

Research Article

Genomic islands and the evolution of livestock-associated *Staphylococcus aureus* genomes

 Relangi Tulasi Rao¹, Shivani Sharma², Natesan Sivakumar³ and  Kannan Jayakumar¹

¹Department of Animal Behaviour and Physiology, School of Biological Sciences, Madurai Kamaraj University, Madurai 625021, Tamil Nadu, India; ²Research Intern, Dr. Praveen K. Lab, Centre for Ecological Sciences, Indian Institute of Science, Bengaluru 560012, Karnataka, India; ³Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai 625021, Tamil Nadu, India

Correspondence: Kannan Jayakumar (jklabinbox@gmail.com)



Background: Genomic Islands (GIs) are commonly believed to be relics of horizontal transfer and associated with specific metabolic capacities, including virulence of the strain. Horizontal gene transfer (HGT) plays a vital role in the acquisition of GIs and the evolution and adaptation of bacterial genomes.

Objective: The present study was designed to predict the GIs and role of HGT in evolution of livestock-associated *Staphylococcus aureus* (LA-SA).

Methods: GIs were predicted with two methods namely, Ensemble algorithm for Genomic Island Detection (EGID) tool, and Seq word Sniffer script. Functional characterization of GI elements was performed with clustering of orthologs. The putative donor predictions of GIs was done with the aid of the pre_GI database.

Results: The present study predicted a pan of 46 GIs across the LA-SA genomes. Functional characterization of GI sequences revealed few unique results like the presence of metabolic operons like *leuABCD* and *folPK* genes in GIs and showed the importance of GIs in the adaptation to the host niche. The developed framework for GI donor prediction results revealed *Rickettsia* and *Mycoplasma* as the major donors of GI elements.

Conclusions: The role of GIs during the evolutionary race of LA-SA could be concluded from the present study. Niche adaptation of LA-SA enhanced presumably due to these GIs. Future studies could focus on the evolutionary relationships between *Rickettsia* and *Mycoplasma* sp. with *S. aureus* and also the evolution of Leucine/Isoleucine mosaic operon (*leuABCD*).

Introduction

The gene transfer between the different bacterial species has a great impact on the evolution and transformation of bacterial pathogens. Genomic Islands (GIs) are unfamiliar gene blocks present in pathogenic and non-pathogenic prokaryotic genomes apart from the core genome [1]. They play a role in determining accessory functions, such as antibiotic resistance, secondary metabolic activities, symbiosis and other special functions related to sustaining during adverse environmental conditions. These GIs supposedly associate with the virulence of the pathogenic bacteria and are known as pathogenicity islands (PAIs) [2]. GIs have evidence of horizontal origins, i.e. Horizontal Gene Transfer (HGT), which means that an interchange of genetic information between phylogenetically distant organisms takes place [3]. The advantage of acquisition of GIs has an evolutionary edge and that is a large number of genes can be transferred and incorporated into the recipient genome. This transfer may lead to dramatic changes in an organism and ultimately result in a quantum leap in evolution [4].

During the evolution, many bacteria have equipped their genomes with DNA from other bacterial species or even genus with the help of mobile genetic elements (MGEs). These MGEs also referred to as

Received: 21 July 2020
Revised: 23 September 2020
Accepted: 07 October 2020

Accepted Manuscript online:
13 November 2020
Version of Record published:
25 November 2020

accessory genetic elements, are therefore a potential resource for bacteria that provide adaptive strengths to improve the fitness and, potentially, pathogenicity and drug resistance [5].

For instance, in *Pseudomonas aeruginosa*, accessory gene elements have enhanced the virulence toward invertebrate *Caenorhabditis elegans* [6]. GIs' another class of accessory gene elements contribute to the evolution and adaptation to the niche. *Staphylococcus aureus* also possesses several MGEs, including GIs but the knowledge on GIs of livestock-associated *S. aureus* (LA-SA) is limited.

Computational analysis of whole-genome sequences of several bacterial strains predicted that GIs are present in any given strain [7]. In the present genomic era, the number of sequenced bacterial whole genomes is increasing rapidly. Comparative genomic studies became the initial step in analyzing the microbial genomics and it helps in deciphering the evolutionary relationships [8]. Public accessibility of data of whole-genome sequences and development of bioinformatics methods makes it possible to study the evolution of bacteria, pathogenicity and other associated characteristics. Recognition and analysis of GIs contribute toward a better understanding of the evolution of the disease and the development of bacterial pathogenicity and even in understanding the evolution of host-specific strains [4,9]. Moreover, zoonotic transfer of *S. aureus* is becoming a major possible way of spreading resistant *S. aureus*. Hence, it is imperative to predict and analyze the evolution of the GIs in the livestock-associated strains, to comprehend the evolutionary developments. Pertaining to this context, the present study was planned-out in order to predict the GIs and the genetic elements' coding by GIs in the available completed genomes of *S. aureus* which have livestock-association. Furthermore, an in-depth bioinformatics comparative analysis of predicted GIs and their functions was performed.

Materials and methods

Acquisition of genome sequences

Complete genome sequences of *S. aureus* strains available at NCBI's FTP server (ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Staphylococcus_aureus) were retrieved only after manual verification of the Livestock association or origin. Chromosomal sequences alone were analyzed and annotated with Prokka: rapid prokaryotic genome annotation tool in local machine [10] for maintaining uniformity in the annotation files. The details of genome sequences were presented as supplementary data (Supplementary Table S1).

