58/84) were classified as strong biofilm formers, while 23.81% (20/84) and 5.95% (5/84) exhibited moderate or weak biofilm production, respectively. Only one isolate (1/84, 1.19%) was categorized as non-biofilm producers. Strong biofilm-forming isolates exhibited significantly faster autoagglutination than non-strong isolates ($P < 0.05$). Pan-GWAS identified multiple candidate genes significantly associated with biofilm phenotypes ($P < 0.05$), which were enriched in pathways involved in cell envelope remodeling, metabolic adaptation, nutrient acquisition, transport systems, and environmental sensing. SNP-GWAS further identified multiple core genome SNPs significantly associated with enhanced biofilm formation ($P < 0.05$). Functional analysis revealed significant enrichment of genes in functions related to lipoprotein maturation, autotransporter function, amino acid metabolism, and transport systems. Integration of pan-GWAS and SNP-GWAS further identified the genetic determinants correlated with biofilm phenotypic heterogeneity in *B. pertussis*.

Conclusions

Strong biofilm-forming isolates consistently exhibited higher levels of biofilm biomass and faster aggregation kinetics than non-strong isolates. GWAS analysis revealed that the observed phenotypic divergence is driven by coordinated remodeling of multiple biological pathways rather than a single genetic determinant. These findings provide new insights into the phenotype–genotype association signatures of biofilm formation in circulating *B. pertussis*.

Background

Bacterial biofilms are multicellular surface-adherent microbial communities encased in a self-produced or extracellular polymeric substances matrix that facilitate persistent colonization of the host respiratory tract[1]. Biofilm formation promotes bacterial resistance to antimicrobial agents, environmental survival, stress response and evaded innate host immune responses, and leads to persistent chronic infections[2]. Biofilms is a sequential and tightly regulated developmental process involving multiple phenotypically distinct phases, each marked by unique patterns of stage-specific protein expression and transcriptional profiles[3, 4]. This process has been observed on a variety of artificial surfaces, under static, shaking, and

fluid flow conditions[5, 6]. Microscopically, *B. pertussis* forms discrete cellular aggregates that progressively mature into three-dimensional microcolonies encased in an opaque matrix composed of DNA and polysaccharide[5, 7, 8]. In vivo, biofilms formed by *B. pertussis* and the closely related species *B. bronchiseptica* have been observed in the murine nasal cavity and trachea, which contribute to bacterial persistence and inter-host transmission[2, 9]. The biological importance of biofilm formation is further evidenced by studies showing that enhanced biofilm-forming strains exhibit increased epithelial cell adhesion and higher levels of respiratory tract colonization in animal models[10, 11]. Furthermore, proteomics studies conducted on planktonic and biofilm cells revealed time-dependent differential gene expression patterns and key metabolic pathway essential for the maturation of *B. pertussis* biofilms [4, 12].

Biofilm formation represents a pivotal survival strategy that allows enhanced respiratory tract colonization, persistence, transmission, and circulation of *B. pertussis* in humans[8]. Recent studies have indicated that biofilm formation has been associated with *B. pertussis* pathogenesis and antimicrobial tolerance[8, 13]. Biofilm grown *B. pertussis* clinical isolates confer increased tolerance to antimicrobial agents compared to planktonic cultures. Hyperbiofilm formation by *B. pertussis* correlates with enhanced virulence traits[8]. Although increased levels of biofilm formation by circulating strains from Europe and other high-income regions has been reported, very little is known about the variability in biofilm-forming capacity of circulating strains from China, particularly for the hypervirulent, macrolide-resistant *B. pertussis* (MRBP) strains[8, 13, 14]. It is also not known whether the biofilm-forming capacity varies among different allelic genotype strains in China. Furthermore, the genetic determinants underlying biofilm formation remained poorly characterized. To date, only a limited number of studies have investigated the roles of individual genes in biofilm formation, and most have relied on single-gene mutations or core-genome SNP analyses to characterize the associated phenotypes[15, 16].

Genome-wide association study (GWAS) has emerged as a powerful approach designed to identify genetic variations associated with specific traits or diseases by scanning the genome across multiple individuals[17, 18]. Indeed, GWAS has proven to be a formidable tool for deciphering the genetic architecture of complex traits, successfully transitioning from human genetics to microbial genomics[19, 20]. In recent years, GWAS has proven highly successful in microbial genomics for identifying genetic determinants underlying phenotypic characteristics[21, 22]. By leveraging large-scale detection of genetic variants, GWAS facilitates the reconstruction of evolutionary trajectories in pathogens, as exemplified by the host-adaptive evolution observed in Campylobacter[23]. Despite its potential for elucidating the genetic basis of biofilm formation, GWAS has not yet been systematically applied to investigate phenotype–genotype associations underlying biofilm formation among circulating *B. pertussis* isolates. Whether specific genomic signatures or accessory loci profiles are associated with enhanced biofilm formation has not been thoroughly investigated. Given the highly clonal population structure of epidemic *B. pertussis*, identifying how subtle genetic variation contribute to phenotypic heterogeneity remains crucial for elucidating the molecular mechanisms underlying biofilm formation.

In this study, we systematically quantified biofilm biomass and autoagglutination in clinical *B. pertussis* isolates from Beijing, China that were previously characterized by leveraging available WGS data. We next sought to identify multiple candidate genes and core genome SNPs associated with phenotypic features shared by distinct biofilm-forming groups using multiple GWAS approaches. Overall, our findings reveal associations between candidate loci and biofilm phenotypes, providing new insights into the genetic determinants underlying biofilm phenotypic heterogeneity and its contribution to biofilm development in *B. pertussis*.

Materials and methods

Bacterial Strains and Culture Conditions

A total of 84 *B. pertussis* isolates were analyzed in this study from the National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention[24]. All isolates were recovered from nasal swabs of respiratory outpatients and hospitalized patients between 2021 and 2024. Frozen glycerol stocks were streaked on charcoal agar supplemented with 10% sterile defibrinated sheep blood (BMT090, CHROMagar, Shanghai, China) and incubated at 37℃ for 5 days to obtain isolated colonies. All isolates were characterized by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (EXS 2600, Zybio, Chongqing, China) and quantitative real-time polymerase chain reaction (qPCR) (QuantGene9600, Bioer Technology, Hangzhou, Zhejiang, China), as previous described[25]. Detailed strain information is provided in **Table S1**. Single colonies were cultured for additional 72 h and inoculated into Stainer-Scholte (SS) broth (M2343; Tuopu Bioengineering, Shandong, China), followed by incubation at 37°C for 24 h with shaking at 180rpm and then used for biofilm assays. The reference strains *B. pertussis* ATCC1925, ATCC9340, and ATCC9797, were prepared as controls.

Biofilm Quantification with Microtiter Plate Assay

Biofilm biomass was quantified using a crystal violet (CV) microtiter plate assay as previously described[13, 14]. Briefly, 200μL of bacterial suspension adjusted to an OD600 of 0.1 was inoculated into polystyrene microtiter plates and incubated at 37°C under static conditions for 96 h. After incubation, the planktonic cultures were removed and wells were gently washed three times with sterile distilled water. Subsequently, the attached biofilms were stained with 200μL 0.1% crystal violet (CV) (Innochem, Beijing, China) for 30 min. The staining solution was discarded by washing three times and 200μL of 95% ethanol was added to dissolve the crystal violet. Absorbance was measured at OD540nm using a multimode microplate reader (EnSight, PerkinElmer, Suzhou, China). To evaluate biofilm formation kinetics, biofilm biomass was measured at 12 intervals over 96 h, respectively. Isolates were classified into four categories based on the optical density cut-off value as follows: strong (OD > 4ODc; red circles), moderate (4ODc ≥ OD > 2ODc; blue circles), weak (2ODc ≥ OD > ODc; green circles), and non- biofilm formers (ODc ≥ OD; orange circles). The cutoff value (ODc) was defined as the mean OD of the negative

control plus three standard deviations. All assays were performed in triplicate. The reference strains *B. pertussis* ATCC1925, ATCC9340 and ATCC9797 were used as positive control and sterile SS broth served as the negative control.

Autoagglutination assay

Autoagglutination was evaluated as described previously with minor modifications [8, 15]. Briefly, overnight cultures grown in SS broth were harvested by centrifugation, washed, and resuspended in sterile saline solution to an OD600 of 1.0. Then, 4mL aliquot of bacterial suspension was incubated statically at 37°C for 24h. The OD600 values of the upper suspension were measured at 2, 4, 8, 12 and 24 h by using a spectrophotometer. The Sterile saline solution served as the blank. The aggregation index (AI) was calculated as follows: $(OD_{t0}\text{-}OD_t)/OD_{t0}$, where OD_{t0} is the initial OD measurement and OD_t is the OD measured at a designated time point. All assays were conducted in triplicate with three independent biological replicates.

