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.

Comparative genomics of the streptococcal mobilome reveals class-specific integration, mobility and antimicrobial resistance cargo

Kaiwei Nie
Sanya Institute of Nanjing Agricultural University
Xingyang Dai
Sanya Institute of Nanjing Agricultural University
Jiaqi Zhao
Sanya Institute of Nanjing Agricultural University
Mengjie Qi
Sanya Institute of Nanjing Agricultural University
Cai Yin
Ningxia Hui Autonomous Region Center for Animal Disease Control and Prevention
Xiaoming Wang
Nanjing Agricultural University
Liping Wang
Nanjing Agricultural University
Jinhu Huang
jhuang@njau.edu.cn

Sanya Institute of Nanjing Agricultural University

Research Article

Keywords: Streptococcus, mobilome, ICE, IME, prophage, plasmid, antimicrobial resistance

Posted Date: May 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9681180/v1

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

Additional Declarations: No competing interests reported.

Abstract

Mobile genetic elements (MGEs) drive genome plasticity, horizontal gene transfer, and antimicrobial resistance (AMR) dissemination in *Streptococcus*, yet genus-wide comparisons across major MGE classes remain limited. Here, we analyzed 1,961 complete genomes from 60 *Streptococcus* species and compared integrative and conjugative elements (ICEs), integrative and mobilizable elements (IMEs), prophages, and plasmids in terms of host distribution, integration-site preference, mobility-associated modules, representative backbones, and AMR cargo. We identified 1,172 ICEs, 2,362 IMEs, 3,397 prophages, and 225 plasmids, and found that the streptococcal mobilome was strongly partitioned by host lineage, with ICEs and IMEs enriched in *S. dysgalactiae*, prophages in *S. pyogenes*, and plasmids in *S. suis*. Integrated MGEs also displayed class-specific hotspot hierarchies: ICEs were highly concentrated at *rplL*, with secondary hotspots at *rlmD* and *rpmH*; IMEs were dominated by tRNA-associated sites together with *rpsI* and *rpmG*, whereas prophages occupied a broader hotspot spectrum centered on tRNA, *rpmE*, *rpsD*, *hlpA* and *mutL*. Mobility analyses further distinguished the classes, showing that ICEs retained a narrow repertoire of conjugative backbones dominated by typeFATA, IMEs displayed the broadest relaxase diversity, and plasmids were mainly non-mobile or mobilizable. Representative family analyses resolved recurrent backbone types within each class, whereas AMR cargo was concentrated in ICEs and plasmids and was further stratified by host species. Together, these findings reveal a previously underappreciated class- and host-dependent organization of the streptococcal mobilome and its AMR cargo.

Introduction

The genus *Streptococcus* encompasses a wide spectrum of species, including prominent human and animal pathogens such as *S. pneumoniae*, *S. pyogenes*, *S. agalactiae*, and *S. suis*, as well as commensal and food-associated members like the viridans group streptococci and *S. thermophilus*. This extensive broad ecological and pathogenic diversity makes the genus a valuable model for investigating bacterial genome plasticity, host adaptation mechanisms, and evolutionary dynamics across different ecological niches (Krzyściak et al., 2013; Gao et al., 2014; Lu et al., 2019; Salvadori et al., 2019). Horizontal gene transfer constitutes a principal mechanism driving host specialization, diversification of virulence factors, and the dissemination of antimicrobial resistance (AMR) within streptococci, thereby significantly contributing to their ecological and functional diversity (Lefébure and Stanhope, 2007; Choi et al., 2012; McNeilly and McMillan, 2014; Andam and Hanage, 2015; Chaguza et al., 2015).

Mobile genetic elements (MGEs) serve as the primary vehicles for these processes by facilitating the acquisition, excision, and transfer of DNA segments, as well as promoting genome rearrangements (Fariás and Espinosa, 2000; Lu et al., 2019; Ares-Arroyo et al., 2023). Among these, integrative and conjugative elements (ICEs) represent major self-transmissible chromosomal elements that frequently carry adaptive genes, including those conferring AMR (Wozniak and Waldor, 2010; Roberts and Mullany, 2011; Johnson and Grossman, 2015; Lu et al., 2019). Integrative and mobilizable elements (IMEs) form a more diverse group of integrated elements that lack a complete conjugation apparatus and instead rely

on transfer functions supplied *in trans*; they are increasingly recognized as widespread contributors to genome plasticity in streptococci (Coluzzi et al., 2017; Guédon et al., 2017; Lao et al., 2022). Prophages constitute another major component of the streptococcal mobilome, playing roles in lysogenic conversion, genome remodeling, and phage- mediated gene transfer (Scott et al., 2008; Rezaei Javan et al., 2019; Huang et al., 2023). Additionally, plasmids serve as extrachromosomal elements facilitating the dissemination of replication, mobilization, and AMR modules across strains and species (Guzmán and Espinosa, 1997; Smillie et al., 2010; Lorenzo-Díaz et al., 2014; Garcillán-Barcia et al., 2022). Collectively, ICEs, IMEs, prophages, and plasmids represent the principal vehicles of DNA mobility, whether integrated, phage-associated, or extrachromosomal elements, all capable of interbacterial transfer.

Previous studies have revealed that these elements are non-randomly distributed within the genome. ICEs and IMEs frequently integrate into specific chromosomal hotspots (Lorenzo-Díaz et al., 2016; Coluzzi et al., 2017; Lao et al., 2022). Prophages may occupy functionally important site such as *mutL* (Rezaei Javan et al., 2019), while conserved mobilizable plasmid backbones, like that of pMV158, can spread across distinct streptococcal hosts (Guzmán and Espinosa, 1997; Lorenzo-Díaz et al., 2014; Garcillán-Barcia et al., 2022). However, most prior studies have focused on individual MGE classes or on a limited number of streptococcal species. A comprehensive comparative framework encompassing ICEs, IMEs, prophages, and plasmids simultaneously remains absent (Lefébure and Stanhope, 2007; Garcillán-Barcia et al., 2009; Smillie et al., 2010; Andam and Hanage, 2015; Coluzzi et al., 2017; Guédon et al., 2017; Lu et al., 2019; Rezaei Javan et al., 2019; Lao et al., 2022).

In this study, we conducted a genus-wide comparative analysis of ICEs, IMEs, prophages, and plasmids using publicly available complete *Streptococcus* genomes. We investigated whether these MGE classes occupy distinct chromosomal niches, encode divergent repertoires of mobility-associated genes, and contribute differentially to the AMR dissemination. Our results provide a comparative framework that elucidates how different genetic elements differentially shape genome diversification and AMR spread across the streptococcal mobilome.

Materials and methods

Genome dataset collection

A *Streptococcus* genome dataset was retrieved from NCBI Assembly database (last accessed in July 2025). To minimize biases introduced by fragmented assemblies, only complete chromosomal sequences were included in chromosome-based screening analyses, whereas plasmid sequences were analyzed separately for plasmid-associated mobility and AMR features. The final dataset comprised 1,961 complete chromosomes and 225 plasmids, covering 60 *Streptococcus* species (Table S1 & S2).

Identification of ICEs, IMEs and prophages

The 1,961 complete chromosomal sequences were screened for ICEs and IMEs using ICEfinder 3.0 through the ICEberg 3.0 platform (https://tool2-mml.sjtu.edu.cn/ICEberg3/index.html) (Wang et al., 2024). This tool detects candidate chromosomally integrated elements based on the presence of characteristic mobility-associated genes, including genes of integrase, relaxase, type IV coupling protein, and type IV secretion system (Wang et al., 2024). Elements predicted to encode a complete conjugation module were classified as ICEs, whereas those lacking a complete conjugation apparatus but retaining mobilization-related genes were classified as IMEs (Johnson and Grossman, 2015; Lao et al., 2022). Using this workflow, we identified 1,172 ICEs and 2,362 IMEs.

Prophages were identified from the same set of complete chromosomal sequences using geNomad (v1.12.0) (https://github.com/apcamargo/genomad), which combines marker-based annotation and machine-learning-assisted sequence classification to detect viral regions and delimit integrated proviral segments (Camargo et al., 2024). This analysis identified 3,397 prophages across the dataset.

Definition of integration sites for ICEs, IMEs and prophages

Integration sites of ICEs, IMEs and prophages were inferred from the predicted integrase family together with BLASTn analysis of host chromosomal sequences flanking each element. For each putative MGE, the 3'- or 5'-end of a potentially interrupted coding sequence (CDS) or tRNA gene was used as the query sequence, and the corresponding left and right boundaries of the predicted element were inspected to identify the most likely insertion site. For downstream comparative analyses, all tRNA-associated insertion sites were grouped into a single tRNA category.