Prediction of GIs

Initial comparative analysis was done with GView Server [11], to draw circular chromosomes based on the BLAST-core to check the similarity. The following existing methods were deployed to predict GIs of the genomes. These methods are Ensemble algorithm for Genomic Island Detection ((EGID)—[12]), and SeqWord Sniffer-python language script [13]. EGID: a tool for improved GI detection in genomic sequences, is based on the predicted results of five existing GI programs, namely Alien Hunter [14], COLOMBO SIGI-HMM [15], INDeGenIUS [16], IslandPath [17] and PAI-DA [18]. The framework of this program includes (i) collection of prediction results from existing five programs; (ii) analysis and filtering of predicted results; and (iii) generating final consensus GI results. SeqWord Gene Island Sniffer program is based on the analysis of oligonucleotide usage variations in DNA sequences and detects putative horizontally transferred gene clusters. Chromosomal sequences alone were analyzed in the present study as GIs were integrated part of chromosomes.

Clusters of orthologous group enrichment analysis

Clusters of Orthologous Group (COG) of proteins enrichment analysis was performed in order to establish the functional characterization of GIs. COG categorization was carried out online using the Batch version of Conserved Domain Database [19], searching against the COG database while using other default parameters. 'NA' was used instead when the genes did not match any COG accession numbers.

Virulent genes analysis

To identify possible virulence factors, the curated and experimentally validated virulence factors of Virulence Factors Database (VFDB) were aligned with the ORF protein sequences. BLAST-based virulence gene analysis was performed in local machine, and virulence factor protein sequences were downloaded from VFDB (<http://www.mgc.ac.cn/VFs/> [20]). A local database was built in the local machine with VFDB.SetA (curated and validated) protein sequences. Proteins encoded by GIs were aligned against this database with minimum e-value as $1e^{-10}$ using Blastp and were filtered with 75% identity and 95% aligned length.

Distribution analysis of GIs

The predicted GI elements clustering analysis was performed with ClustAGE software [21] as described in the manual of the software. Heat map of Bray–Curtis (BC) [22] similarity values to the neighbor-joining tree were visualized in the online tree visualization software Interactive Tree Of Life [23].

Credible donors of predicted GIs

A novel framework was proposed and developed to detect the GIs with the aid of the concept of island ontology and proposed island flow from the pre_GI database [24]. Initially, all available Oligonucleotide Usage Pattern (OUP) neighbors data, i.e. proposed host, compositional similarity (CS) and island distance (D) for Query Island were retrieved from the pre_GI database, except the data for the proposed island flow from the query to subject with the compositional similarity cutoff of 80%. The values of D were normalized and the data with values > 0 were removed from the analysis ($x_i > 0$, regarded as false positives); lower D values indicate the most probable donor. The finalized data were manually curated and the bacterial donors were removed with which there is no possible interaction, e.g. marine thermophiles and archaea have a low possibility of terrestrial bacteria to interact. Finally, all the probable donors were grouped into three categories based on the normalized values of D. Group I was regarded as potent probable neighbors, Group II had moderately probable neighbors and group III had the least probable donors (Supplementary Data S2). The credible donor relationship was established by comparing the two GIs [Query and Subject (probable donor)] with LingvoCom [24].

Results

Properties and comparative studies of genomes

All the points of interest of the genomes used as a part of the present study have been classified (Table 1). Among these strains, *S. aureus* ST398 has the large genome with the size of 2.87 Mb, which has bovine origin but isolated from human and the smallest genome is of *S. aureus* 71193 strain with 2.71 Mb. The genomes demonstrated a noteworthy deviation in size from one another and the GC content of all genomes appeared to be same with an estimation of 33%.

GView server, a BLAST-based approach, for genome comparisons was employed in order to further analyze the collinearity between genomes. It demonstrated a similarity between genomes at approximately 90–98% on sequence level and can be seen in Figure 1. Similar results of Mauve analysis were repeated here with a better understanding of the unique regions of genomes when compared with that of the reference genome. The loss and gain of genes could be seen from the whole-genome BLAST results, wherein ED133 strain had much more similar nucleotide content while RF122 strain had the least similarity. The regions around 1.8 and 2.1 Mb were unique to the majority of strains in the study. The GView server results of Blast-based comparisons were similar to the results obtained from Mauve alignment.

Prediction of GIs

GIs were predicted with the EGID tool, and then followed by the SeqWord Gene Island Sniffer and. EGID uses the resulted GI coordinates of five existing GI prediction programs mentioned in the methods section, followed by filtering statistically significant and consensus GI coordinates [12]. SeqWord Gene Island Sniffer program analyzes the oligonucleotide usage variations in DNA sequences and detects putative horizontally transferred gene clusters [13]. The overlapping GIs from these methods was removed manually and examined the HGT mechanisms of GIs based on the annotations. Further, the coordinates of GIs were adjusted on the basis of the HGT mechanism. Additionally, the false-positive predicted GIs (e.g. ribosomal genes) and essential genes were removed from the analyses.

It was concluded that 46 different GIs prevailed across the study genomes, most GIs were 17 in number as observed in *S. aureus* ST398 strain and least were only 11 in number as found in 08S00974 strain, in spite of its larger genome-size elevated coverage by the GIs was also in accordance with the predicted number of GIs and was found to be 15.5% in whole-genome sequence and 15.3% in proteome in case of *S. aureus* ST398 (large genome) strain, whereas the least coverage was seen in *S. aureus* 08S00974 (Table 1). LA-SA strains possess an average of 13 GIs in their genomes and the least number of GIs was 11.

Some of the known islands from other *S. aureus* strains were identified in the LA-SA genomes and represented in Table 2. They were regarded as PAIs since these Islands coded for virulence factors like enterotoxins and clumping factors [25]. The visualization of GIs of 08BA02176, RF122 and ST398 strains was done with *DNAplotter*, a Java-based program [26] as seen in Figure 2.