Pan-genome-wide association analysis

Pan-GWAS was conducted based on the accessory gene content of 84 de novo-assembled *B. pertussis* genomes[25]. Genomes were annotated using Prokka (v1.15.6) with default parameters[26]. The annotated GFF3 files were input into Roary (v3.13.0) to construct the pangenome using a minimum BLASTP identity of 95% and a core gene threshold of 99% (-i 95 -cd 99)[27]. Associations between gene presence/absence patterns and biofilm phenotypes were evaluated using PLINK (v1.90) [28] and Pyseer (v1.4.0) [29]. In PLINK, the Roary-derived gene presence/absence matrix was analyzed against binary biofilm phenotypes using Fisher's exact test, followed by Bonferroni and Benjamini–Hochberg correction for multiple comparisons. Genes with adjusted *P* values < 0.05 were considered significantly associated with the phenotype. In Pyseer, quantitative biofilm phenotypes were analyzed using a linear mixed model (LMM), with allele frequency thresholds of 0.01 and 0.99 (--min-af 0.01, --max-af 0.99). Associations were evaluated based on estimated effect sizes (β) and likelihood ratio test (lrt) *P* values. Given the extensive expansion of IS481 and IS110 insertion sequences in *B. pertussis*, transposase-associated loci were excluded from downstream biological interpretation to reduce the potential impact of lineage-associated confounding and spurious associations arising from mobile genetic elements.

Core genome SNPs association analysis

To further identify the genome variants associated with biofilm formation, SNP-based GWAS was performed using Pyseer (v1.4.0)[29]and GEMMA (v0.98.5)[30]. Briefly, core genome SNPs were identified by mapping sequencing reads to the *B. pertussis* Tohama I reference genome (NZ_CP039022.1) using Snippy (v4.6.0) with default parameters[31]. Biallelic SNPs with a minor allele frequency (MAF) ≥ 1% and ≤ 99% were retained for downstream analyses. For Pyseer analyses, population structure was controlled

using a phylogeny-derived similarity matrix. Briefly, recombination regions were filtered using Gubbins (v3.4.3), followed by reconstruction of a maximum-likelihood phylogeny with FastTree (v2.2.0) under the GTR substitution model. The resulting phylogenetic relatedness matrix was incorporated into a linear mixed model (LMM) to correct for population stratification. For GEMMA analyses, the filtered SNP matrix was converted to binary PLINK format. A kinship matrix was calculated using the "-gk 2" option to account for population structure. Associations between SNP genotypes and quantitative biofilm phenotypes were subsequently assessed using a linear mixed model implemented with the "-lmm 4" option, with SNP annotations incorporated through the "-a" parameter.

Statistical analysis

All the continuous variables in this study were expressed as number (%). Continuous variables were analyzed among strong, moderate and weak biofilm-forming groups using two-way ANOVA followed by Dunn's multiple-comparison tests. Correlations analysis was determined by the Spearman rank method. Statistical significance was defined as follows: $P < 0.05$ (*), $P < 0.01$(**), $P < 0.001$ (***), and $P < 0.0001$ (****), whereas "ns" represents no significant difference. Data analysis and visualization by using GraphPad Prism (v10.6.0) and R (v4.31).

Results

Characteristics of circulating B. pertussis clinical isolates

A total of 84 *B. pertussis* isolates were collected in Beijing, China between 2021 and 2024, while China is currently experiencing an epidemic of whooping cough. The characteristics of these strains are shown in **Table S1**. All isolated strains belonged to sequence type 2 (ST2) (100%). Based on whole-genome allele profiling, all isolates were classified into six major allelic combinations comprising *ptxP3/prn⁻/fhaB1* (39/84, 47.62%), *ptxP3/prn150/fhaB1* (35/84, 42.86%), *ptxP1/prn1/fhaB3* (7/84, 8.33%), *ptxP3/prn2/fhaB1* (1/84, 1.19%), *ptxP3/prn150/fhaB⁻* (1/84, 1.19%), and *ptxP3/prn⁻/fhaB⁻* (1/84, 1.19%). Overall, 77 isolates carried the *ptxP3* allele (91.67%), whereas only 7 isolates harbored *ptxP1*(8.33%). The *fhaB1* was identified in 75 isolates (89.29%), representing the predominant *fhaB* allele, followed by *fhaB3* in 7 isolates (8.33%) and the remaining 2 isolates exhibited *fhaB* naturally deficient genotypes (*fhaB⁻*) (2.38%). All isolates carried the *ptxA1*, *fim3-1*, and *fim2-1* alleles (100%), indicating genetic conservation of these loci across the studied population.

Biofilm phenotypes exhibited substantial heterogeneity among circulating B. pertussis isolates

The majority of circulating *B. pertussis* isolates formed biofilms, but substantial heterogeneity in biofilm-forming capacity was observed among isolates (Fig. 1a). Quantitative analysis demonstrated that the mean biofilm biomass of strong biofilm-forming isolates was 1.97-fold and 3.67-fold higher than that of moderate and weak producers, respectively. Among all tested isolates, 69.05% (58/84), 23.81% (20/84), 5.95% (5/84) were categorized as strong, moderate and weak biofilm producers, respectively (Fig. 1b, **Fig. S1**). Only one isolate (1/84, 1.19%) was classified as non-biofilm producer. Time-course

quantification of biofilm biomass over 96 h under different absorbance wavelengths revealed distinct phenotypic heterogeneity among isolates **(Fig. S2**). At OD540nm, the strong biofilm-forming group exhibiting the highest biofilm biomass, ranging from 0.85 to 3.59 (median 1.20), followed by moderate group (0.52–0.81, median 0.73), and weak group (0.36–0.38, median 0.37) (Fig. 1c). Furthermore, both *fhaB*$^{-}$ isolates exhibited a moderate biofilm-forming capacity relative to the ATCC reference strain. (**Fig. S3**).

Enhanced and progressive biofilm formation in circulating B. pertussis isolates

To further characterize the dynamics of biofilm development, we compared the biofilm-forming capacities of representative circulating isolates from the strong, moderate, and weak biofilm-forming groups. The result indicated that circulating isolates consistently exhibit an enhanced and progressive biofilm phenotypic characteristics. During roller drum cultivation in sterile glass tubes, we observed compact bacterial rings at the air-liquid interface, particularly in strains with strong biofilm-forming capacity. Similarly, the ATCC reference strains formed thicker and more compact rings at the air–liquid interface at the top of the bacterial suspension. In contrast, the moderate and weak biofilm formers produced thinner or diffused bacterial rings over the glass surface of the bacterial suspension. Notably, the non-biofilm forming isolate ID83 did not form a visible diffused ring at the air-liquid interface (Fig. 2a).

Quantitative analysis showed biofilm biomass increased progressively over time in all groups, with the strong biofilm-forming group consistently exhibiting significantly higher biofilm biomass than the non-strong group throughout the incubation period ($P < 0.05$; Fig. 2b). At 24 h, the mean OD540 values of the strong biofilm-forming group was 0.302 ± 0.043, followed by the moderate (0.229 ± 0.027) and weak group (0.175 ± 0.021). At 48 h, the corresponding values in the strong, moderate, and weak group were 0.765 ± 0.459, 0.328 ± 0.051 and 0.249 ± 0.031, respectively. At 72h, the strong group maintained a markedly higher biofilms level (1.229 ± 0.621) than the moderate (0.458 ± 0.129) and weak (0.303 ± 0.031) group. With prolonged cultivation, the differences between the strong and non-strong groups remained statistically significant ($P < 0.0001$). Biofilm biomass reached its maximum level at 96 h in all groups, with the strong-forming group exhibiting the highest biofilm biomass (2.101 ± 0.973), followed by the moderate (0.639 ± 0.091) and weak groups (0.366 ± 0.014). In line with this trend, the ATCC reference strains exhibited a progressive increase in biofilm biomass over time. In contrast, strain ID83 showed no measurable biofilm formation at any of the assessed time points.

Enhanced biofilm-forming strains display hyperaggregative properties

To evaluate the relationships between autoagglutination and biofilm formation in *B. pertussis*, we compared the aggregation index (AI) between the strong and non-strong biofilm-forming groups. At 2 h of incubation, the strong group exhibited a significantly higher mean AI values (0.100 ± 0.029) than the moderate (0.036 ± 0.013) and weak group (0.031 ± 0.007), corresponding to 2.8-fold and 3.2-fold

increases, respectively ($P < 0.0001$) (Fig. 3a-b). To monitor aggregation kinetics, AI values were determined at selected time points (2, 4, 8, 12, and 24 h) during a 24 h static incubation. At each time point from 4 h onward, the mean AI of the strong group were significantly higher than the non-strong group ($P < 0.0001$) (Fig. 3a-b). At 4 h, the strong group exhibited a mean AI of 0.168 ± 0.032, compared with the moderate (0.083 ± 0.004) and weak groups (0.067 ± 0.006). With prolonged incubation, the mean of AI values exhibited a time-dependent increase, followed by a gradual decline in OD absorbance. At 24 h, the mean AI values increased to 0.435 ± 0.038, 0.317 ± 0.040, and 0.204 ± 0.023 for the strong, moderate, and weak groups, respectively. The strong group exhibited significantly faster aggregation kinetics than the non-strong group across all time points. In contrast, the non-strong group exhibited substantially slower aggregation kinetics throughout the incubation period (Fig. 3a-b). The ATCC reference strains also displayed a time-dependent aggregation profile, as evidenced by a progressive increase in AI values over time (Fig. 3a-b). Notably, strain ID83 exhibited only minimal cell aggregation throughout the incubation period, demonstrating a markedly impaired aggregation phenotype (Fig. 3a-b).