Identification of plasmid mobility features

All 225 plasmid sequences were analyzed using oriTfinder 2.0 to identify candidate origin-of-transfer (*oriT*) regions and plasmid mobility-associated genes (https://bioinfo-mml.sjtu.edu.cn/oriTDB2/oriTfinder.php) (Li et al., 2018). oriTfinder 2.0 predicts putative *oriT* sites by similarity to experimentally characterized *oriT* regions and by evaluating their association with transfer-related genes, including relaxase, type IV coupling protein, and type IV secretion system (Li et al., 2018). Based on these annotations, plasmids were classified into non-mobile, mobilizable, and conjugative categories (Smillie et al., 2010; Li et al., 2018). Plasmids lacking both an identifiable *oriT* region and transfer-related genes were considered non-mobile; plasmids carrying an *oriT* region and/or relaxase genes but lacking a complete conjugation apparatus were classified as mobilizable; and plasmids encoding a more complete set of conjugation-related genes, including type IV coupling protein genes and secretion system genes, were classified as conjugative. These annotations were subsequently used for comparative analyses of plasmid mobility composition and relaxase family distribution across *Streptococcus*.

Phylogenetic analysis of integrases

Integrases associated with different integration sites in streptococcal MGEs were aligned using ClustalW. Phylogenetic reconstruction was performed in MEGA v11.0 using the neighbor-joining and BioNJ algorithms based on pairwise distances estimated by the maximum composite likelihood method. The tree topology with the highest log-likelihood value was selected for downstream analysis.

Representative family-level genetic environment analysis

To characterize recurrent structural backbones within each MGE class, representative genetic environments were selected from the major integration-site- or backbone-associated families identified in the dataset. Representative elements were chosen based on a combination of family abundance, structural completeness, and backbone conservation. For ICEs and IMEs, representative families were defined primarily according to recurrent chromosomal insertion sites together with shared mobility-module organization. For prophages, representative families were defined based on recurrent integration sites and overall structural similarity. For plasmids, representative types were defined according to predicted mobility category, relaxase family, and backbone architecture. To avoid overrepresentation of rare or fragmented structures, only families represented by at least 3 elements and showing a conserved core gene organization were included in the representative comparison. When multiple candidates were available within the same family, the element with the most complete structure and the clearest gene organization was selected as the representative for visualization.

AMR cargo annotation and quantification

Protein-coding sequences predicted within ICEs, IMEs, prophages, and plasmids were screened against the MEGARes database using BLASTP (https://www.meglab.org/) (Doster et al., 2020). Hits were retained when the E-value was ≤ 1e-05 and amino acid identity was > 30%. For each mobile genetic element, antimicrobial resistance gene (ARG) annotations were extracted from the corresponding element-level protein annotation results and manually curated to remove obvious non-resistance homologues when necessary. ARG occurrence was then summarized at both the element-class and host-species levels. For class-level comparisons, the number and proportion of ARG-positive ICEs, IMEs, prophages, and plasmids were calculated. For species-level analyses, ARG-bearing elements were counted within each *Streptococcus* species and compared across MGE classes.

Results

Major MGE classes are unevenly distributed across the streptococcal mobilome

Across 1,961 complete Streptococcus chromosomes and 225 independently collected plasmids, we identified 1,172 ICEs, 2,362 IMEs, 3,397 prophages, and 225 plasmids. Prophages were the most abundant chromosomal MGE class, followed by IMEs and ICEs, whereas plasmids were treated as an

independent extrachromosomal dataset. These differences in abundance, host-species representation, size, and ARG-positive proportion indicate that the streptococcal mobilome is stratified rather than uniform (Fig. 1A; Table 1; Fig. S1; Tables S1–S5, S6A, and S6B).

Table 1
Summary of ICEs, IMEs, prophages, and plasmids identified in the complete *Streptococcus* genomes

MGE class	No. of elements	No. of host species	Mean size (bp)	ARG-positive elements	ARG-positive proportion[a]
ICE	1172	42	55102.51	527	45.0%
IME	2362	53	9248.94	26	1.1%
Plasmid	225	31	17164.14	47	20.90%
Prophage	3397	51	31656.55	88	2.6%

Abbreviations: ICE, integrative and conjugative element; IME, integrative and mobilizable element; ARG, antimicrobial resistance gene.

[a] ARG-positive proportion was calculated as the percentage of elements carrying at least one ARG within each MGE class.

All four MGE classes were widely distributed, but their accumulation patterns differed markedly among host species (Fig. 1A, B; Tables S6A and S6B). After normalizing by genome number, clear host-lineage biases became evident. *S. pyogenes*, the most heavily sampled species, was strongly prophage-dominated, with 3.16 prophages per genome but only 0.26 ICEs and 0.30 IMEs. In contrast, *S. dysgalactiae* showed the highest IME density (3.74 per genome) together with a high ICE density (1.41 per genome), indicating a strong enrichment of ICEs/IMEs in this lineage. A similarly pattern, though on a smaller scale, was observed in *S. anginosus* (1.53 ICEs, 3.67 IMEs, and 2.63 prophages per genome), suggesting that substantial MGE accumulation is not limited to the most intensively studies pathogens and may be underappreciated in other lineages. Among the major host species, *S. suis* showed a high chromosomal MGE burden, with 0.91 ICEs, 1.40 IMEs, and 2.41 prophages per genome (Table S6A). It also represented the principal plasmid-enriched lineage, contributing 73 of the 225 plasmids (32.4%) in the plasmid dataset (Tables S2 and S6B). By comparison, *S. parauberis* was particularly enriched in prophages (3.29 per genome), whereas *S. mutans* lacked detectable ICEs and was largely IME-dominated (1.03 IMEs per genome). Additional plasmid enrichment was observed in *S. thermophilus*, *S. dysgalactiae*, and *S. agalactiae*, whereas the strongest concentration remained in *S. suis*, which contributed 73 plasmids, corresponding to 32.4% of the plasmid dataset (Table S6B). Together, these results reveal marked host-lineage specialization in the accumulation of major MGE classes across the streptococcal mobilome.

The four MGE classes also differed markedly in size architecture (Fig. 1C; Table 1; Tables S2–S5). ICEs were the largest elements, with a mean size of 55.1 kb, whereas IMEs were much smaller, with a mean size of 9.2 kb. Prophages occupied an intermediate range, with a mean size of 31.7 kb, whereas

plasmids were concentrated in the small-size fraction, with a mean size of 17.1 kb. Notably, ICEs, IMEs and prophages each displayed bimodal density curves (Fig. 1C), suggesting additional structural subdivision within these classes. Thus, both species-level density contrasts and class-specific size architectures support the view that the streptococcal mobilome is a highly structured system in which different host species preferentially enrich different classes and structural modes of MGEs.

ICEs, IMEs and prophages exhibit quantitatively distinct hotspot hierarchies

We next examined chromosomal integration-site usage as a quantitative measure of how different classes of integrated MGEs interact with host chromosomes. Detailed integration-site assignments for all ICEs, IMEs and prophages are provided (Table S2-S5). Among MGEs with confidently assigned integration sites, integration sites were resolved for 941/1,172 ICEs (80.3%), 1,626/2,362 IMEs (68.8%) and 1,457/3,397 prophages (42.9%). The three MGE classes showed clearly distinct hotspot hierarchies (Fig. 2B–E). ICEs were strongly concentrated at a small number of sites, with *rplL* alone accounting for 58.7% (552/941) of all assigned ICE integration events, followed by *rlmD* (12.4%, 117/941), *rpmH* (7.3%, 69/941), *galM* (4.4%, 41/941) and *tRNA* sites (4.1%, 39/941) (Fig. 2B; Table S3). By contrast, IMEs displayed a broader spectrum dominated by *tRNA* sites (40.3%, 655/1,626), followed by *rpsI* (12.5%, 204/1,626), *rpmG* (11.4%, 185/1,626), *guaA* (5.8%, 95/1,626), *rpsU* (5.7%, 92/1,626) and *rpsN* (5.2%, 85/1,626) (Fig. 2C; Table S4). Prophages showed a third, distinct hierarchy, led by *tRNA* sites (13.3%, 194/1,457), *rpmE* (12.6%, 183/1,457), *rpsD* (12.4%, 180/1,457), *hlpA* (11.7%, 171/1,457) and *mutL* (6.0%, 88/1,457) (Fig. 2E; Table S5). Thus, ICEs, IMEs, and prophages do not follow a common integration logic, but instead occupy the chromosome through quantitatively distinct hotspot spectra.

Hotspot usage was also strongly host dependent (Fig. 2A, D). In *S. dysgalactiae*, ICEs were almost entirely confined to *rplL* (54.5%, 116/213) and *rlmD* (42.3%, 90/213), whereas IMEs in the same host were distributed much more broadly across *tRNA* sites (39.5%, 224/567), *rpmG* (17.8%, 101/567), *rpsN* (15.0%, 85/567) and *guaA* (12.0%, 68/567) (Fig. 2A). A similar contrast was evident in *S. pneumoniae*, where ICEs were again dominated by *rplL* (63.2%, 108/171), whereas IMEs were partitioned mainly between *tRNA* sites (44.6%, 45/101) and *rpsU* (39.6%, 40/101). In *S. suis*, ICEs remained centered at *rplL* (57.5%, 69/120), whereas IMEs were biased toward *tRNA* sites (52.4%, 109/208) with additional contributions from *rpsI* and other secondary sites (Table S1C). Prophages showed equally strong host dependence (Fig. 2D). In *S. pyogenes*, the largest prophage reservoir in the dataset, *hlpA* accounted for 24.5% (164/669) of assigned integration sites, followed by *tRNA* sites (16.7%, 112/669), *ssrA* (12.1%, 81/669) and *rpmE* (10.9%, 73/669). In *S. agalactiae*, prophage integration was dominated by *rpsD* (43.0%, 80/186), whereas in *S. pneumoniae* it was concentrated at *uvrA* (39.0%, 55/141) and *purA* (22.7%, 32/141).