Table 1 Characteristic features and properties of LA-SA genomes selected for the present study

Strain/Feature	08BA02176	08S00974	71193	E154	ED98	ED133	ISU935	LGA251	Newbould.305	NZ15MR0322	O11	O46	RF122	ST398
Sequence type	st398	st398	st398	st398	st5	st133	st5	st130	st115	st398	st130	st130	st151	st398
Size (Mb)	2.78	2.8	2.71	2.83	2.82	2.83	2.86	2.75	2.8	2.83	2.76	2.79	2.74	2.87
Genes	2593	2613	2531	2671	2681	2710	2683	2564	2662	2750	2596	2614	2656	2738
CDS	2532	2551	2472	2609	2618	2649	2621	2501	2601	2688	2569	2562	2594	2676
tRNA genes	58	61	58	61	62	60	61	61	59	61	29	51	61	61
tmRNA genes	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Repeat regions	2	0	0	0	0	0	0	0	1	0	0	0	0	0
GC%	32.94	32.94	32.93	33	32.84	32.92	32.91	32.96	32.92	32.86	32.78	32.83	32.78	32.92
Predicted GIs	14	11	13	13	12	13	13	14	13	12	11	11	14	17
Coverage in genome (%)	14.3	9.63	9.96	11.4	11.16	13.97	11.42	14.13	13.96	10.9	14.1	10.59	13.64	15.55
Coverage in proteome (%)	13.57	6.97	10.5	8.34	11.76	15.31	9.35	12.24	13.14	9.89	13.04	9.41	13.67	15.23

Abbreviation: CDS, Conserved Domain Search.

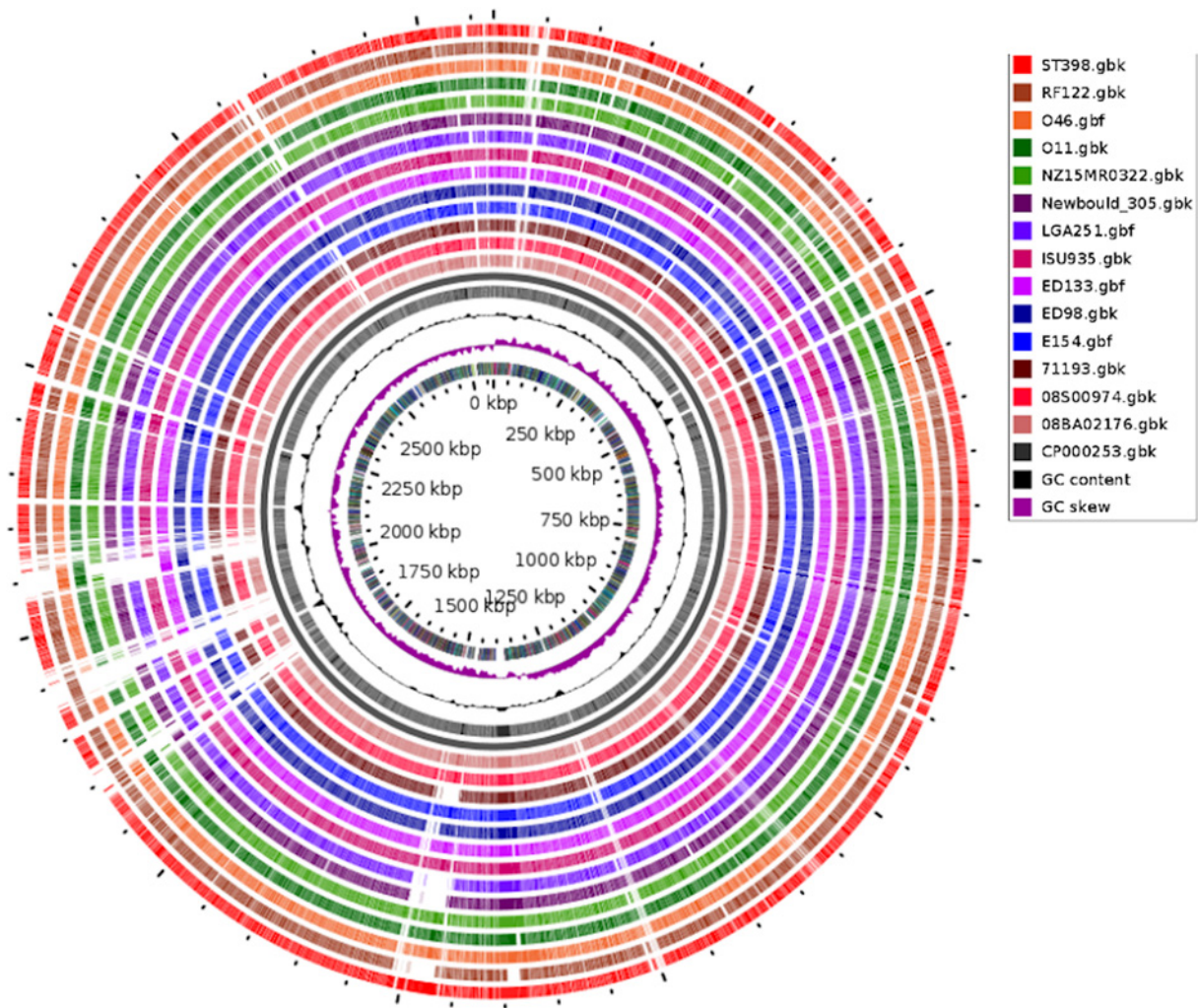


Figure 1. BLAST-based comparative genome maps

The comparative genome map generated with the GView server using *S. aureus* NCTC8325 as a reference sequence to other *S. aureus* genomes. The outermost ring depicts the genes of a reference strain followed by the genomes of other strains based on the BLAST similarity. The genome sequences are colored differently, and regions without colors are absent from the respective strain, and highlight differences in the accessory gene content.