Pan-genome-wide association analysis reveals accessory genes associated with biofilm formation in B. pertussis

To mitigate the limitations of individual association methods, we employed two orthogonal approaches, PLINK and Pyseer, to minimize false positives associations and triangulate potential genetic hits. PLINK analysis identified 300 clusters of orthologous groups (COGs) and genes were significantly associated with biofilm phenotypes at a nominal significance threshold (naïve $P < 0.05$) (**Tables S3-4**). After multiple-testing correlation, 10 associations remained significant following the stringent Bonferroni correction, whereas 62 hits (20.67%) retained significance using the less stringent Benjamini-Hochberg correction (adjusted $P < 0.05$) (Fig. 4a-b). Using the same dataset, Pyseer identified 143 significant associations after adjustment (lrt-$P < 0.05$) (Fig. 4c-d). Among these, 109 genes were positively associated with biofilm formation, whereas 34 showed negative associations (lrt-$P < 0.05$). All candidate genes ranked by adjusted *P* values are shown in **Tables S3-4**. Functional annotation revealed that 73.43% (105/143) of the candidate genes encoded transposases or insertion sequence (IS) elements (**Tables S3-4**). COG classification further indicated that multiple identified genes were associated with cell replication, recombination, and repair (72%), followed by energy production and conversion (4.67%) and cell wall/membrane/envelope biogenesis (4.0%) (**Fig. S4**). Genes involved in signal transduction, nucleotide transport and metabolism, and transcription were 2.67%, 0.67% and 0.67%, respectively (**Fig. S4**).

Genes involved in cell-envelope biogenesis and peptidoglycan remodeling exhibited the most significant associations across both pan-GWAS approaches. These genes include *mrcA*, *rlpA*, *lnt*, and *wbpO*. The *mrcA* gene, encoding a penicillin-binding protein, was associated with 90.48% (76/84) of biofilm-producing isolates and in 94.82% (55/58) of strong biofilm producers, but showed no association with non-biofilm isolate. Similarly, *rlpA*, encoding an endolytic peptidoglycan transglycosylase, was associated with 97.62% (82/84) of isolates and was present in all strong and moderate biofilm-producing strains. In addition, the *wbpO* was associated with 95% (19/20) of moderate biofilm-forming

isolates and 60.34% (35/58) of strong biofilm-forming isolates, whereas the *lnt* gene encoding the lipoprotein maturation ranked among the top loci in PLINK analysis and showed significant association with enhanced biofilm formation (lrt-*P* = 0.0299). Another functional category comprised genes involved in respiratory and metabolic adaptation, including *cydB* (86.90%) and *rubA2* (17.86%), as well as the environmental sensing and chemotaxis gene tap (2.38%), which are associated with processes relevant to biofilm formation and environmental adaptation. Several adhesin-related loci, including *fhaS/fhaB* and *tibA*-associated loci, were also enriched in strong biofilm-producing strains ($P < 0.05$). Additionally, the putative positive transcription regulator gene *bvgA* was uniquely identified by the Pyseer (**Tables S3-4**).

Core genome SNP-based GWAS reveals genetic loci associated with biofilm heterogeneity in B. pertussis

To further investigate the genetic variations associated with biofilm phenotypes, SNP-based GWAS was performed using both GEMMA and Pyseer. GEMMA identified 155 significant SNPs, whereas Pyseer yielded 36 significant SNPs associated with biofilm production (lrt-$P < 0.05$; **Tables S5-6**). All SNPs ranked by significance (lrt-*P* on y-axis) are shown in a Manhattan Plot in Fig. 5a-d with their genomic location in the Tohama I genome. After the multivariable correlation, the most significant hits are shown in **Tables S5-6** with 19 overlapping top-ranked SNPs were consistently identified by both approaches in bold, including 10 non-synonymous mutations comprising missense substitutions (8/19, 42.11%), one in-frame deletion (1/19, 5.27%), and one frameshift mutation (1/19, 5.26%). Functional annotation revealed that significant loci were enriched in genes involved in membrane organization, transport systems, protein secretion, cellular metabolism, environmental sensing and virulence regulation. Quantile–quantile plots showed acceptable control of population stratification in both models (GEMMA: $\lambda = 0.9094$; Pyseer: $\lambda = 1.2655$) (**Fig. S5**).

The most significant SNPs were identified in *sphB2*, which encodes the autotransporter SphB2. GEMMA identified two missense mutations (p.Leu70Trp and p.Asp73Ala) and one in-frame deletion (p.Gly87_Gly90del), whereas Pyseer only detected the same in-frame deletion within *sphB2*. All significant SNPs were enriched in strong biofilm-forming isolates and showed positive β coefficients, indicating a positive association between *sphB2* polymorphisms and enhanced biofilm formation (lrt-$P < 0.05$; **Tables S5-6**). Among genes involved in cell envelope remodeling and peptidoglycan maintenance, multiple SNP associations were detected in *lolA*, encoding an outer membrane lipoprotein chaperone. Two GEMMA-identified significant SNPs (p.Phe36Ile and p.Ala27_Thr28delinsProPro) were exclusively detected in isolate ID82, which exhibited the strongest biofilm-forming phenotype. Other genes involved in respiratory and metabolic adaptation, including *dadA* and a gene encoding a FUSC family protein, were significantly associated with SNPs. Notably, SNPs in *dadA* were correlated with enhanced biofilm formation ($P < 0.05$). Additional SNP associations were detected in genes involved in transport and nutrient acquisition, including an MFS transporter gene, *bugD* and *ugpA*. Significant SNPs within these genes were positively associated with biofilm formation ($P < 0.05$; **Tables S5-6**).

Discussion

In this study, we characterized substantial heterogeneity in biofilm-forming capacity among circulating *B. pertussis* strains isolated from Beijing, China. By integrating pan GWAS and SNP GWAS approaches with biofilm phenotype data, we identified accessory genes and core genome SNPs associated with variation in biofilm formation. We further investigated genotype–phenotype associations between biofilm phenotypic diversity and their underlying genetic determinants. Finally, we determined distinct signatures in biofilm formation and candidate genes, as well as core genome SNPs across the analyzed strains.

Consistent with what was previously reported[8, 13, 14], we observed that over 98.81% of the analyzed isolates produced biofilms, with 69.05% exhibiting strong, 23.81% moderate, 5.95% weak, and 1.19% non-biofilm-forming phenotypes, respectively. These findings suggested that circulating *B. pertussis* isolates exhibited heterogeneity in biofilm formation, marked by a predominance of the strong biofilm phenotype. To further characterize biofilm development, we monitored biofilm biomass over time. We observed that the biofilm biomass in the strong biofilm-forming group was significantly higher than the non-strong group, and the difference remained statistically significant throughout the entire experimental period. The relationship between allelic genotypes and biofilm formation remains poorly characterized. Although *prn* has been reported to facilitate host cell adhesion in earlier studies, *prn* deficiency alone did not significantly affect adhesion and invasion in our previous studies[32, 33]. Here, we found that the *prn150* and *prn*$^{-}$ isolates showed similarly high biofilm formation, with no significant correlations between the two groups ($P > 0.05$). These findings suggest that the *prn* gene may not be an essential determinant of biofilm formation in *B. pertussis*, as strains lacking this gene still formed biofilms.

Previous studies have shown that *fhaB* contributes to biofilm development by promoting microcolony formation[9, 34]. Here, we demonstrated that two *fhaB*$^{-}$ mutants displayed a markedly reduction in biofilm biomass and exhibited moderate biofilm phenotype. Consistent with Serra et al findings [9], our results provide strong evidence that *fhaB* plays a crucial role in the biofilm biosynthesis of *B. pertussis*. Recent studies have shown that *fim3* isolates harboring the *fim3*-2 allele displayed enhanced biofilm formation capacity but less autoagglutination compared to *fim2* isolates[35, 36]. Given that all isolates in our study harbored *fim3-1* and *fim2-1* alleles, the correlation between fimbriae serotype diversity and biofilm formation warrants further investigation in more genetically diverse populations. Collectively, these findings indicate that enhanced biofilm-forming capacity represents the predominant characteristic of recently circulating isolates, which may contribute to their sustained circulation and host colonization.

It has been reported that hyperbiofilm-forming strains displayed hyperaggregative properties[8]. In this study, we found a positive correlation between bacterial aggregation and enhanced biofilm-forming capacity within the biofilm-forming isolates. Strong biofilm-forming isolates exhibited greater autoagglutination than non-strong isolates, indicating enhanced cell–cell interactions. Notably, we observed the sole non-biofilm forming isolate displayed minimal cell aggregation throughout the incubation period, further supporting a close association between autoaggregation and biofilm formation in circulating *B. pertussis* isolates. Our results suggest that both bacterial aggregation and biofilm formation are mediated by shared physical adhesive forces and potentially overlapping

extracellular matrix components that promote cell-cell interactions essential for biofilm formation. In our previous study, we observed that two *fhaB*$^{-}$ isolates exhibited markedly reduced bacterial aggregation[32]. Our data suggest that *fhaB* plays a crucial role in the synthesis of the biofilm matrix. Taken together, these findings indicated that the currently circulating strains utilize aggregation as a means to enhance their biofilm-forming capacities.

Previous GWAS and pan-GWAS studies have identified multiple accessory genes are significantly associated with genotypes of multidrug-resistant *Pseudomonas aeruginosa* that exhibit enhanced biofilm production[37]. In the present study, by integrating pan-GWAS and SNP-GWAS analyses, we found that biofilm formation in *B. pertussis* is a multifactorial adaptive phenotype shaped by coordinated adaptive remodeling of multiple biological pathways rather than being dominated by a single genetic determinant. Although two GWAS algorithms interrogate distinct genomic dimensions, both independently converged on overlapping functional pathways related to cell envelope remodeling, respiratory adaptation, metabolic flexibility, nutrient acquisition, transport systems and environmental sensing, reinforcing the robustness and biological relevance of the identified associations. The functional convergence of significant association signals suggested that heterogeneous biofilm phenotypes arise through coordinated physiological adaptation rather than isolated genetic changes.