Notably, several additional sites highlighted here, including *galM* for ICEs, *rpsN* for IMEs, and *rpmE*, *rpsD*, *hlpA*, *uvrA*, and *purA* for prophages, extend the set of genus-level hotspots beyond the best-characterized integration sites reported previously. Consistent with these hotspot patterns, the integrase phylogeny

further showed that representatives sharing the same dominant integration category often clustered into related branches, supporting a linkage between hotspot preference and integrase lineage (Fig. 2F). Together, these results indicate that chromosomal hotspot usage across the streptococcal mobilome is shaped jointly by MGE class, host background, and integrase-associated integration specificity.

ICEs, IMEs and plasmids encode distinct mobility strategies

ICEs, IMEs and plasmids represent the major conjugation-related MGE classes, and comparison of their mobility-associated modules, including mating pair formation (MPF) architectures and relaxase repertoires, revealed clearly distinct transfer strategies (Garcillán-Barcia et al., 2009; Li et al., 2018; Smillie et al., 2010). In streptococci, only two MPF architectures were detected in ICEs, namely typeFATA and typeFA, with the former clearly predominating overall (71.8%, 841/1,172) than typeFA (28.2%, 331/1,172) (Fig. 3A). This predominance was particularly marked in *S. dysgalactiae, S. suis*, and *S. thermophilus*, whereas *S. pneumoniae* showed a relative enrichment of typeFA. Plasmids displayed a contrasting pattern: most were non-mobile (60.0%, 135/225), 37.8% (85/225) were mobilizable (pMOB), and only 2.2% (5/225) were predicted to be conjugative (pCONJ) (Fig. 3B). These results indicate that streptococcal ICEs retain a restricted but conserved set of conjugative backbones, whereas plasmids are dominated by non-self-transmissible and mobilizable forms.

Relaxase profiles further separated the three MGE classes (Fig. 3C–D). Across the top 15 streptococcal species with the highest representation of conjugation-related MGEs, the overall relaxase landscape was dominated by MOBP1, *firmi_Rep_2*, MOBT, *profile_MOBT* and MOBV, but the class-specific distributions differed markedly. ICEs were mainly associated with MOBP1 and *profile_MOBT*, indicating a comparatively restricted relaxase repertoire. By contrast, IMEs showed the broadest and most heterogeneous spectrum, with *firmi_Rep_2*, MOBT, MOBV, *firmi_Viral_Rep_B1*, *PHA_IME_B*, *PHA_IME_A1* and MOBQ all making substantial contributions, consistent with their modular diversity and reliance on transfer functions supplied in trans. Plasmids were simpler and more polarized, being dominated by MOBV, with smaller contributions from MOBT, MOBP1 and MOBQ. Species-level heatmaps supported the same pattern (Fig. 3D): ICE repertoires were usually dominated by one or two relaxase families, IMEs varied much more strongly among species, and plasmids were largely centered on MOBV-like backbones. Together, these results show that ICEs, IMEs and plasmids do not encode equivalent mobility strategies: ICEs are linked to a narrow set of conjugative modules, IMEs possess the most diverse mobilization repertoire, and plasmids are largely non-mobile or mobilizable, with only a minor conjugative fraction.

Representative MGE families retain conserved backbones across species

Representative family analyses showed that each major MGE class in *Streptococcus* contains recurrent backbone types with conserved modular organization across hosts (Fig. 4). Among ICEs, the dominant large-type architecture was represented by the *rplL*-associated ICESa2603-like family (Fig. 4A),

consistent with the major ICE hotspot, because *rplL*-associated ICEs accounted for 47.1% (552/1,172) of all ICEs. These elements retained a conserved backbone centered on integration, conjugation, and maintenance modules, with most variation confined to accessory regions. By contrast, the second major ICE type was represented by a smaller Tn*916*-like family (Fig. 4B). These Tn916-like ICEs, typically carrying typeFA/profile_MOBT modules, accounted for 16.1% (189/1,172) of all ICEs under the stringent size-based definition and 24.0% (281/1,172) under the broader family-level classification, consistent with the smaller ICE size mode identified in the overall size distribution (Fig. 1C).

Among IMEs, two recurrent families were selected as representative examples. The *rpsI*-associated family accounted for 8.6% (204/2,362) of all IMEs and retained a conserved mobilization-centered backbone across multiple host species (Fig. 4C). By contrast, the *rpsN*-associated family accounted for 3.6% (85/2,362) of all IMEs and was detected exclusively in *S. dysgalactiae*, highlighting *rpsN* as a recurrent integration hotspot for a highly conserved, host-restricted IME lineage in this dataset (Fig. 4D). All identified *rpsN*-associated IMEs displayed the same gene organization, further indicating that this lineage is highly conserved despite its restricted host distribution.

Plasmids were grouped into mobilizable, conjugative, and other types, with the pMV158-like mobilizable family and the conjugative backbones selected as representative structures for comparison (Fig. 4E, F). The pMV158-like MOBV-bearing family represented the dominant mobilizable plasmid backbone and accounted for 23.1% (52/225) of all plasmids, with a conserved mobilization-centered organization across host species (Fig. 4E). S. suis was the major plasmid reservoir in the dataset, carrying 32.4% (73/225) of all plasmids (Fig. 1B). The two conjugative plasmid backbones represented the rare pCONJ fraction, which comprised 2.2% (5/225) of all plasmids (Fig. 4F). These backbones were identified in *Streptococcus* and in other genera, indicating a broader taxonomic distribution (Fig. 4F). A further analysis was conducted to examine the structural relationship among the S. suis plasmid pSH0918, related plasmids, and the ICEs (Fig. S2). Our results show that although pSH0918 lacks a complete conjugation system and the relaxase gene, it was still capable of transferring from the donor strain SH0918 to the recipient strain YZDH1, and subsequently from YZDH1 to additional recipient cells. This mobility is attributable to the presence of functionally active ICEs in both strains and the high similarity of pSH0918 sequences to the ICE *oriT* region. Thus, such plasmids retain the potential for horizontal transfer capacity and constitute a plausible route for resistance dissemination in *S. suis*.

Prophages were broadly classified here into SMphage-like elements and other prophage types, with one representative backbone shown for each category (Fig. 4G, H). The *rlmD*-associated family accounted for 2.4% (80/3,397) of all prophages and retained a conserved phage-specific organization across host species (Fig. 4G). The *rpsD*-associated family accounted for 5.3% (180/3,397) of all prophages and represented a second recurrent prophage backbone type (Fig. 4H). Together, these comparisons show that the major class-level patterns described above are underpinned by recurrent family-level backbones, while distinct subtypes within each MGE class differ in abundance and host distribution (Fig. 4). The additional plasmid observations in *S. suis* further extend the significance of this framework by suggesting that, beyond chromosomally integrated MGEs, plasmid-associated mobilization may also

contribute substantially to resistance-gene circulation in streptococci, especially in livestock-associated lineages.

AMR cargo is concentrated in ICEs and plasmids and stratified by host species

We next examined ARG burden at both the MGE and species levels to determine whether ARG-bearing MGEs were distributed uniformly across *Streptococcus* or instead concentrated in particular host backgrounds (Fig. 5). ARG carriage differed sharply among the four major MGE classes. ICEs showed the highest ARG burden, with 45.0% (527/1,172) of elements carrying at least one ARG, followed by plasmids at 20.9% (47/225). In contrast, ARGs were much less frequent in prophages (2.6%, 88/3,397) and rare in IMEs (1.1%, 26/2,362) (Fig. 5A). Across all four MGE classes, we detected 1,183 ARG–MGEs associations involving 48 distinct ARGs. ICEs carried by far the largest overall burden, with 965 ARG occurrences distributed across 527 ARG-bearing ICEs, compared with 111 occurrences in prophages, 57 in plasmids, and 50 in IMEs. The mean ARG burden per ARG-bearing element was 1.83 in ICEs, 1.92 in IMEs, 1.26 in prophages, and 1.21 in plasmids, whereas the burden normalized to all elements remained highest in ICEs (0.82 ARGs per element), followed by plasmids (0.25), prophages (0.03), and IMEs (0.02). The highest ARG load was observed in a single ICE from *S. pluranimalium* SP28, which carried 9 ARGs [*tet*(O), *erm*(B), *dfrG*, *lsa*(E), *lnu*(B), *ant(6)-Ia*, *aac(6')-aph(2")*, *tet*(L) and *cat*], whereas the most heavily loaded IME, prophage, and plasmid carried 7, 4, and 3 ARGs, respectively. These results indicate that ARG-bearing MGEs differ not only in frequency, but also in their average cargo load and upper limits of ARG accumulation.