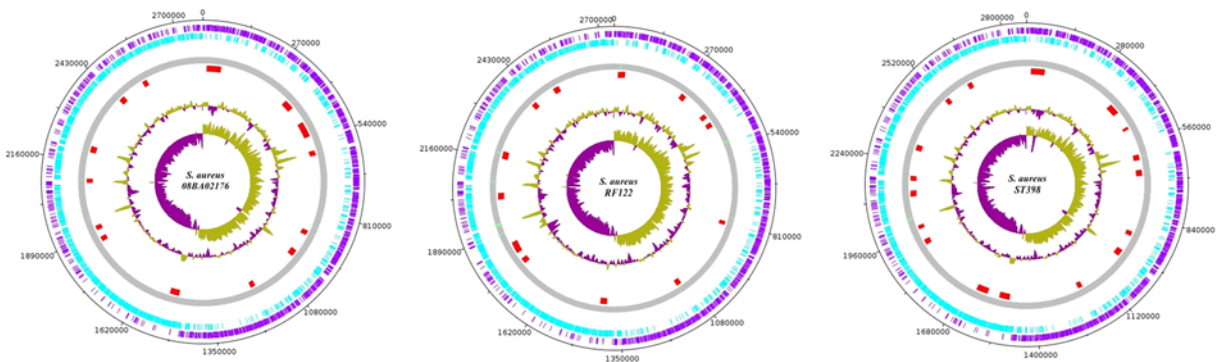


Figure 2. Location of GI coordinates on the circular genome

The circular genome of *S. aureus* strains represent the predicted GI locations. The inner circle represents the GC skew and the next represents the GC content plot and the red dots represent the predicted GI location on the circular genome (plotted by *DNAplotter*).

Table 2 Previously published Staphylococcal GIs predicted in these LA-SA' genomes*

GI	LASA-GI	Coordinates	Function	08BA20716	08S00974	71193	E154	ED98	ED133	ISU935	LGA251	N305 [†]	NZ15	O11	O46	RF122	ST398
φSa2	08BA02176_GI2	354723-392677	Virulence	+	+	-	+	+	+	+	-	+	+	+	+	-	+
νSaβ	LGA251_GI11	1870085-1902095	Virulence	-	-	-	-	-	+	-	+	-	-	-	-	+	-
φSa3	77193_GI10	1934043-1976257	Virulence	-	-	+	-	-	+	-	-	-	-	+	+	-	-
νSaγ	08BA02176_GI7	1182896-1202263	Virulence	+	-	+	-	+	+	-	+	+	-	-	-	+	+
Salbov	08BA02176_GI3	458784-517982	Virulence	+	+	+	-	+	+	-	+	+	-	-	-	+	+
Type-V SSCmec	08BA02176_GI1	16107-72549	Resistance	+	+	+	-	-	+	-	+	-	-	-	-	+	+

*From PAI-DB [32].
[†]Newbould_305.

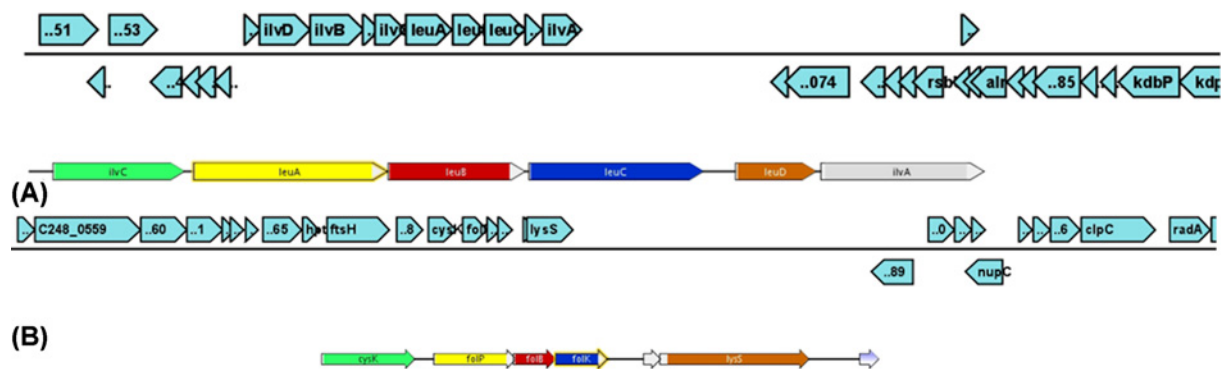


Figure 3. Genetic maps of the operons present in the GIs

Gene maps of GIs indicating the presence of genes involved in metabolism (A) 08BA02176_GI11 with *leuABCD*, (B) 08BA02176_GI14 with folic acid biosynthesis genes.

Mechanisms involved in the transfer of GIs

The three common mechanisms mainly meant for HGT are through Phage integrase (Transduction), Transposon – Transposase (Insertion Sequences) and tRNA [27]. Mostly, phage-related integrase genes are present on these predicted GIs, suggesting that they are integrated and excised in a method similar to prophages. *S. aureus* Pathogenicity Island (SaPI) is mobilizable with the assistance of the transducing phages [28]. The majority of the predicted GIs (13/46) were noted to flank by transposon – transposase genes and followed by phage integrases suggesting that transposase is majorly responsible for the acquisition of GIs (Supplementary Table S2). The present study also suggests that transducing phages are responsible for the acquisition of GIs. Finally, it was concluded that from the mechanisms involved in gene transfer, transposon – transposase and phage infections are important mechanisms involved in the HGT events of the studied genomes.

Homologous GIs

Homologs from the predicted GIs were retrieved with GET_HOMOLOGUES software [29], for understanding the relationship among LA-SA genomes, and also for performing comparative analysis. Further, the conserved domain superfamily analysis was also carried out for the retrieved homologs representing each strain [19]. These findings showed that only eight ORFs of all GIs are common to all strains under the study. Unique proteins of each strain are majorly noted as hypothetical proteins that belong to an unknown family of proteins. The common proteins are mainly Phage proteins and virulent proteins. The Conserved Domain Search (CDS) analysis also showed that many hypothetical proteins are assigned to the unknown family and other hypothetical proteins belong to proteins. The presence of phage proteins indicates frequent phage infections. RF122 strain's GIs harbor the highest unique ORFs and lowest in 08S00974.