In our study, the prominent functional categories identified by both GWAS approaches involved genes associated with cell-envelope biogenesis and peptidoglycan remodeling, including *mrcA*, *rlpA*, *lnt*, and *wbpO*. Among these, *mrcA* is involved in peptidoglycan synthesis and cell wall integrity, whereas *rlpA* contributes to peptidoglycan turnover and envelope remodeling, which both exhibited the strongest association signals for biofilm formation[38, 39]. These findings are consistent with previous reports in Gram-negative bacteria indicating that peptidoglycan remodeling contributes to surface attachment, microcolony formation, and stress adaptation during biofilm development [40]. Previous transcriptomic studies have reported that *wbpO* is upregulated in biofilm cells[12]. Here, we also detected a significant association signal between the presence of *wbpO* and enhanced biofilm formation in *B. pertussis*, suggesting that this locus may contribute to biofilm stability and surface adherence. The *lnt*-associated loci further highlighted the contribution of membrane lipoprotein maturation to biofilm stability, likely by modulating extracellular matrix organization and surface properties. Additional cell wall-associated loci were identified by SNP-GWAS, including *lolA*, *amiD*, *alr*, *fabI*, and *envC*. Previous studies have demonstrated that *lolA* plays an important role in biofilm formation[41, 42]. In this work, we identified significant SNPs within *lolA* gene by both SNP-GWAS. Strains harboring these genetic variants exhibited significantly higher biofilm levels. Although multiple SNPs were detected in different isolate subsets, their clustering within *lolA* suggests that allelic heterogeneity at this locus may contribute to biofilm phenotypic variation in *B. pertussis*. Together, these findings indicate that coordinated modulation of envelope composition and surface polysaccharide architecture plays an important role underlying biofilm formation in *B. pertussis*.

Adhesion-associated pathways were enriched in biofilm-associated loci, highlighting a key role of bacterial adhesion in biofilm development. Previous studies have indicated that filamentous hemagglutinin family proteins play a role in mediating epithelial attachment and bacterial aggregation in *Bordetella* species[9, 43]. In this study, we identified adhesin-associated loci involving *fhaB*, *tibA*, *fimV*-associated pilus assembly proteins and *sphB2*-related variants were enriched among strong biofilm-forming isolates, suggesting that enhanced surface attachment and aggregation ability are critical during the initial biofilm synthesis. The *sphB2* gene encodes a Type V secretion system-associated protease involved in FHA maturation[44, 45]. SNP-GWAS identified significant variants in the *sphB2* gene, all exhibiting positive β coefficients, which were associated with enhanced biofilm formation. We found these SNPs were present exclusively in isolate ID82, which demonstrated the strongest biofilm-forming capacity. Our data further support a robust association between *sphB2* polymorphisms and biofilm formation. Overall, these findings are consistent with the functional role of filamentous hemagglutinin and autotransporter proteins in mediating host interaction, aggregation, and biofilm initiation in *Bordetella* species[9].

Another prominent functional category jointly identified by GWAS involved pathways related to respiratory adaptation and redox homeostasis. Associated loci included *cydB/cioB, fdx*, and *rubA*, suggesting that metabolic flexibility may contribute to biofilm formation in *B. pertussis*. The enrichment of these loci together with D-amino acid metabolic genes suggests that biofilm-forming strains are adapted to hypoxic and oxidative stress conditions characteristic of mature biofilms. Previous studies have demonstrated that the *cydB* deletion mutant produced weaker, more loosely structured, and lower-volume biofilm in *Campylobacter jejuni* [46]. In this study, the *cydB* gene was enriched among strong and moderate biofilm-forming isolates, which supports the key role for respiratory adaptation in facilitating survival under the oxygen-limited conditions characteristic of producing biofilms. The D-amino acid dehydrogenase has previously been implicated in peptidoglycan remodeling and biofilm regulation in multiple bacterial species[47–49]. Our analysis revealed that multiple SNPs within *dadA* exhibited the strongest association signals with the phenotype, highlighting the critical involvement of D-amino acid metabolism in biofilm adaptation. In addition, *fdx*-associated loci were detected in nearly all strong and moderate biofilm formers, further highlighting the importance of redox homeostasis in biofilm adaptation. Together, these findings suggest that biofilm formation in *B. pertussis* involves coordinated metabolic and physiological remodeling rather than isolated metabolic changes.

Environmental sensing and virulence-related pathways emerged as important components of the biofilm-associated genomic landscape. In this work, we identified several significant loci encoding a Cache domain protein, PhoU, and an ArsR/SmtB family transcriptional regulator, all of which were associated with biofilm formation. Chemotaxis-associated loci, including *tar* and *tap*, were identified, suggesting roles in environmental sensing and spatial colonization during biofilm development, potentially contributing to bacterial persistence under nutrient-limited conditions[50, 51]. Previous studies have shown that the BvgAS two-component regulatory systems plays a central role in regulating virulence and biofilms formation in *B. pertussis* [14, 52]. Consistent with these observations, our GWAS

identification of *bvgA* further supports the involvement of Bvg-mediated signaling in biofilm heterogeneity. Evidence has indicated that deletion of *bpeGReg* gene, encoding a globin-coupled diguanylate cyclase involved in c-di-GMP synthesis, significantly reduced biofilm formation in *B. pertussis*[53]. We also found phosphodiesterase-related loci, including a 3'3'-cGAMP phosphodiesterase, which suggest that cyclic nucleotide signaling pathways contribute to regulating sessile growth and biofilm homeostasis. Together, these findings indicate that environmental sensing and signal transduction pathways are important determinants of biofilm formation in *B. pertussis*.

Nutrient acquisition and metabolic flexibility pathways represented another important adaptive layer. We observed multiple associated loci encoding surface proteins or transporters, including components of MFS, ABC, TRAP, UgpA, as well as systems involved in oligopeptide, branched-chain amino acid, phosphate, and C4-dicarboxylate transport systems. Among these, an ABC transporter gene harbored the largest number of SNPs, including several non-conservative missense mutations that may influence transporter function and biofilm formation. The MFS transporter warrants further investigation, as mutations in this gene has been shown to reduce biofilm formation in both *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis*[54, 55]. In this study, we also identified multiple missense variants and two complex indels within the MFS transporter gene *EQ877_RS00175*, all of which were exclusively present in two strong biofilm-forming isolates. Additional metabolism-associated loci, including *livH*, *nicX*, *nicF*, *maeB2*, *folP*, *yngG*, and *glnA*-related genes, were enriched in strong biofilm-forming isolates and exhibited relatively large effect sizes despite low allele frequencies. The consistent identification of these loci across independent GWAS analyses highlights the importance of nutrient utilization, energy metabolism, and substrate transport during biofilm development. Together, these findings suggest that enhanced nutrient scavenging and metabolic adaptation facilitate bacterial persistence under nutrient-limited biofilm conditions and contribute to lineage-associated variation in biofilm phenotypes.

Several limitations of this study should be acknowledged. Firstly, the strong clonality and limited sample size reduced the statistical power to detect rare or subtle genetic loci with modest effects. Secondly, several associated loci were identified at relatively low frequencies and therefore require validated in larger strain collections, as some accessory loci may reflect lineage-specific adaptation rather than conserved determinants of biofilm formation. Thirdly, most genome assemblies analyzed in this study were based on short-read sequencing data, which may limit assembly continuity and resolution in repetitive genomic regions enriched with IS481 elements. This limitation could reduce the sensitivity of GWAS analyses for detecting structural variants and accessory genomic features. Future integration of long-read sequencing technologies will help to improve genome completeness and facilitate more comprehensive characterization of complex genomic regions. Moreover, GWAS analyses identify statistical associations rather than direct causal mechanisms, and further functional studies, including targeted mutagenesis and transcriptomic profiling will be essential to clarify both the mechanistic contributions and biological relevance of candidate loci involved in *B. pertussis* biofilm formation.

Despite the limitations acknowledged above, a major strength of this study is that two independent GWAS approaches converged on the same biological pathways. Pan-GWAS primarily captured accessory

genome gain or loss events, whereas SNP-GWAS additionally detected allelic polymorphisms within the conserved core genome. Although these associate loci identified by each approach were not completely overlapping, the convergence of biological pathways across independent analyses strengthens confidence in the identified associations and reduces the likelihood that the observed signals are driven solely by population structure or lineage-specific artifacts. Overall, our integrated genomic analyses support a multifactorial adaptive model in which accessory genome diversification and fine-scale allelic variation contribute to biofilm formation in *B. pertussis*.

Conclusion

In this study, we found significant differences in biofilm-forming capacity between strong and non-strong biofilm formers. Strong biofilm formers increased progressively over time and remained consistently higher than that of non-strong biofilm formers. Moreover, the strong biofilm formers also displayed significantly enhanced auto-aggregation kinetics, as reflected by higher average AI values across all tested time points. Through multiple GWAS analyses, we identified multiple genetic loci associated with enhanced biofilm production, which supporting the reliability of these associations. GWAS further revealed that the biofilm phenotype arises not from a single dominant genetic determinant, but rather from the coordinated adaptive remodeling of multiple biological pathways. Collectively, these findings illuminate the genetic basis underlying biofilm phenotypic diversity and provide new insights into genotype–phenotype relationships that govern biofilm formation in circulating *B. pertussis* strains.