ARG composition was strongly structured by MGE class and host species (Fig. 5B-E). ICEs mainly carried tetracycline- and MLS-associated resistance genes, particularly *tet*(M), *erm*(B), and *tet*(O). These genes were detected in 26.6% (312/1,172), 19.1% (224/1,172), and 11.0% (129/1,172) of all ICEs, respectively, corresponding to 59.2%, 42.5%, and 24.5% of ARG-bearing ICEs. At the class level, tetracycline genes were present in 458 ICEs (39.1% of all ICEs; 86.9% of ARG-bearing ICEs) and MLS/efflux-associated genes in 284 ICEs (24.2%; 53.9%). IMEs carried far fewer ARGs overall, but when present they were concentrated in a narrow subset of determinants, particularly *erm*(A) and *optrA* (10 IMEs each; 38.5% of ARG-bearing IMEs), followed by *erm*(B) (5 IMEs; 19.2%) and *tet*(O) / *tet*(S) (4 IMEs each; 15.4%). Consistent with this, MLS/efflux-associated genes occurred in 17 IMEs (65.4% of ARG-bearing IMEs) and oxazolidinone-related genes in 10 IMEs (38.5%). Prophages were dominated by tetracycline and MLS/efflux-associated genes, with *tet*(S) as the single most frequent determinant (28 prophages; 0.8% of all prophages; 28.0% of ARG-bearing prophages), and *msr*(D), *tetA*(46) and *tet*(O) (10 prophages each; 10.0%). Plasmids showed a different pattern, being enriched mainly for *tet*(O) (14 plasmids; 6.2% of all plasmids; 29.8% of ARG-bearing plasmids), *erm*(B) (10 plasmids; 21.3% of ARG-bearing plasmids) and *erm*(T) (8 plasmids; 17.0%), while MLS/efflux-associated genes overall were detected in 26 plasmids (11.6% of all plasmids; 55.3% of ARG-bearing plasmids). Species-level heatmaps further showed that these ARGs were highly host stratified (Fig. 5E). In ICEs, *tet*(M) and *erm*(B) were particularly concentrated in *S. salivarius* (70.0% each), while *tet*(M) also reached 61.3% in *S. pneumoniae*; by contrast, *S. suis* ICEs

were distinguished by a strong *tet*(O) signal (48.0%). In IMEs, the strongest signals were restricted to a few hosts, including *S. pasteurianus* for *erm*(B) (8.3%) and *tet*(O) / *tet*(L) (4.2% each), *S. parauberis* for *tet*(S) (3.4%) and *S. suis* for *erm*(A) (3.4%). Prophage-associated ARGs were again concentrated in a limited set of hosts, with *msr*(D) reaching 14.3% in *S. pluranimalium*, *erm*(B) reaching 4.5% in *S. parasuis*, and *tet*(S) reaching 2.9% in *S. dysgalactiae*. Plasmid-associated ARGs were most evident in *S. mutans* for *erm*(B) (33.3%), *S. iniae* for *erm*(B) (25.0%), and *S. suis* for *tet*(O) (17.8%) and *erm*(B) (9.6%). These results, together with the class-biased enrichment patterns, indicate that ARG distribution across MGEs cannot be explained by host taxonomy alone (Fig. 5D). Instead, different host-associated ecological or therapeutic settings likely enrich different ARG-bearing element types, suggesting that host species mainly provide the ecological background, whereas antimicrobial selection drives the contrasting distribution of resistance-associated MGEs across the streptococcal mobilome.

Discussion

The streptococcal mobilome is often viewed as a general pool of MGEs that facilitate horizontal gene transfer (Wozniak and Waldor, 2010; Johnson and Grossman, 2015; Guédon et al., 2017; Lu et al., 2019). However, our findings reveal that different classes of MGEs (ICEs, IMEs, prophages, and plasmids) exhibit marked differences in host distribution, integration site preferences, mobility-associated modules, backbone organization and ARG content. These differences indicate that the evolutionary impact of a given MGE depends not just on its presence, but on its structural classification and the host species it inhabits. Therefore, lumping diverse MGEs into a single mobilome pool may obscure important biological distinctions and hinder accurate predictions of resistance or virulence spread (Coluzzi et al., 2017; Huang et al., 2023).

A prominent pattern emerging from our analysis is the uneven distribution of MGE classes across streptococcal lineages (Lefébure and Stanhope, 2007; Choi et al., 2012; Chaguza et al., 2015). Specifically, *S. pyogenes* is predominantly characterized by prophage abundance, *S. dysgalactiae* exhibits the highest densities of ICEs and IMEs, while *S. suis* serves as the primary reservoir for plasmids. The marked contrast between *S. pyogenes* and *S. dysgalactiae* is particularly given that both species are human pathogens occupying similar ecological niches. This disparity implies that host-specific factors, such as CRISPR-Cas system repertoires, phage receptor availability, or recombination machinery efficiency, may influence the acquisition of MGEs. In the case of *S. suis*, the unusually high plasmid burden may be related to its livestock-associated ecology, where frequent exposure to antimicrobial exerts selective pressure favoring plasmid-encoded resistance determinants (Lefébure and Stanhope, 2007; Andam and Hanage, 2015; Chaguza et al., 2015). This hypothesis is corroborated by the pronounced presence of the *tet*(O) gene within ICEs and plasmids of *S. suis* (Dai et al., 2023).

Chromosomal integration of MGEs in *Streptococcus* species exhibits a strongly non-random distribution (Coluzzi et al., 2017; Lao et al., 2022). ICEs, IMEs and prophages tend to accumulate at recurrent genomic hotspots, but their preferred integration sites differ markedly (Lorenzo-Díaz et al., 2016; Coluzzi et al., 2017; Rezaei Javan et al., 2019). ICEs are highly concentrated at *rplL*, a ribosomal protein gene that

may be represent a permissive integration site due to either the minimal fitness cost associated with its disruption or its transcriptional accessibility. IMEs primarily integrate at *tRNA*-associated sites but also use a broader set of sites, including *rpsI*, *rpmG*, *guaA*, and *rpsU*, reflecting their dependence on diverse integrase families. Prophages showed a distinct integration pattern characterized by a wider hotspot spectrum, includes *tRNA* genes as well as *rpmE*, *rpsD*, *hlpA*, and *mutL*. Notably, several sites identified here, *galM* for ICEs, *rpsN* for IMEs, and *rpmE*, *rpsD*, *hlpA*, *ssrA*, *uvrA* and *purA* for prophages, have not been previously recognized as major integration sites in streptococci, thereby expanding the current understanding of the genomic landscape of MGE integration hotspots. Furthermore, the utilization of these hotspots exhibits pronounced host specificity. For instance, prophages in *S. pyogenes* preferentially integrate at *hlpA*, whereas in *S. agalactiae*, *rpsD* is favored, and in *S. pneumoniae*, integration is concentrated at *uvrA* and *purA*. These observations suggest that the local genomic context plays a critical role in modulating hotspot accessibility and integration patterns across different streptococcal species (Coluzzi et al., 2017; Rezaei Javan et al., 2019).

Mobility analyses provide additional evidence supporting this differentiation (Smillie et al., 2010; Li et al., 2018). The three classes of conjugation-related MGEs exhibit different transfer strategies. ICEs in streptococci use only two MPF architectures, typeFA and typeFATA, with typeFATA being predominantly prevalent (71.8%). This limited repertoire implies that ICEs have evolved a small number of highly optimized self-transmissible backbones. In contrast, IMEs displayed the greatest diversity in relaxase types (Coluzzi et al., 2017; Guédon et al., 2017; Lao et al., 2022), including firmi_Rep_2, MOBT, MOBV, and several IME-specific families. This diversity is consistent with the fact that IMEs lack their own conjugation machinery and rely on mobilization *in trans*; thus, a broad relaxase spectrum may enable them to exploit a wider range of helper systems. Plasmids exhibit a third distinct pattern: the majority are either non-mobile or mobilizable, with only a minimal proportion (2.2%) predicted to be conjugative. Consequently, plasmids in Streptococcus primarily function as compact, mobilizable cargo platforms rather than as principal self-transfer agents (Lorenzo-Díaz et al., 2014; Garcillán-Barcia et al., 2022).

These class level distinctions are underpinned by a limited set of recurrent backbone architectures. Within ICEs, the large *rplL*-associated ICE*Sa*2603 family and the more compact Tn*916* family explain the observed bimodal size distribution seen in the results (Roberts and Mullany, 2011; Ciric et al., 2012; Santoro et al., 2014). Both families maintain a conserved core comprising genes involved in integration, conjugation and maintenance; however, the ICE*Sa*2603 family ICEs contain multiple variable regions capable of accommodating additional cargo genes, whereas the Tn*916* family ICEs exhibit a smaller and more streamlined structure. Among IMEs, the widely distributed *rpsI*-associated family contrasts with the newly recognized *rpsN*-associated lineage, which is found only in *S. dysgalactiae* and shows identical gene organization across isolates, suggesting a recent local expansion rather than an ancient, broadly disseminated lineage (Wang et al., 2022; López de Egea et al., 2023). Plasmids also separate into multiple recurrent types, including the dominant pMV158-like MOBV-bearing architecture and rarer conjugative forms that may facilitate intergeneric transfer (Guzmán and Espinosa, 1997; Smillie et al., 2010; Garcillán-Barcia et al., 2022). Notably, *S. suis* represented the major plasmid reservoir in the dataset, suggesting that plasmid-associated transfer may play a disproportionately important role in this

species. Prophages showed the greatest structural heterogeneity but nonetheless contain recurrent hotspot-associated backbones, such as those linked to *rlmD*- and *rpsD*-associated families, indicating that even highly variable elements preserve conserved organizational frameworks (Huang et al., 2023). Collectively, these findings suggest that the streptococcal mobilome is structured around a finite set of recurring backbone architectures that differ in their dissemination range, hotspot preferences, and structural plasticity.