Functional categorization and COG enrichment analysis

Based on the annotation results of genomes, the predicted GIs mostly possess phage and hypothetical proteins (Supplementary Data S1). Besides the usual hypothetical proteins, several GIs encode for some virulent proteins, and also confer antibiotic resistance, as expected. Apart from these functions, some GIs encode for metabolism-related functions as well, e.g. LASA-GI4 genomes encode genes for folate synthesis and LASA-GI19 encodes for trehalose metabolism. This unique finding from the study shows the mosaic operon cluster transfer of LASA-GI11, encoding for Leucine/Isoleucine biosynthesis (Figure 3).

It was found that potent toxin genes like enterotoxins and exotoxins genes were associated with GIs and were found to have HGT origin because of their anomalous GC content (Supplementary Data S1). Phage infections were responsible for the transfer of these toxin genes. The other important genes, which have a horizontal origin, are fibronectin-binding proteins. Fibronectin-binding proteins are adhesins, cell wall-associated proteins involved in critical host–pathogen interactions [30]. These genes, found across all GIs of the study genomes and other putative fibronectin-binding proteins, are also associated with GIs. Furthermore, these proteins were reported to evolve in *S. aureus* in the ruminant host habitat [31]. When we compare the GIs including PAIs of other *S. aureus* strains (human pathogenic *S. aureus*) published in PAI-DB [32], it was observed that the GIs majorly encoded for toxins and antibiotic resistance genes but not for fibronectin-binding proteins in major. Since all the strain genomes used in the

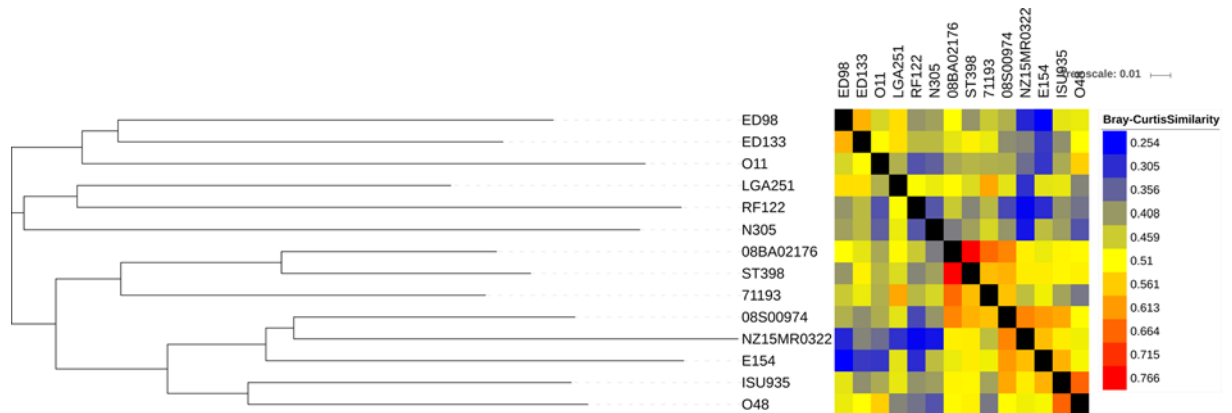


Figure 4. Cluster analysis of GI elements

BC distances (d) calculated for every pair-wise comparison of shared GI element content between strains. Neighbor-joining tree (left) is a consensus across 1000 bootstrap resampling of distributions. The heatmap shows relative pairwise similarity (1 - d) between strains.

present study have animal association, the association of fibronectin-binding proteins with GIs is significant in the view of evolutionary strategies adopted by *S. aureus*.

For approximately 1250 homologous protein sequences representing the overall GI elements of LA-SA genomes of the present study, only 440 sequences were assigned with COG accession numbers and categorized into different functional groups (Supplementary Data S1), since the majority of the sequences are annotated as hypothetical proteins which do not belong to the existing functional classes. Hence, the number of sequences assigned to the COG functional class was less. The overall enrichment analysis resulted in the GI elements that were with molecular functions-related elements like replication, transcription and translation regulatory proteins (25%), followed by general prediction class and proteins with unknown functions (COG class R and S – 14%), amino acid transport and metabolism-related proteins (COG class E – 9%), Phage-related elements (COG class X – 8%) and Defense-related proteins (COG class V – 5%).

Virulent genes across the GIs

The Blastp analysis of GIs against VFDB resulted in identifying the major genes contributing the virulence in LA-SA (Supplementary Table S3). The enrichment analysis against a database indicated that these GIs carried important virulent genes. This analysis identified that the GIs predominantly encode for toxin genes, especially different kinds of enterotoxins followed by adherence-related genes like fibronectin and fibrinogen binding proteins, serine–aspartate rich fibrinogen-binding proteins (*sdr* genes) and clumping factor. Exotoxins and Type VII Secretion System (T7SS) genes that have been associated with virulence in *S. aureus* were also part of GIs. These results suggested that the genes associated with virulence in LA-SA were hustled through HGT mechanisms.

The GI elements distribution and their relatedness

The homologous GI elements were distributed among the *S. aureus* genomes in the present study. The distribution and relatedness of these GI elements help to understand the implications of GIs on the strain-based relatedness, and the evolution of LA-SA genomes. The sequence similarity and the distribution of related GI elements among the LA-SA genomes were depicted in Figure 4 with BC distance-based heat map and Neighbor-joining tree. The heat map based results suggested that the GI elements among the LA-SA genomes have intermediate relatedness. During clustering, the GI elements resulted in BC distance ranging only from 0.25 to 0.76. This was presumably due to the differences in the composition of GI elements. The highest BC (0.766) similarity was seen between 08BA20176 and ST398, presumably, because they both belonged to st398 type. But the average BC similarity coefficient was found to be only 0.5, a moderate similarity coefficient likely due to different phage infections (frequency of certain phage groups varied between *S. aureus* Clonal Lineages – [33]), barriers such as restriction–modification systems (R–M systems) and niche separation reducing the opportunities for HGT [34]. Sequence-type based clustering was not observed when the phylogenetic tree was constructed with GI elements; and this observation suggests similar HGT events that were, presumably, not dependent on STs of strains. The highest BC similarity was seen between 08BA20176 and

ST398 strains while least was found in between E154 and ED98 strains and also in NZ15MR0322 and RF122 strains. These results indicated the uniqueness of GI composition of each strain.