Abbreviations

B. pertussis
Bordetella pertussis
BPagST
BP vaccine antigens sequence types
BLAST
Basic local alignment search tool
COGs
Clusters of orthologous groups
GWAS
Genome-wide association study
lrt
likelihood-ratio test
MAST
multilocus antigen sequence typing
MLST
Multilocus Sequence Typing
OD
Optical density

SNP
Single Nucleotide Polymorphism
ST
Sequence Type
WGS
Whole Genome Sequencing.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Beijing Center for Disease Prevention and Control (Approval No. 2021001, 6 December 2021). All procedures involving human participants were performed in accordance with the principles of the Declaration of Helsinki and applicable institutional guidelines and regulations. Written informed consent was obtained from all participants or their legal guardians/next of kin prior to their enrollment in the study.

Consent for publication

Not applicable.

Competing Interests

The authors declare no competing interests.

Funding

This work was supported by the Beijing Natural Science Foundation–Haidian Original Innovation Joint Fund (L192060), the National Natural Science Foundation of China (81973106), and the National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention (32108).

Author Contribution

D.S.Y., J.C., and Z.J.S. conceived and designed the experiments. D.S.Y. wrote the main manuscript and performed the experiments. J.C., H.Y.D., and L.Y.G. revised the manuscript. J.L., B.C., and Y.G. performed the formal analysis. H.W., X.L., L.X., F.H., Y.Q., and M.J.Z. analyzed the data and provided reagents. Z.J.S. contributed to study design, project administration, and funding acquisition. All authors reviewed the manuscript and approved the final version.

Acknowledgement

The authors would like to thank all participants for their participation in this study

Data Availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

References

1. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol. 2004;2(2):95–108; doi: 10.1038/nrmicro821.
2. Cattelan N, Dubey P, Arnal L, Yantorno OM, Deora R. Bordetella biofilms: a lifestyle leading to persistent infections. Pathog Dis. 2016;74(1):ftv108; doi: 10.1093/femspd/ftv108.
3. Malviya J, Alameri AA, Al-Janabi SS, Fawzi OF, Azzawi AL, Obaid RF, et al. Metabolomic profiling of bacterial biofilm: trends, challenges, and an emerging antibiofilm target. World J Microbiol Biotechnol. 2023;39(8):212; doi: 10.1007/s11274-023-03651-y.
4. Suyama H, Luu LDW, Zhong L, Raftery MJ, Lan R. Proteomic comparison of epidemic Australian Bordetella pertussis biofilm cells. Microbiol Spectr. 2025;13(11):e0171525; doi: 10.1128/spectrum.01715-25.
5. Conover MS, Mishra M, Deora R. Extracellular DNA is essential for maintaining Bordetella biofilm integrity on abiotic surfaces and in the upper respiratory tract of mice. PLoS One. 2011;6(2):e16861; doi: 10.1371/journal.pone.0016861.
6. Mirzaei R, Mohammadzadeh R, Alikhani MY, Shokri Moghadam M, Karampoor S, Kazemi S, et al. The biofilm-associated bacterial infections unrelated to indwelling devices. IUBMB Life. 2020;72(7):1271–85; doi: 10.1002/iub.2266.
7. Conover MS, Sloan GP, Love CF, Sukumar N, Deora R. The Bps polysaccharide of Bordetella pertussis promotes colonization and biofilm formation in the nose by functioning as an adhesin. Mol Microbiol. 2010;77(6):1439–55; doi: 10.1111/j.1365-2958.2010.07297.x.
8. Cattelan N, Jennings-Gee J, Dubey P, Yantorno OM, Deora R. Hyperbiofilm Formation by Bordetella pertussis Strains Correlates with Enhanced Virulence Traits. Infect Immun. 2017;85(12); doi: 10.1128/iai.00373-17.
9. Serra DO, Conover MS, Arnal L, Sloan GP, Rodriguez ME, Yantorno OM, et al. FHA-mediated cell-substrate and cell-cell adhesions are critical for Bordetella pertussis biofilm formation on abiotic surfaces and in the mouse nose and the trachea. PLoS One. 2011;6(12):e28811; doi: 10.1371/journal.pone.0028811.
10. van den Berg BM, Beekhuizen H, Willems RJ, Mooi FR, van Furth R. Role of Bordetella pertussis virulence factors in adherence to epithelial cell lines derived from the human respiratory tract. Infect Immun. 1999;67(3):1056–62; doi: 10.1128/iai.67.3.1056-1062.1999.

11. Conover MS, Redfern CJ, Ganguly T, Sukumar N, Sloan G, Mishra M, et al. BpsR modulates Bordetella biofilm formation by negatively regulating the expression of the Bps polysaccharide. J Bacteriol. 2012;194(2):233–42; doi: 10.1128/jb.06020-11.
12. Suyama H, Luu LDW, Zhong L, Raftery MJ, Lan R. Integrating proteomic data with metabolic modeling provides insight into key pathways of Bordetella pertussis biofilms. Front Microbiol. 2023;14:1169870; doi: 10.3389/fmicb.2023.1169870.
13. Dorji D, Graham RM, Richmond P, Keil A, Mukkur TK. Biofilm forming potential and antimicrobial susceptibility of newly emerged Western Australian Bordetella pertussis clinical isolates. Biofouling. 2016;32(9):1141–52; doi: 10.1080/08927014.2016.1232715.
14. Arnal L, Grunert T, Cattelan N, de Gouw D, Villalba MI, Serra DO, et al. Bordetella pertussis Isolates from Argentinean Whooping Cough Patients Display Enhanced Biofilm Formation Capacity Compared to Tohama I Reference Strain. Front Microbiol. 2015;6:1352; doi: 10.3389/fmicb.2015.01352.
15. Otsuka N, Koide K, Goto M, Kamachi K, Kenri T. Fim3-dependent autoagglutination of Bordetella pertussis. Sci Rep. 2023;13(1):7629; doi: 10.1038/s41598-023-34672-0.
16. de Jonge EF, Vogrinec L, van Boxtel R, Tommassen J. Inactivation of the Mla system and outer-membrane phospholipase A results in disrupted outer-membrane lipid asymmetry and hypervesiculation in Bordetella pertussis. Curr Res Microb Sci. 2022;3:100172; doi: 10.1016/j.crmicr.2022.100172.
17. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219–24; doi: 10.1038/s41588-018-0183-z.
18. Zhao C. Genome wide association study gateway-transitioning variants from association to causality in complex diseases. Sleep. 2024;47(7); doi: 10.1093/sleep/zsae116.
19. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–84; doi: 10.1038/s41576-019-0127-1.
20. Lee BH, Cole S, Badel-Berchoux S, Guillier L, Felix B, Krezdorn N, et al. Biofilm Formation of Listeria monocytogenes Strains Under Food Processing Environments and Pan-Genome-Wide Association Study. Front Microbiol. 2019;10:2698; doi: 10.3389/fmicb.2019.02698.
21. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53(2):156–65; doi: 10.1038/s41588-020-00763-1.
22. Munday RM, Haque R, Jan NJ, Wojcik GL, Marie C, Duchen D, et al. Genome-Wide Association Study of Campylobacter-Positive Diarrhea Identifies Genes Involved in Toxin Processing and Inflammatory Response. mBio. 2022;13(3):e0055622; doi: 10.1128/mbio.00556-22.
23. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. Proc Natl Acad Sci U S A. 2013;110(29):11923–7; doi: 10.1073/pnas.1305559110.

24. Wang Y, Wang H, Li Y, Yan A, Kang Y, Liu L, et al. Analysis of pertussis surveillance data (2022-2023) in a district of Beijing. BMC Infect Dis. 2025;25(1):1834; doi: 10.1186/s12879-025-12225-2.
25. Zhou G, Li Y, Wang H, Wang Y, Gao Y, Xu J, et al. Emergence of Erythromycin-Resistant and Pertactin- and Filamentous Hemagglutinin-Deficient Bordetella pertussis Strains - Beijing, China, 2022-2023. China CDC Wkly. 2024;6(20):437–41; doi: 10.46234/ccdcw2024.085.
26. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9; doi: 10.1093/bioinformatics/btu153.
27. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3; doi: 10.1093/bioinformatics/btv421.
28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7; doi: 10.1186/s13742-015-0047-8.
29. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. Bioinformatics. 2018;34(24):4310–2; doi: 10.1093/bioinformatics/bty539.
30. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44(7):821–4; doi: 10.1038/ng.2310.
31. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825–9; doi: 10.1093/molbev/msab293.
32. Dai H, Sui X, Che J, Zhang C, Zhao J, Li J, et al. Pathogenicity and vaccine protection of circulating pertactin- and filamentous hemagglutinin-deficient Bordetella pertussis strains. Emerg Microbes Infect. 2026;15(1):2640283; doi: 10.1080/22221751.2026.2640283.
33. Leininger E, Roberts M, Kenimer JG, Charles IG, Fairweather N, Novotny P, et al. Pertactin, an Arg-Gly-Asp-containing Bordetella pertussis surface protein that promotes adherence of mammalian cells. Proc Natl Acad Sci U S A. 1991;88(2):345–9; doi: 10.1073/pnas.88.2.345.
34. Hoffman C, Eby J, Gray M, Heath Damron F, Melvin J, Cotter P, et al. Bordetella adenylate cyclase toxin interacts with filamentous haemagglutinin to inhibit biofilm formation in vitro. Mol Microbiol. 2017;103(2):214–28; doi: 10.1111/mmi.13551.
35. Niinikoski V, Barkoff AM, Mertsola J, He Q. Bordetella pertussis isolates in Finland after acellular vaccination: serotype change and biofilm formation. Clin Microbiol Infect. 2024;30(5):683.e1–.e3; doi: 10.1016/j.cmi.2024.01.021.
36. Matczak S, Bouchez V, Leroux P, Douché T, Collinet N, Landier A, et al. Biological differences between FIM2 and FIM3 fimbriae of Bordetella pertussis: not just the serotype. Microbes Infect. 2023;25(7):105152; doi: 10.1016/j.micinf.2023.105152.
37. Redfern J, Wallace J, van Belkum A, Jaillard M, Whittard E, Ragupathy R, et al. Biofilm associated genotypes of multiple antibiotic resistant Pseudomonas aeruginosa. BMC Genomics. 2021;22(1):572; doi: 10.1186/s12864-021-07818-5.