ARGs are distributed unevenly across MGEs, revealing distinct patterns (Lu et al., 2019). ICEs carry the highest proportion of ARGs, accounting for 45% of ARG-bearing elements, followed by plasmids (21%). In contrast, only 2.6% of prophages and 1.1% of IMEs contain ARGs. This distribution aligns with the relative cargo capacity and autonomous mobility of these MGEs. ICEs, characterized by their large size, stability, and self-transmissibility, serve as optimal platforms for the accumulation of multiple resistance determinants (Wang et al., 2022; Yang et al., 2023). Plasmids, although smaller, are mobilizable and their extrachromosomal nature facilitates rapid acquisition and loss of ARGs under selective pressure (Santoro et al., 2014; Yang et al., 2023; Coccitto et al., 2024). IMEs, despite being mobilizable, exhibit limited compatibility with most resistance genes, with the notable exceptions of *erm*(A) and *optrA*, which are disproportionately represented in IMEs, suggesting a specific co-adaptation between these resistance genes and IME backbones (Wang et al., 2022; Dai et al., 2023). The involvement of prophages in ARG transmission has been documented recently, particularly concerning the SMphage family, which is prevalent in *S. suis* and *S. agalactiae* (Huang et al., 2023). Nevertheless, given the extensive diversity of prophages within bacterial genomes, their role in ARG dissemination remains relatively minor and is overshadowed by the more significant contribution of chromosomal ICE-mediated transfer. Plasmids are predominantly enriched for *tet*(O), *erm*(B), and *erm*(T) genes, underscoring their role as compact yet effective vectors of resistance (Guzmán and Espinosa, 1997; Smillie et al., 2010; Garcillán-Barcia et al., 2022). These class-specific cargo profiles suggest that the spread of ARGs is constrained not only by opportunities for horizontal transfer but also by the compatibility between specific resistance determinants and particular MGE backbones.

At the species level, ARG distribution exhibits marked stratification. For instance, *tet*(M) and *erm*(B) are highly prevalent in ICEs from *S. salivarius* and *S. pneumoniae*, whereas ICEs from *S. suis* ICEs are characterized by a pronounced presence of *tet*(O). This pattern is unlikely to be explained by host taxonomy alone. A more plausible explanation is that antimicrobial selection pressures across host species drive these differences (Lefébure and Stanhope, 2007; McNeilly and McMillan, 2014; Andam and Hanage, 2015; Chaguza et al., 2015). In livestock-associated *S. suis*, the use of tetracyclines in animal husbandry may select for *tet*(O)-carrying elements. Conversely, in human respiratory pathogens, clinical use of macrolides and tetracyclines likely promotes the prevalence of *tet*(M) and *erm*(B) within ICEs. A similar rationale applies to plasmids, which harbor distinct ARG repertoires in *S. mutans*, *S. iniae*, and *S. suis*. These observations suggest that ecological factors and antibiotic exposure, rather than taxonomy alone, shape the resistance gene landscape.

Conclusion

In summary, the streptococcal mobilome can be conceptualized as a complex network of backbone types, specific insertion sites, and ecological selection pressures. ICEs represent the primary chromosomal resistance platforms, while IMEs exhibit considerable structural diversity but generally harbor few ARGs. Plasmids function as compact mobilizable backbones with occasional capacity for wider dissemination, and prophages contribute to genome diversification through structural heterogeneity and the carriage of selected resistance determinants. This conceptual framework provides a basis for understanding the interplay among MGEs, host genomic context, and antimicrobial selective pressures in driving the evolutionary dynamics of *Streptococcus* species.

Declarations

Funding Declaration

This work was supported by the National Key R&D Program of China (2023YFD1301800), the National Natural Science Foundation of China (32573429), the Natural Science Foundation of Ningxia Hui Autonomous Region (2024AAC05098), the Fundamental Research Funds for the Central Universities (KYCXJC2025001), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD)

Data availability

All genome sequences analyzed in this study are publicly available from the NCBI Assembly/GenBank databases, and the corresponding accession numbers are listed in the supplementary materials (Table S1). The genome- and element-level summary datasets generated in this study are provided in the supplementary materials, including element-level annotations for plasmids, ICEs, IMEs, and prophages, as well as species-level summaries of chromosomal MGE density and plasmid distribution (Tables S1–S5, S6A, and S6B). Additional data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Kaiwei Nie (Writing–original draft, Investigation, Formal analysis, Data curation). Xingyang Dai (Writing–original draft, Formal analysis, Data curation). Jiaqi Zhao (Data curation). Mengjie Qi (Investigation). Cai Yin (Investigation). Xiaoming Wang (Validation). Liping Wang (Validation). Jinhu Huang (Writing – review & editing, Methodology, Funding acquisition, Data curation).

References

1. Andam CPHanage WP. Mechanisms of genome evolution of *Streptococcus*. Infect Genet Evol 2015;33:334-342. https://doi.org/10.1016/j.meegid.2014.11.007
2. Ares-Arroyo M, Coluzzi C, Rocha EPC. Origins of transfer establish networks of functional interactions among mobile genetic elements. Nucleic Acids Res 2023;51(7):3001-3016. https://doi.org/10.1093/nar/gkac1079
3. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC. Identification of mobile genetic elements with geNomad. Nat Biotechnol 2024;42(8):1303-1312. https://doi.org/10.1038/s41587-023-01953-y
4. Chaguza C, Cornick JE, Everett DB. Mechanisms and impact of genetic recombination in the evolution of *Streptococcus* pneumoniae. Comput Struct Biotechnol J 2015;13:241-247. https://doi.org/10.1016/j.csbj.2015.03.007
5. Choi SC, Rasmussen MD, Hubisz MJ, Gronau I, Stanhope MJ, Siepel A. Replacing and additive horizontal gene transfer in *Streptococcus*. Mol Biol Evol 2012;29(11):3309-3320. https://doi.org/10.1093/molbev/mss138
6. Ciric L, Ellatif M, Sharma P, Patel R, Song X, Mullany P, Roberts AP. Tn916-like elements from human, oral, commensal streptococci possess a variety of antibiotic and antiseptic resistance genes. Int J Antimicrob Agents 2012;39(4):360-361. https://doi.org/10.1016/j.ijantimicag.2011.12.007
7. Coccitto SN, Massacci FR, Cinthi M, Albini E, Cucco L, Paniccià M. Novel optrA-carrying integrative and conjugative element in a porcine *Streptococcus* suis, Italy. J Antimicrob Chemother 2024;79(12):3356-3358. https://doi.org/10.1093/jac/dkae354
8. Coluzzi C, Guédon G, Devignes MD, Ambroset C, Loux V, Lacroix T. A glimpse into the world of integrative and mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. Front Microbiol 2017;8:443. https://doi.org/10.3389/fmicb.2017.00443
9. Dai X, Zhao J, Sun J. ICESpsuAH0906, a novel optrA-carrying element conferring resistance to phenicols and oxazolidinones from *Streptococcus parasuis*, is transferable to *Streptococcus suis*. Vet Microbiol 2023;283:109795. https://doi.org/10.1016/j.vetmic.2023.109795
10. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, Belk KE, Noyes NR, Morley PS. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res 2020;48(D1):D561-d569. https://doi.org/10.1093/nar/gkz1010
11. Fariás MAEEspinosa M. Conjugal transfer of plasmid pMV158: uncoupling of the pMV158 origin of transfer from the mobilization gene mobM, and modulation of pMV158 transfer in Escherichia coli mediated by IncP plasmids. Microbiology 2000;146(Pt 9):2259-2265. https://doi.org/10.1099/00221287-146-9-2259
12. Gao XY, Zhi XY, Li HW, Klenk HP, Li WJ. Comparative genomics of the bacterial genus *Streptococcus* illuminates evolutionary implications of species groups. PLoS One 2014;9(6):e101229. https://doi.org/10.1371/journal.pone.0101229