Credible donors of GI elements

From the above results, it was obvious that HGT played an important role in the acquisition of GIs and the evolution of virulence and adaptation of LA-SA strains. Thus, it is practically significant to explore the donors of GIs in order to understand the interactions with donors, as well as to interpret the HGT events. Predictions were performed for all 14 genomes of the present study and all the results were cumulated for understanding the donors for LA-SA GIs better. The donor–recipient island ontology further confirmed with LingvoCom (Supplementary Data S3).

This framework predicted several probable donors of GIs and the data were plotted as a network with gephi tool [35]. The results showed that *Rickettsia* sp. and *Mycoplasma* sp. were over-represented as donors of GI elements (Figure 5A). The family-level representation of the donors' list resulted in the same observation that Rickettsiaceae followed by Mycoplasmataceae were dominant donors of the GI elements of the LA-SA strains (Supplementary Figure S1). Further, these results compared with donors of GI elements of human-associated *S. aureus* (HA-SA), revealed that these strains presented with *Mycobacterium* sp. were also part of donors while some species of *Rickettsia* and *Mycoplasma* were absent from donors list of HA-SA (Figure 5B). These results suggested LA-SA and HA-SA strains were presented with diverse GIs, in the due course of evolution.

Discussion

GIs and their importance in LA-SA genomes

To have a better idea of the evolution of LA-SA, the roles of the accessory genome, and HGT events in pathogenicity, drug resistance and epidemiological information, GI prediction, and functional categorization were carried out with the available LA-SA's 14 complete genomes. Several studies emphasized the importance of HGT and GIs in the evolution of *S. aureus* and identified GIs including PAIs [36,37]. A different study reported the detection of 13 known GIs including PAIs across the 5 MRSA strains [38], comparable with the present study where 13 GIs have been predicted, on average, in the 14 LA-SA strains. Some GIs were widely distributed across these 14 LA-SA genomes while some were confined to two or three strains (Supplementary Data S1). This was so perhaps due to the barriers such as DNA R–M systems and niche separation which reduce the opportunities for HGT [34].

Phages and Transposon – Transposases were found to be predominantly involved in the exchange of genetic material through HGT in LA-SA genomes in the present study. Moon et al. [39] in 2016 suggested that phages involved in mediating GIs which confers virulence and resistance in some *S. aureus* strains. While transposons are not able to transpose within the bacterium, but they are capable to integrate with various sites of the other host's genome and are therefore able to transpose in the host genome [40]. Primarily, Transposon – Transposase-mediated GIs confer resistance to host strains [34]; but here we report that Transposon – Transposase-mediated GIs were predicted to encode virulence and some metabolism-related functions, apart from antibiotic resistance. A recent study by Jani et al. [38] also reported the Transposase-mediated GIs.

Functional characterization of GIs revealed that apart from the virulence and antibiotic resistance functions, certain GIs encode some metabolism-related functions. The Leucine/Isoleucine biosynthesis genes are a mosaic *leuABCD* operon and known to have horizontal transfer origin and probable source would be archaea [41]. This operon in *S. aureus* is involved in leucine and pyruvate metabolism. There were no available reports on the acquisition of *leuABCD* operon through HGT in *S. aureus*. The folic acid synthesis genes in *S. aureus* also seem to be a part of HGT events. The genes *folP*, *folB* and *folK*, involved in folate synthesis in *S. aureus*, were encoded by a GI in LA-SA genomes. This functional characterization of GIs also suggests the role of HGT in host-niche adaptation and the similar finding was well established in *Prochlorococcus* spp. and *Klebsiella pneumoniae* as well [42,43]. Another recent study on *S. aureus* also suggests the importance of GIs in the host specificity [44].

Even though many genes encoded by GIs were hypothetical proteins and phage-related protein, COG enrichment analysis revealed that certain genes encoding vital molecular functions, such as transcriptional regulatory genes, were enriched in GIs. These genes were primarily flanked with phage elements like integrase and capsid proteins, suggesting that temperate phages may responsible for this enrichment. Apart from these regulatory functions, phage proteins are known to increase the virulence of host bacterium, through regulation of expression of the virulence genes [45].

The other important genes which have horizontal origin are fibronectin-binding proteins. Fibronectin-binding proteins are adhesins, cell wall-associated proteins, involved in critical host–pathogen interactions [30]. The insights into the predicted GIs of strains clearly state that important virulence factors are associated with GIs. Sui et al. [46] reported that GIs inconsistently harbor greater number of virulence factors than the rest of the host genome, and