38. Zhao J, Yang W, Deng H, Li D, Wang Q, Yi L, et al. Matrine reverses the resistance of Haemophilus parasuis to cefaclor by inhibiting the mutations in penicillin-binding protein genes (ftsI and mrcA). Front Microbiol. 2024;15:1364339; doi: 10.3389/fmicb.2024.1364339.
39. Palaniappan S, Gunasekaran V, Muthu T, Venkatesan LS, Sathishkumar P. Prevalence and effective treatment of drug-resistant Pseudomonas aeruginosa biofilm using flavonoid naringin. Bioprocess Biosyst Eng. 2025;48(12):2125–36; doi: 10.1007/s00449-025-03232-3.
40. Fidopiastis PM, Mariscal V, McPherson JM, McAnulty S, Dunn A, Stabb EV, et al. Vibrio fischeri Amidase Activity Is Required for Normal Cell Division, Motility, and Symbiotic Competence. Appl Environ Microbiol. 2021;87(3); doi: 10.1128/aem.02109-20.
41. Kaplan E, Greene NP, Crow A, Koronakis V. Insights into bacterial lipoprotein trafficking from a structure of LolA bound to the LolC periplasmic domain. Proc Natl Acad Sci U S A. 2018;115(31):E7389–e97; doi: 10.1073/pnas.1806822115.
42. Liao CT, Chiang YC, Hsiao YM. Functional characterization and proteomic analysis of lolA in Xanthomonas campestris pv. campestris. BMC Microbiol. 2019;19(1):20; doi: 10.1186/s12866-019-1387-9.
43. Nash ZM, Inatsuka CS, Cotter PA, Johnson RM. Bordetella filamentous hemagglutinin and adenylate cyclase toxin interactions on the bacterial surface are consistent with FhaB-mediated delivery of ACT to phagocytic cells. mBio. 2024;15(5):e0063224; doi: 10.1128/mbio.00632-24.
44. Suzuki K, Shinzawa N, Ishigaki K, Nakamura K, Abe H, Fukui-Miyazaki A, et al. Protective effects of in vivo-expressed autotransporters against Bordetella pertussis infection. Microbiol Immunol. 2017;61(9):371–9; doi: 10.1111/1348-0421.12504.
45. Khazani Asforooshani M, Noori Goodarzi N, Shahbazi B, Rezaie Rahimi N, Mahdavian K, Rohani M, et al. Exploiting subtractive genomics to identify novel drug targets and new immunogenic candidates against Bordetella pertussis: an in silico study. Front Bioinform. 2025;5:1570054; doi: 10.3389/fbinf.2025.1570054.
46. Korkus J, Sałata P, Thompson SA, Paluch E, Bania J, Wałecka-Zacharska E. The role of cydB gene in the biofilm formation by Campylobacter jejuni. Sci Rep. 2024;14(1):26574; doi: 10.1038/s41598-024-77556-7.
47. Feng W, Chittò M, Xie W, Ren Q, Liu F, Kang X, et al. Poly(d-amino acid) Nanoparticles Target Staphylococcal Growth and Biofilm Disassembly by Interfering with Peptidoglycan Synthesis. ACS Nano. 2024;18(11):8017–28; doi: 10.1021/acsnano.3c10983.
48. Zhang H, Mi Z, Wang J, Zhang J. D-histidine combated biofilm formation and enhanced the effect of amikacin against Pseudomonas aeruginosa in vitro. Arch Microbiol. 2024;206(4):148; doi: 10.1007/s00203-024-03918-4.
49. Miyamoto T, Saitoh Y, Katane M, Sekine M, Homma H. YgeA is involved in L- and D-homoserine metabolism in Escherichia coli. FEMS Microbiol Lett. 2022;369(1); doi: 10.1093/femsle/fnac096.
50. Huang Z, Wang YH, Zhu HZ, Andrianova EP, Jiang CY, Li D, et al. Cross Talk between Chemosensory Pathways That Modulate Chemotaxis and Biofilm Formation. mBio. 2019;10(1); doi:

10.1128/mBio.02876-18.

51. Dwivedi R, Nothaft H, Garber J, Xin Kin L, Stahl M, Flint A, et al. L-fucose influences chemotaxis and biofilm formation in Campylobacter jejuni. Mol Microbiol. 2016;101(4):575–89; doi: 10.1111/mmi.13409.
52. Mishra M, Parise G, Jackson KD, Wozniak DJ, Deora R. The BvgAS signal transduction system regulates biofilm development in Bordetella. J Bacteriol. 2005;187(4):1474–84; doi: 10.1128/jb.187.4.1474-1484.2005.
53. Wan X, Tuckerman JR, Saito JA, Freitas TA, Newhouse JS, Denery JR, et al. Globins synthesize the second messenger bis-(3'-5')-cyclic diguanosine monophosphate in bacteria. J Mol Biol. 2009;388(2):262–70; doi: 10.1016/j.jmb.2009.03.015.
54. Bock LJ, Ferguson PM, Clarke M, Pumpitakkul V, Wand ME, Fady PE, et al. Pseudomonas aeruginosa adapts to octenidine via a combination of efflux and membrane remodelling. Commun Biol. 2021;4(1):1058; doi: 10.1038/s42003-021-02566-4.
55. Li D, Zhang X, Yao Y, Sun X, Sun J, Ma X, et al. Structure and function of Mycobacterium tuberculosis EfpA as a lipid transporter and its inhibition by BRD-8000.3. Proc Natl Acad Sci U S A. 2024;121(44):e2412653121; doi: 10.1073/pnas.2412653121.

Figures

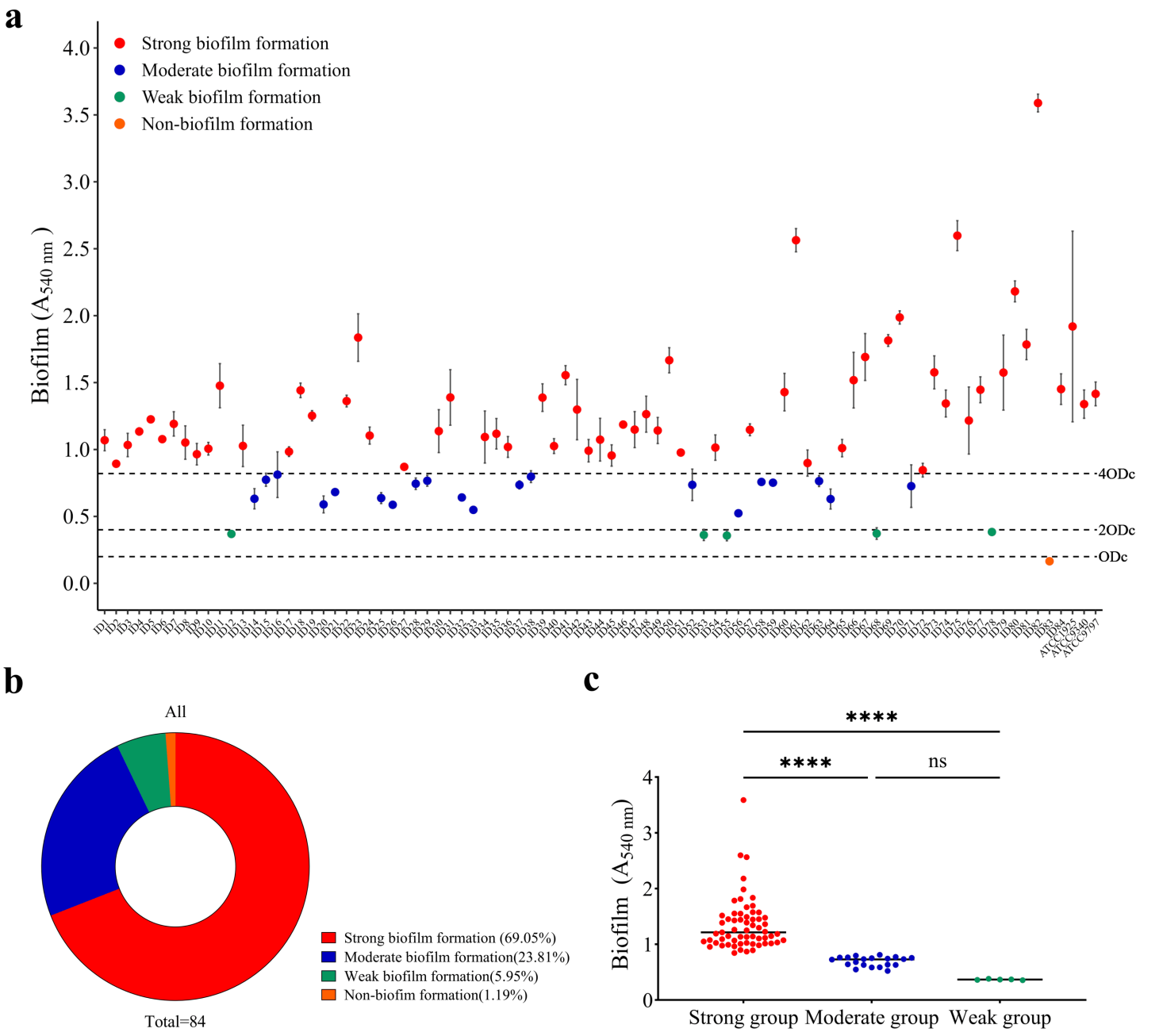


Figure 1

Quantification and phenotypic classification of biofilm formation capacity among *B. pertussis* isolates. (a) Biofilm production was quantified using the crystal violet staining method, and absorbance was measured at OD540nm. (b) Pie charts showed the proportion of strains classified as strong (red), moderate (blue), weak (green) or non-biofilm formation (orange). (c) Correlation analyses of biofilm formation among strong, moderate, and weak biofilm-forming strains. Each dot represents an individual isolate and horizontal bar indicate mean values. Statistical significance was determined using the two-way ANOVA followed by Dunn's multiple-comparison tests; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$, and ns, $P > 0.05$, respectively.