13. Garcillán-Barcia MP, Francia MV, de la Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev 2009;33(3):657-687. https://doi.org/10.1111/j.1574-6976.2009.00168.x
14. Garcillán-Barcia MP, de la Cruz F, Espinosa M. The facts and family secrets of plasmids that replicate via the rolling-circle mechanism. Microbiol Mol Biol Rev 2022;86(1):e00222-20. https://doi.org/10.1128/MMBR.00222-20
15. Guédon G, Libante V, Coluzzi C, Payot S, Leblond-Bourget N. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. Genes 2017;8(11):337. https://doi.org/10.3390/genes8110337
16. Guzmán LMEspinosa M. The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid oriT. J Mol Biol 1997;266(4):688-702. https://doi.org/10.1006/jmbi.1996.0824
17. Huang J, Dai X, Wu Z. Conjugative transfer of streptococcal prophages harboring antibiotic resistance and virulence genes. ISME J 2023;17(9):1467-1481. https://doi.org/10.1038/s41396-023-01463-4
18. Johnson CMGrossman AD. Integrative and conjugative elements (ICEs): what they do and how they work. Annu Rev Genet 2015;49:577-601. https://doi.org/10.1146/annurev-genet-112414-055018
19. Krzyściak W, Pluskwa KK, Jurczak A, Kościelniak D. The pathogenicity of the *Streptococcus* genus. Eur J Clin Microbiol Infect Dis 2013;32(11):1361-1376. https://doi.org/10.1007/s10096-013-1914-9
20. Lao J, Lacroix T, Guédon G. ICEscreen: a tool to detect Firmicute ICEs and IMEs, isolated or enclosed in composite structures. NAR Genom Bioinform 2022;4(4):lqac079. https://doi.org/10.1093/nargab/lqac079
21. Lefébure TStanhope MJ. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol 2007;8(5):R71. https://doi.org/10.1186/gb-2007-8-5-r71
22. Li X, Xie Y, Liu M, Tai C, Sun J, Deng Z, Ou HY. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. Nucleic Acids Res 2018;46(W1):W229-W234. https://doi.org/10.1093/nar/gky352
23. López de Egea G, González-Díaz A, Guédon G, Lao J, Berbel D, Casabella A. A new integrative and mobilizable element is a major contributor to tetracycline resistance in *Streptococcus dysgalactiae* subsp. equisimilis. Antibiotics 2023;12(3):579. https://doi.org/10.3390/antibiotics12030579
24. Lorenzo-Díaz F, Fernández-López C, Douarre PE. Streptococcal group B integrative and mobilizable element IMESag-rpsI encodes a functional relaxase involved in its transfer. Open Biol 2016;6(10):160084. https://doi.org/10.1098/rsob.160084
25. Lorenzo-Díaz F, Fernández-López C, Lurz R, Bravo A, Espinosa M. Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective. Plasmid 2014;74:15-31. https://doi.org/10.1016/j.plasmid.2014.05.004

26. Lu M, Gong T, Zhang A, Tang B, Chen J, Zhang Z, Li Y, Zhou X. Mobile genetic elements in streptococci. Curr Issues Mol Biol 2019;32:123-166. https://doi.org/10.21775/cimb.032.123
27. McNeilly CLMcMillan DJ. Horizontal gene transfer and recombination in *Streptococcus* dysgalactiae subsp. equisimilis. Front Microbiol 2014;5:676. https://doi.org/10.3389/fmicb.2014.00676
28. Rezaei Javan R, Ramos-Sevillano E, Akter A, Brown J, Brueggemann AB. Prophages and satellite prophages are widespread in *Streptococcus* and may play a role in pneumococcal pathogenesis. Nat Commun 2019;10:4852. https://doi.org/10.1038/s41467-019-12825-y
29. Roberts APMullany P. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. FEMS Microbiol Rev 2011;35(5):856-871. https://doi.org/10.1111/j.1574-6976.2011.00283.x
30. Salvadori G, Junges R, Morrison DA, Petersen FC. Competence in *Streptococcus* pneumoniae and close commensal relatives: mechanisms and implications. Front Cell Infect Microbiol 2019;9:94. https://doi.org/10.3389/fcimb.2019.00094
31. Santoro F, Vianna ME, Roberts AP. Variation on a theme: an overview of the Tn916/Tn1545 family of mobile genetic elements in the oral and nasopharyngeal streptococci. Front Microbiol 2014;5:535. https://doi.org/10.3389/fmicb.2014.00535
32. Scott J, Thompson-Mayberry P, Lahmamsi S, King CJ, McShan WM. Phage-associated mutator phenotype in group A *Streptococcus*. J Bacteriol 2008;190(19):6290-6301. https://doi.org/10.1128/JB.01569-07
33. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. Microbiol Mol Biol Rev 2010;74(3):434-452. https://doi.org/10.1128/MMBR.00020-10
34. Wang J, Qi K, Bai X, Wu Z, Kang W, Liang P, Zheng H, Xu J. Characterization of integrative and conjugative elements carrying antibiotic resistance genes of *Streptococcus* suis isolated in China. Front Microbiol 2022;13:1074844. https://doi.org/10.3389/fmicb.2022.1074844
35. Wang M, Liu G, Liu M, Tai C, Deng Z, Song J, Ou HY. ICEberg 3.0: functional categorization and analysis of the integrative and conjugative elements in bacteria. Nucleic Acids Res 2024;52(D1):D732-D737. https://doi.org/10.1093/nar/gkad935
36. Wozniak RAFWaldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 2010;8(8):552-563. https://doi.org/10.1038/nrmicro2382
37. Yang Q, Zhu Y, Schwarz S, Wang L, Liu W, Yang W. Integrative and conjugative elements in streptococci can act as vectors for plasmids and translocatable units integrated via IS1216E. Int J Antimicrob Agents 2023;61(5):106793. https://doi.org/10.1016/j.ijantimicag.2023.106793

Figures

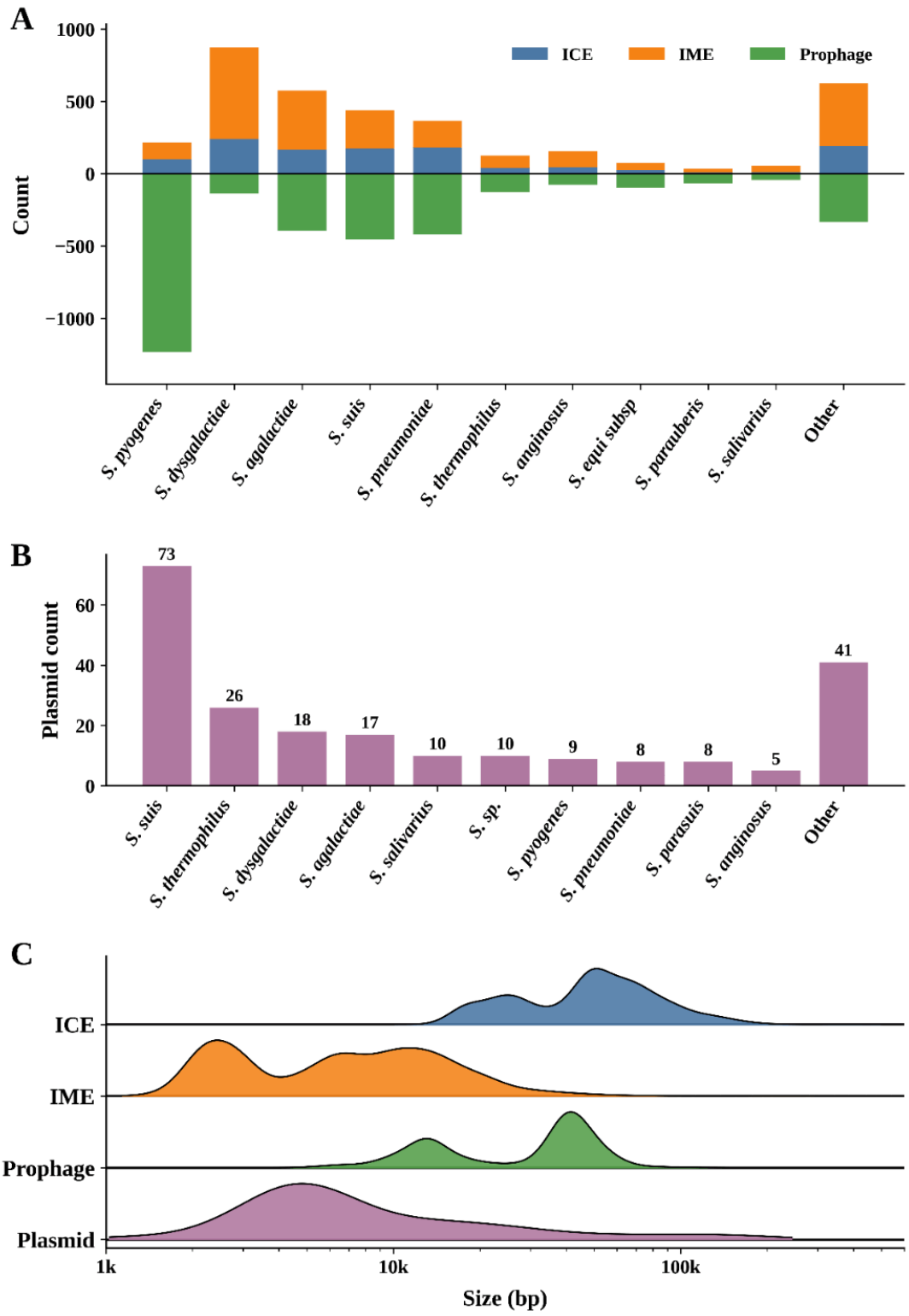

Figure 1

The four major MGE classes are abundant but unevenly distributed across the streptococcal mobilome. **(A)**Distribution of ICEs, IMEs and prophages across the top 10 *Streptococcus*species ranked by total chromosomally associated MGE burden, with all remaining species combined as Other. ICEs and IMEs are shown above the zero line, whereas prophages are shown below the zero line. **(B)** Distribution of plasmids across the top 10 *Streptococcus* species ranked by plasmid count, with all remaining species

combined as Other. **(C)** Size-density distributions of ICEs, IMEs, prophages and plasmids. Curves were generated from element sizes after kernel density estimation.