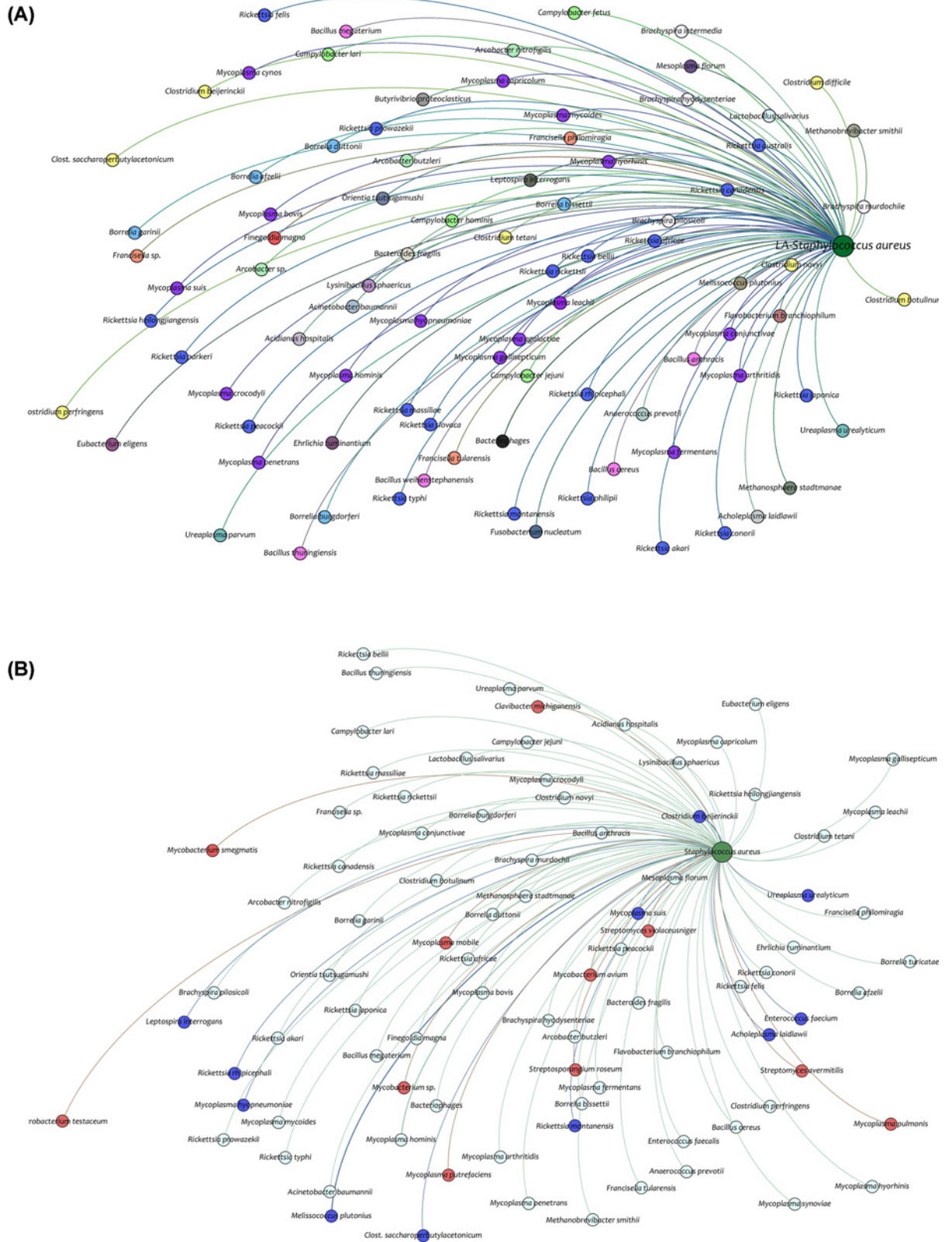


Figure 5. Representation of GI donors in the network model

(A) Donors of GIs of LA-SA strains and each colored node represent a different genus. (B) This network depicts the GI donors of both LA-SA strains and HA-SA strains. The red color nodes represent the unique donors to HA-SA strains while blue color for LA-SA strains and cyan colored nodes represents the common donors.

are enriched for proteins like toxins or hypothetical pathogen-associated genes. Similar results were obtained in the present study, which located proximity to the earlier studies. Analyzing the closely related pathogenic genomes have suggested that genes involved in virulence are apparently associated with PAIs, a subset of GIs [9,46]. This again supports the notion of the role that GIs play in the transformation of non- virulent strain to virulent strain.

GIs and their credible donors

GIs played a vital role in adaptation, survival and virulence of strains but which organisms contributed to acquiring such traits was yet to discover in the field of accessory genomics of bacteria. In the present study, an attempt was made to predict the donors of such traits, and also the relationships between host and donors. Predictions were performed for the 14 genomes of the present study and *Rickettsia* sp. and *Mycoplasma* sp. were found to have been over-represented as donors. Wan and Che [47] in 2014, also developed a similar kind of approach to predict the GI donors. They reported that *Gordonia*, *Nocardia* and *Rhodopseudomonas* species as major donors of GI elements of *Mycobacterium tuberculosis*. The predicted donor results also suggested the frequent phage infections and incorporated phage related genes in the genomes of LA-SA strains. There were few unique donors for LA-SA GIs and HA-SA GIs, presumably, because of niche separation.

There were not enough literature or reports available to support these findings of GI donors. While previous experimental research suggested that capsular polysaccharide genes of *S. aureus* were found to have homologs in the *Rickettsia* genome [48]. And there could be a possible membrane fusion between *Mycoplasma* and *S. aureus*, *Bacillus subtilis* for facilitating the transfer of conjugative elements between hosts and recipient cells [49,50]. The other plausible reasons for these results could be the interactions between *S. aureus* and other donors (or part of the microbiome of hosts) during the due course of infection. It was also reported that *Mycoplasma* and *S. aureus* are believed to coexist and cause bovine mastitis [51]. This study paves the way to realize a need to understand the evolutionary relationships with other bacterial species and in particular between *Rickettsia* and *Mycoplasma* sp. with *S. aureus*.

Conclusions

In conclusion, the established the fact that GIs plays a vital role in niche adaptation and evolution of LA-SA strains. The predicted GIs were observed to enhance the virulence capacity of strain as they primarily encoded for adherence-related proteins, like the fibronectin-binding proteins, clumping factors and also for most of the toxin genes, like entero- and exotoxins. The GIs functional enrichment and horizontal transfer of *leuABCD* mosaic operon and folic acid biosynthesis genes (*folBPK*) in *S. aureus* suggests their role in the niche adaptation. The donor prediction results of GI elements showed that *Rickettsia* sp. and *Mycoplasma* sp., over-represented as donors of GI elements in LA-SA strains and *Mycobacterium* sp., were unique donors for the HA-SA associated GI elements. Based on these findings, future studies could focus on the evolutionary relationships between *Rickettsia* and *Mycoplasma* sp. with *S. aureus*, and also the evolution of Leucine/Isoleucine mosaic operon (*leuABCD*) in *S. aureus*.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the UGC-BSR-SRF [grant number F.25-1/2013-14(BSR)/8-1/2007(BSR) (to R.T.R.)];

Author Contribution

R.T.R. and S.S. conceived and designed the experiments. R.T.R. conducted the experiments, and R.T.R. and S.S. analyzed the data. R.T.R. wrote the manuscript. S.S., N.S., and K.J. corrected and revised the manuscript.