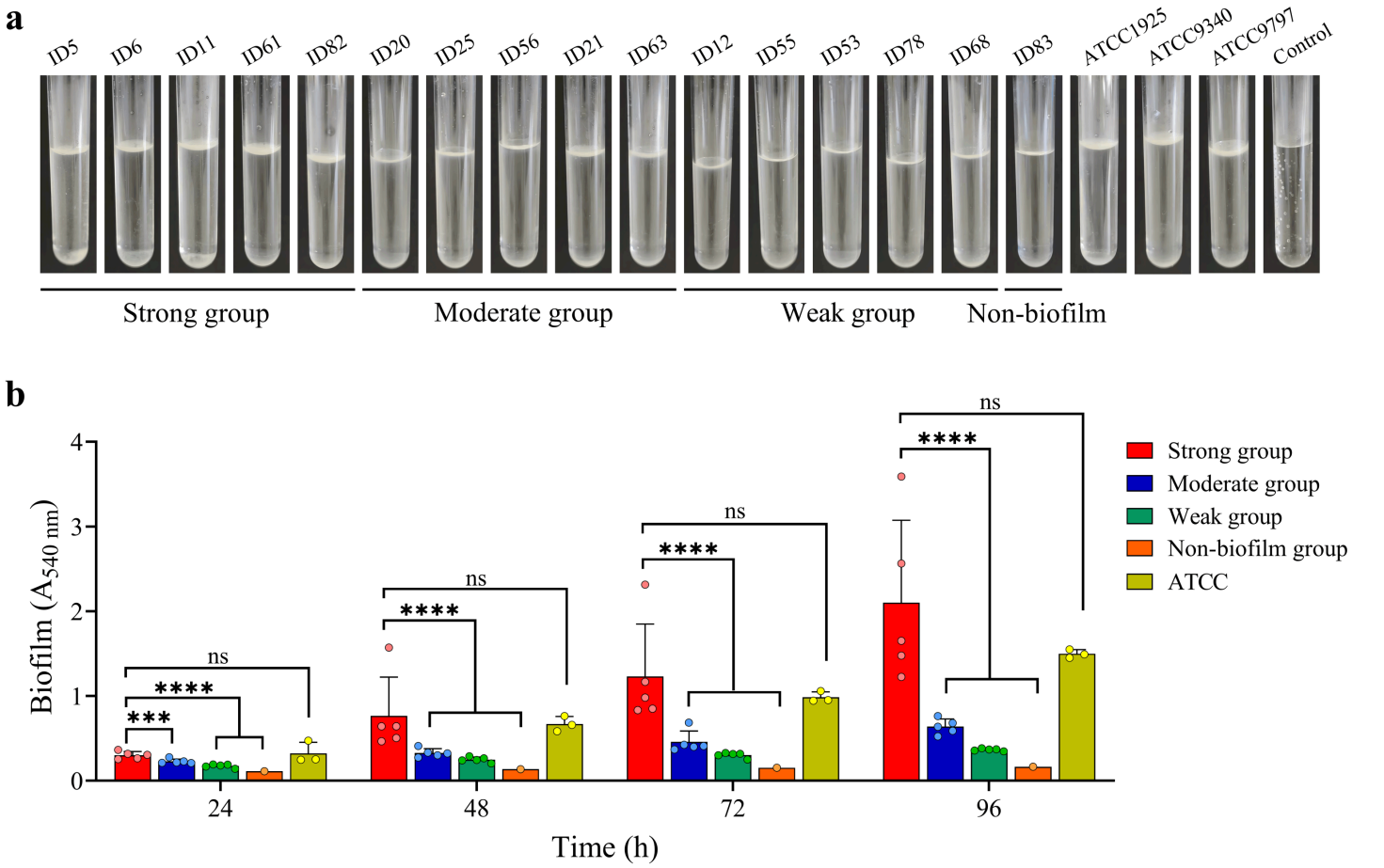


Figure 2

Biofilm formation and biomass accumulation in circulating *B. pertussis* strains. (a) Static biofilm formation of a bacterial ring at the air-liquid interface in glass culture tubes. (b) Quantification of biofilm biomass over time by crystal violet staining. Data are presented as the mean ± SD of three independent experiments, each performed in triplicate. Statistical significance was determined using two-way ANOVA test followed by Dunnett's test multiple comparisons. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$, and ns, $P > 0.05$, respectively.

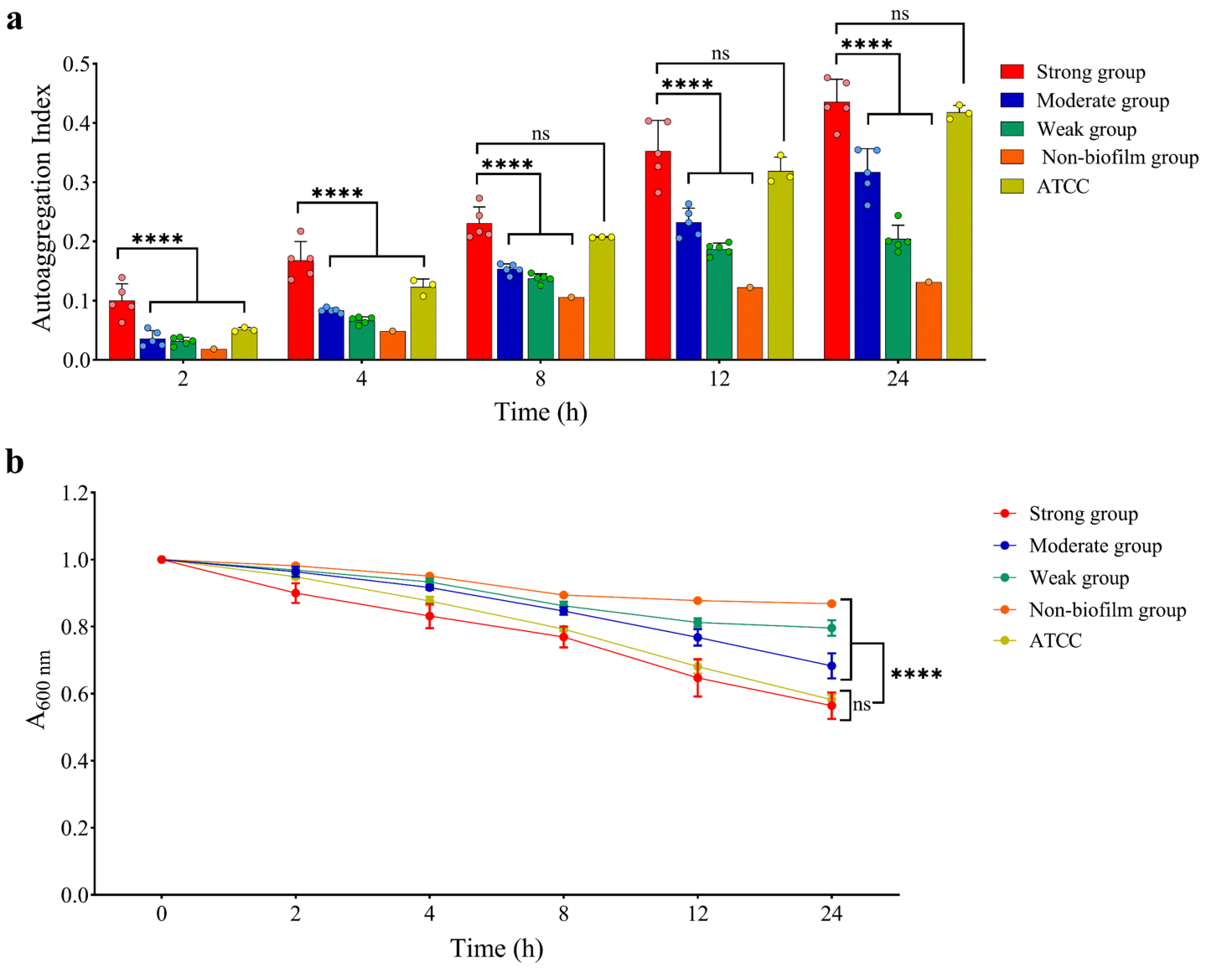


Figure 3

Quantification of autoagglutination of circulating *B. pertussis* strains. (a) Quantification of cell aggregation between different biofilm-forming groups and the ATCC reference strains. The x axis indicates incubation time (h), and the y axis indicates the aggregation index. (b) Autoaggregation kinetics of different biofilm-forming groups and the ATCC reference strains. The x-axis represents incubation time (h), and the y axis represents the OD600 of cells remaining in suspension. Each bar represents the mean ± SD from three independent experiments performed in duplicate. Statistical significance was determined using two-way ANOVA by Dunnett's multiple comparisons test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; and ****,$P < 0.0001$, respectively.