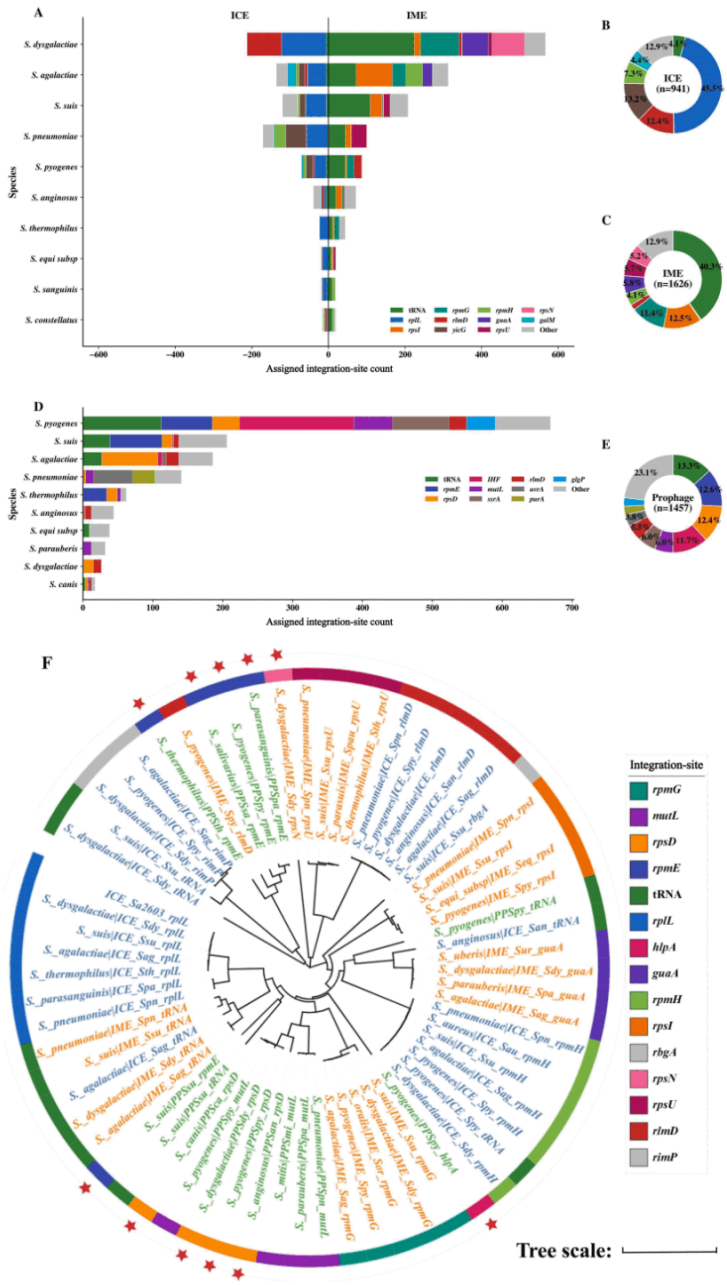

Figure 2

ICEs, IMEs and prophages exhibit quantitatively distinct hotspot hierarchies. **(A)** Species-level distribution of ICE and IME integration sites across the top 10 *Streptococcus* species with the highest

numbers of assigned ICE/IME insertion events. ICE-associated integration sites are shown on the left and IME-associated integration sites on the right. **(B)** Overall hotspot composition of ICEs. **(C)** Overall hotspot composition of IMEs. **(D)** Species-level distribution of prophage integration sites across the top 10 *Streptococcus*species with the highest numbers of assigned prophage insertion events. **(E)**Overall hotspot composition of prophages. **(F)** Phylogenetic relationships of representative integrases associated with major integration-site categories in the streptococcal mobilome. Tip labels indicate representative element identity and integration-site association. Red star symbols mark the integration sites highlighted in this study as newly emphasized recurrent hotspots.

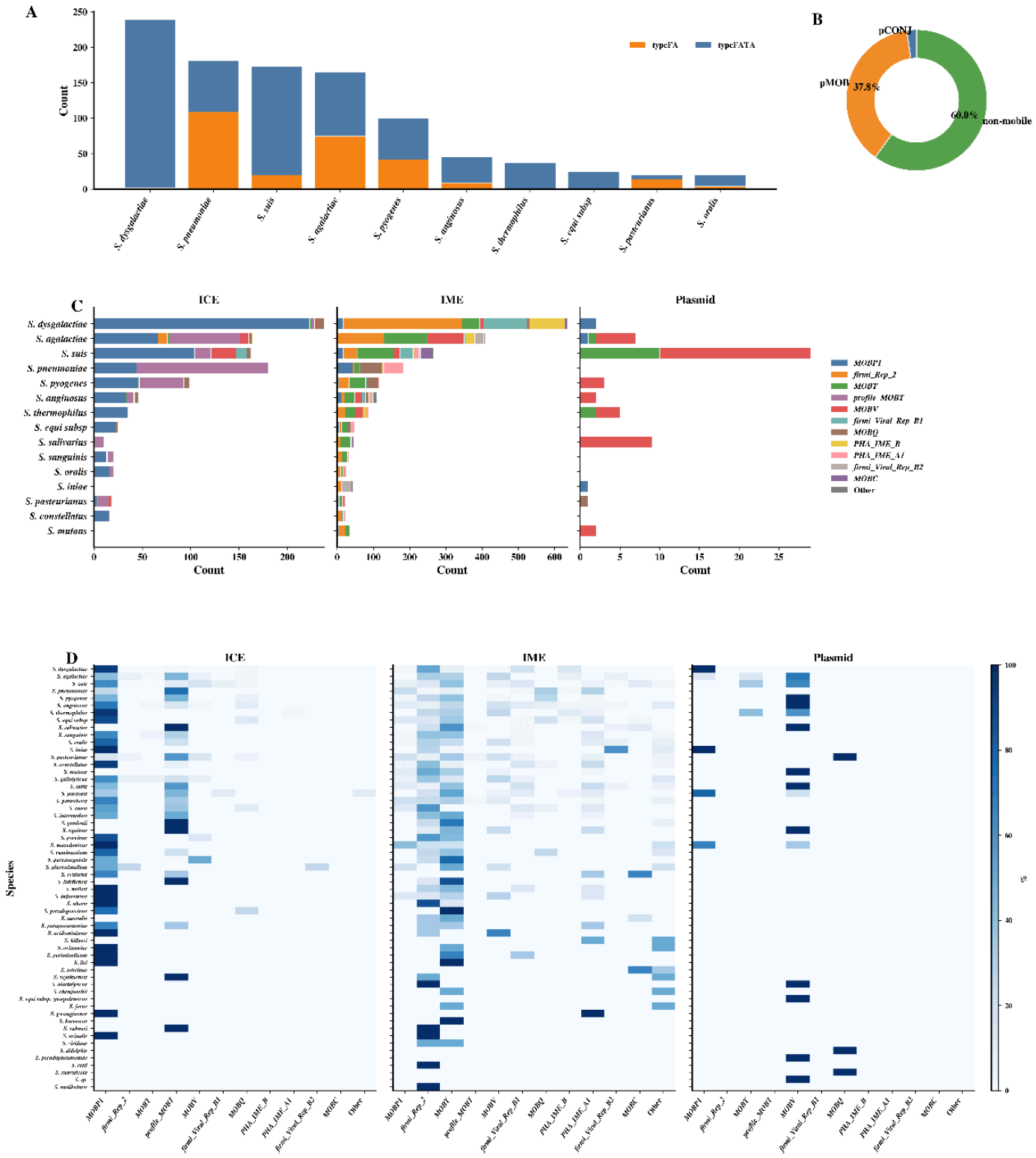

Figure 3

ICEs, IMEs and plasmids encode distinct mobility strategies. **(A)** Distribution of MPF types among ICEs in the top 10 ICE-rich *Streptococcus* species. **(B)** Mobility classification of plasmids into non-mobile, mobilizable (pMOB) and conjugative (pCONJ) types. **(C)** Relaxase family composition in ICEs, IMEs and plasmids across the top 15 species with the highest representation of conjugation-related MGEs. **(D)**

Species-level heatmaps showing the within-species percentage distribution of major relaxase families in ICEs, IMEs and plasmids.