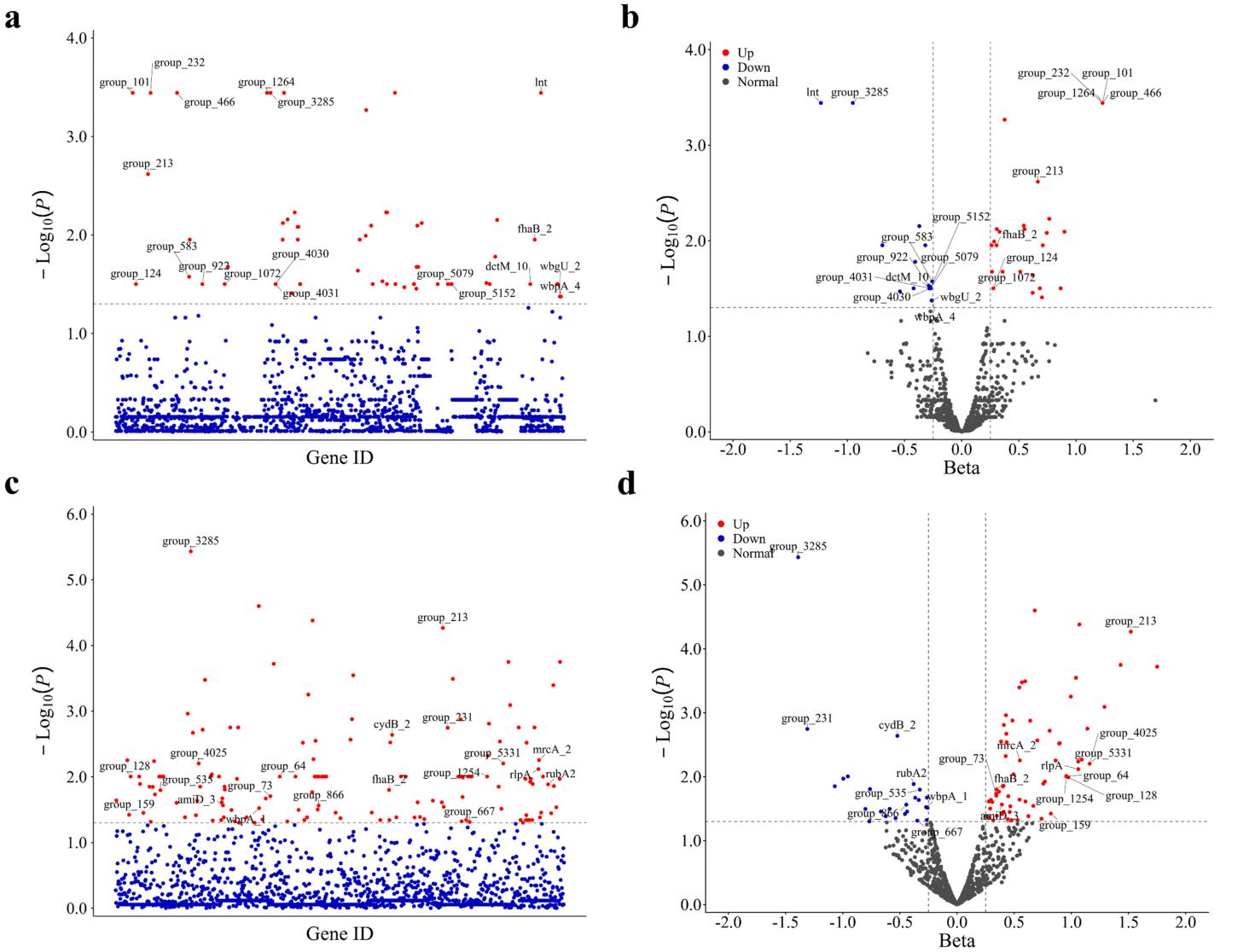


Figure 4

Pan-GWAS analysis for identifying biofilm-associated genes. Manhattan plots illustrating the statistical significance (-$\log_{10}P$) of gene associations identified by PLINK(a) and Pyseer (c). The x-axis represents virtual genomic positions arranged according to Roary IDs for visualization purposes. The horizontal dashed lines indicate the genome-wide significance threshold after Benjamini-Hochberg correction. Volcano plots depicting the relationship between statistical significance (-$\log_{10}P$) and effect size |β| for PLINK (b) and Pyseer (d). Candidate genes with $P < 0.05$ and absolute effect size |β|>0.25 are highlighted.

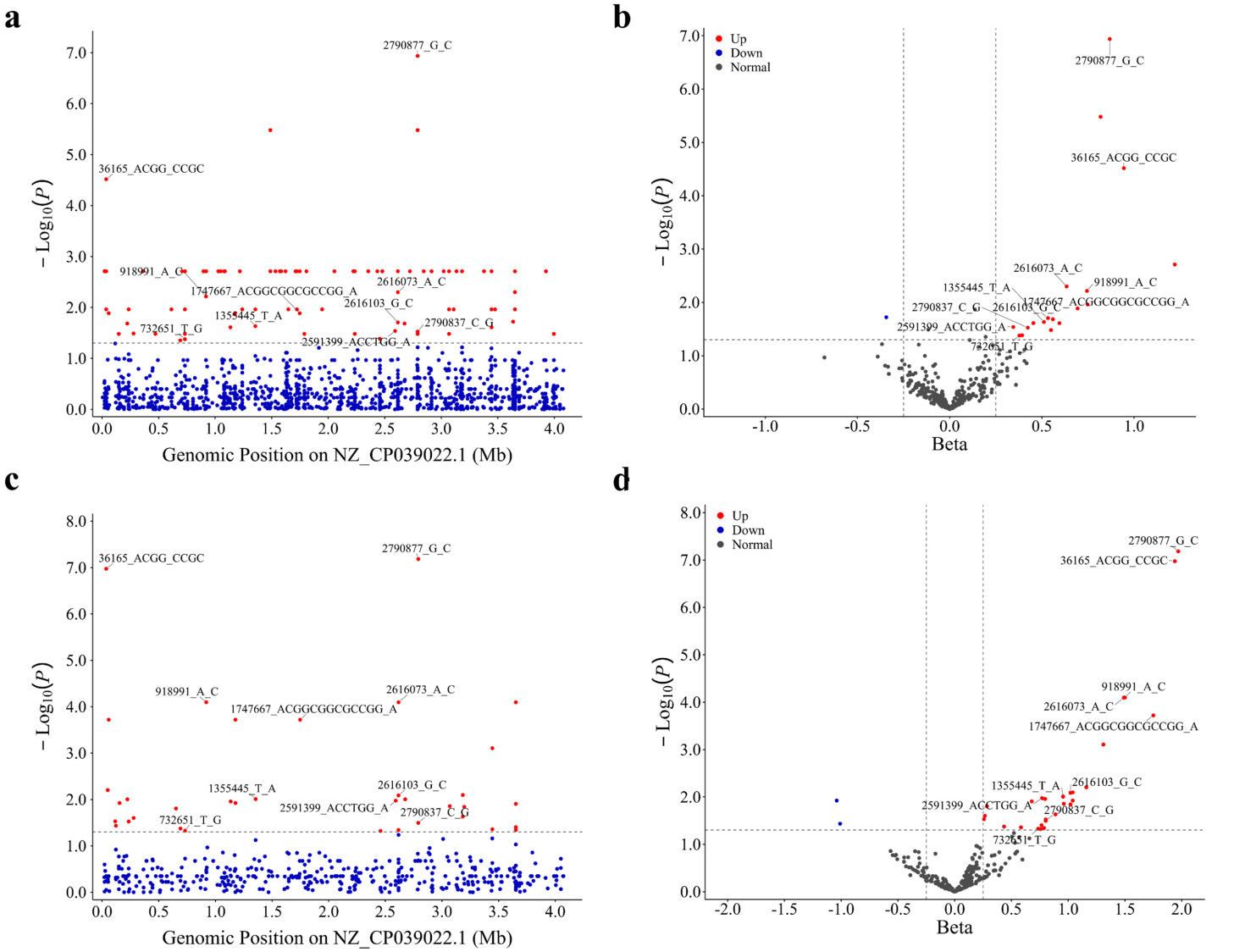


Figure 5

SNP-GWAS analysis of biofilm-associated loci using GEMMA and Pyseer. Manhattan plots of SNP association significance (-$\log_{10}P$) identified by GEMMA (a) and Pyseer (c). The x-axis indicates physical genomic positions relative to the Tohama I reference genome (NZ_CP039022.1). The horizontal dashed line indicates the significance threshold after Benjamini-Hochberg correction. Volcano plots showing the relationship between association significance (-$\log_{10}P$) and effect size |β| for GEMMA (b) and Pyseer (d), with significant SNPs highlighted ($P < 0.05$; |β| > 0.25).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfigure.docx

- Supplementarytable.xlsx