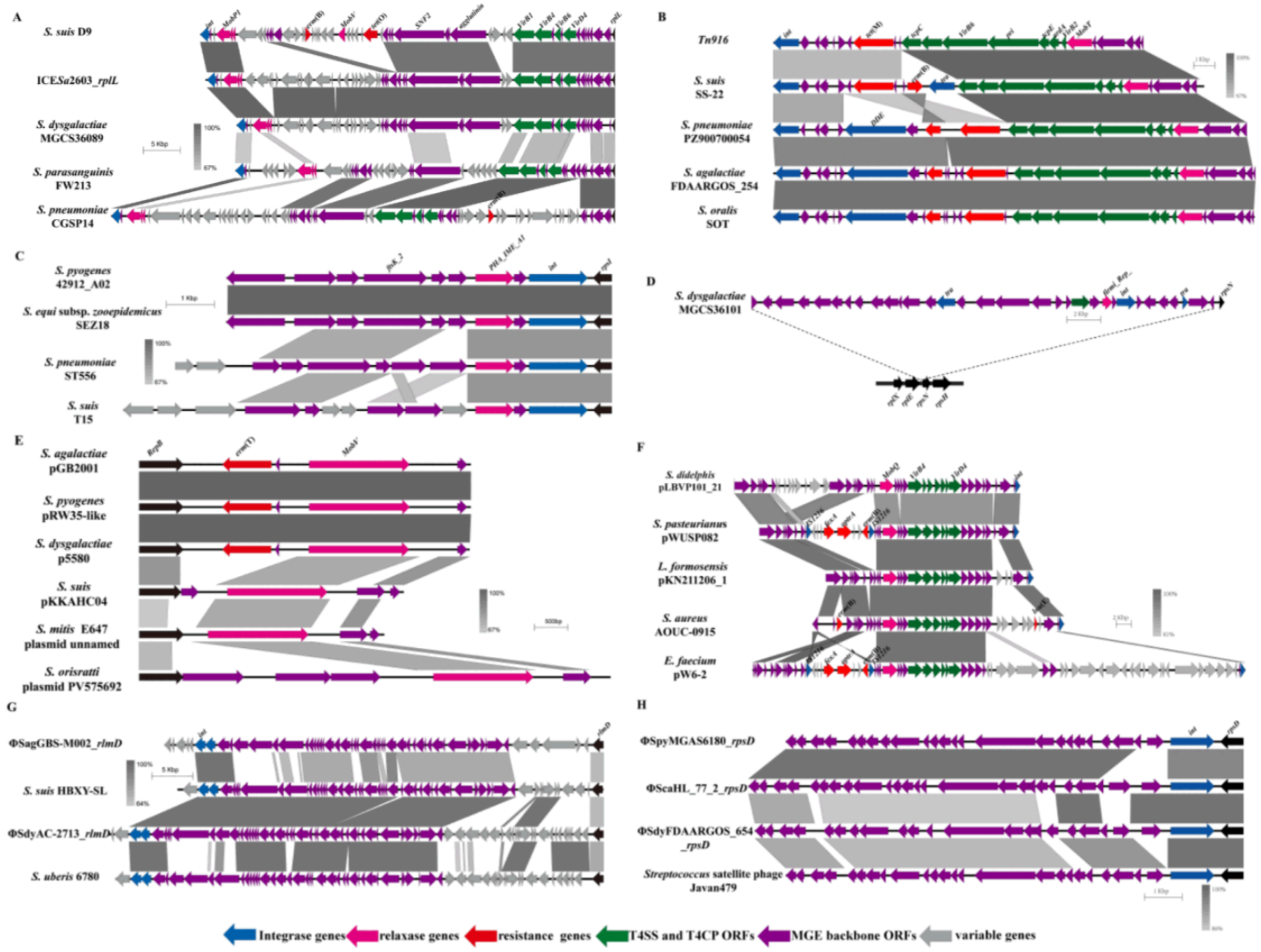

Figure 4

Representative MGE families retain conserved backbones across species. **(A)** Representative *rplL*-associated ICESa2603-like ICE family. **(B)** Representative compact *Tn916*-like ICE family. **(C)** Representative *rpsI*-associated IME family. **(D)**Representative *rpsN*-associated IME family. **(E)** Representative pMV158-like MOBV-bearing plasmid family. **(F)** Representative conjugative plasmid backbones detected in Streptococcus and related genera. **(G)**Representative *rlmD*-associated prophage family. **(H)**Representative *rpsD*-associated prophage family. Arrows indicate predicted coding sequences and are colored according to gene category: blue, integrase genes; pink, relaxase genes; red, antimicrobial resistance genes; green, T4SS- and T4CP-related genes; purple, MGE backbone genes; gray, variable or accessory genes. Shaded links indicate homologous regions between representative elements.

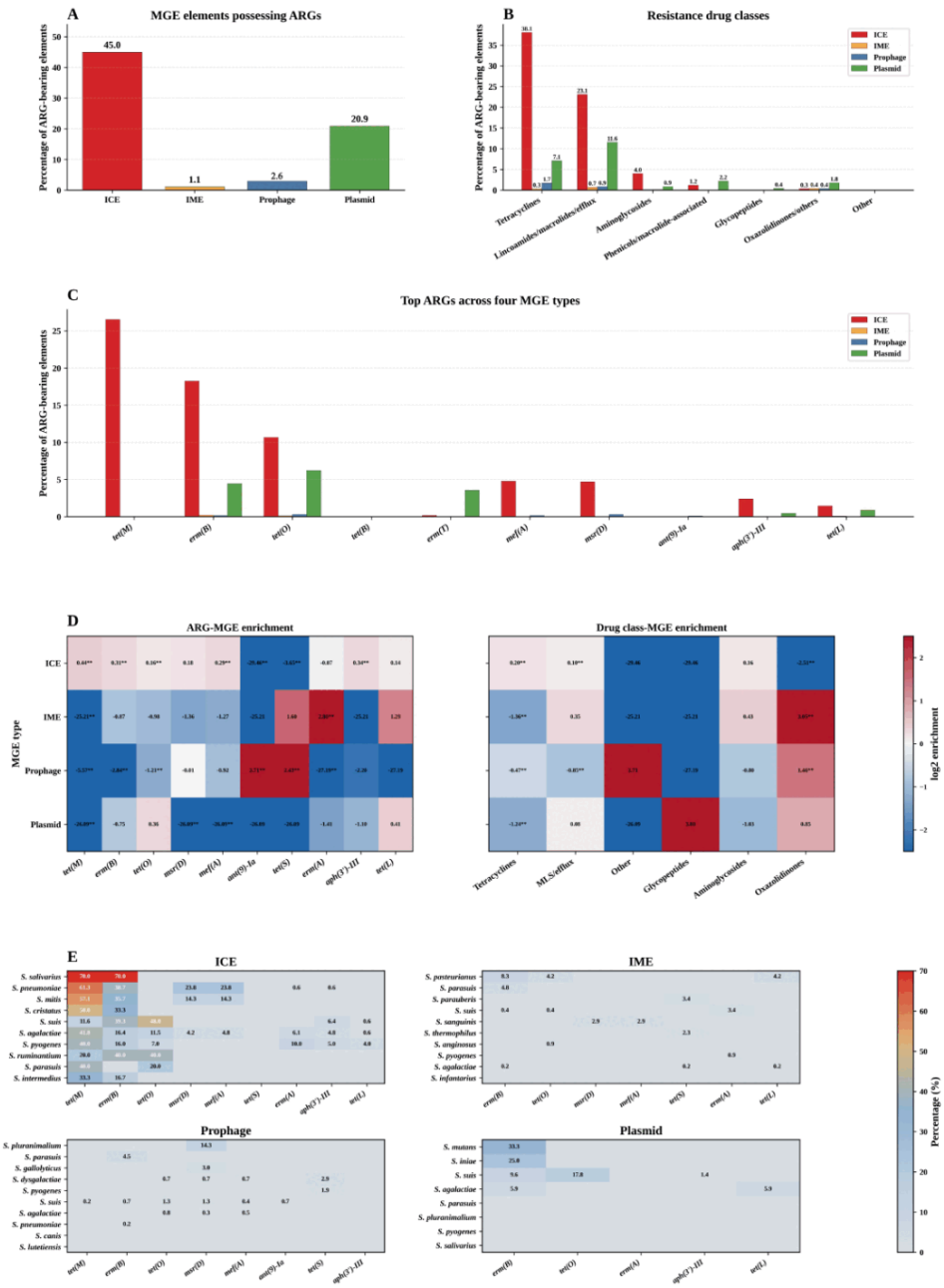

Figure 5

Host species shape the distribution of ARG-bearing MGEs. **(A)** Percentage of elements carrying at least one ARG in ICEs, IMEs, prophages and plasmids. Values above bars indicate the percentage of ARG-bearing elements within each MGE class. **(B)**Distribution of major resistance drug classes across ARG-bearing ICEs, IMEs, prophages and plasmids. **(C)** Distribution of the most abundant ARGs across the four major MGE classes. **(D)** Enrichment heatmaps showing the relative association of major ARGs (left) and

resistance drug classes (right) with different MGE classes. In both panels, the y-axis indicates MGE type and the x-axis indicates ARGs or drug classes, respectively. Positive values indicate relative enrichment and negative values indicate relative depletion compared with the overall ARG-bearing MGE background. **(E)** Species-level heatmaps showing the distribution of major ARGs in ARG-bearing ICEs, IMEs, prophages and plasmids. Species names are shown in abbreviated form.

Supplementary Files

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

- FigS12.docx
- TablesS16.xlsx