Acknowledgements

We acknowledge UGC-NRCBS, UGC-CAS, DST-SERB, and DBT-IPLS for Infrastructure and instrumental facilities at Madurai Kamaraj University.

Abbreviations

BC, Bray–Curtis; COG, Cluster of Orthologous Group; EGID, Ensemble algorithm for Genomic Island Detection; GI, genomic island; HA-SA, human-associated *Staphylococcus aureus*; HGT, horizontal gene transfer; LA-SA, livestock-associated *Staphylococcus aureus*; MGE, mobile genetic element; PAI, pathogenicity island; R–M system, restriction–modification system; VFDB, Virulence Factors Database.

References

- Hacker, J. and Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641–679, <https://doi.org/10.1146/annurev.micro.54.1.641>
- Weinstock, G.M., Sodergren, E.J., Smajs, D. and Norris, S.J. (2000) Identification of virulence genes *in silico*: infectious disease genomics. *Virulence Mechanisms of Bacterial Pathogens*, 3rd, pp. 251–261, American Society of Microbiology
- Jain, R., Rivera, M.C., Moore, J.E. and Lake, J.A. (2002) Horizontal gene transfer in microbial genome evolution. *Theor. Popul. Biol.* **61**, 489–495, <https://doi.org/10.1006/tpbi.2002.1596>
- Dobrindt, U., Hochhut, B., Hentschel, U. and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2**, 414–424, <https://doi.org/10.1038/nrmicro884>
- Jackson, R.W., Vinatzer, B., Arnold, D.L., Dorus, S. and Murillo, J. (2011) The influence of the accessory genome on bacterial pathogen evolution. *Mob. Genet. Elements* **1**, 55–65, <https://doi.org/10.4161/mge.1.1.16432>
- Vasquez-Rifo, A., Veksler-Lublinsky, I., Cheng, Z., Ausubel, F.M. and Ambros, V. (2019) The *Pseudomonas aeruginosa* accessory genome elements influence virulence towards *Caenorhabditis elegans*. *Genome Biol.* **20**, 207, <https://doi.org/10.1186/s13059-019-1890-1>
- Lloyd, A.L., Henderson, T.A., Vigil, P.D. and Mobley, H.L. (2009) Genomic islands of uropathogenic *Escherichia coli* contribute to virulence. *J. Bacteriol.* **191**, 3469–3481, <https://doi.org/10.1128/JB.01717-08>
- Bentley, S.D. and Parkhill, J. (2004) Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* **38**, 771–791, <https://doi.org/10.1146/annurev.genet.38.072902.094318>
- Schmidt, H. and Hensel, M. (2004) Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* **17**, 14–56, <https://doi.org/10.1128/CMR.17.1.14-56.2004>
- Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069, <https://doi.org/10.1093/bioinformatics/btu153>
- Petkau, A., Stuart-Edwards, M., Stothard, P. and Van Domselaar, G. (2010) Interactive microbial genome visualization with GView. *Bioinformatics* **26**, 3125–3126, <https://doi.org/10.1093/bioinformatics/btq588>
- Che, D., Hasan, M.S., Wang, H., Fazekas, J., Huang, J. and Liu, Q. (2011) EGID: an ensemble algorithm for improved genomic island detection in genomic sequences. *Bioinformatics* **7**, 311, <https://doi.org/10.6026/007/97320630007311>
- Bezuidt, O., Lima-Mendez, G. and Reva, O.N. (2009) SeqWord Gene Island Sniffer: a program to study the lateral genetic exchange among bacteria. *World Acad. Sci. Eng. Technol.* **58**, 1169–1174
- Vernikos, G.S. and Parkhill, J. (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. *Bioinformatics* **22**, 2196–2203, <https://doi.org/10.1093/bioinformatics/btl369>
- Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W.F. et al. (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* **7**, 142, <https://doi.org/10.1186/1471-2105-7-142>
- Shrivastava, S., Reddy, C.V. and Mande, S.S. (2010) INDeGenIUS, a new method for high-throughput identification of specialized functional islands in completely sequenced organisms. *J. Biosci.* **35**, 351–364, <https://doi.org/10.1007/s12038-010-0040-4>
- Hsiao, W., Wan, I., Jones, S.J. and Brinkman, F.S. (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* **19**, 418–420, <https://doi.org/10.1093/bioinformatics/btg004>
- Tu, Q. and Ding, D. (2003) Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis. *FEMS Microbiol. Lett.* **221**, 269–275, [https://doi.org/10.1016/S0378-1097\(03\)00204-0](https://doi.org/10.1016/S0378-1097(03)00204-0)
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y. et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226, <https://doi.org/10.1093/nar/gku1221>
- Chen, L., Zheng, D., Liu, B., Yang, J. and Jin, Q. (2016) VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697, <https://doi.org/10.1093/nar/gkv1239>
- Ozer, E.A. (2018) ClustAGE: a tool for clustering and distribution analysis of bacterial accessory genomic elements. *BMC Bioinformatics* **19**, 150, <https://doi.org/10.1186/s12859-018-2154-x>
- Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G. et al. (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48–51, <https://doi.org/10.1126/science.1218198>
- Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245, <https://doi.org/10.1093/nar/gkw290>
- Pierneef, R., Cronje, L., Bezuidt, O. and Reva, O.N. (2015) Pre.GI: a global map of ontological links between horizontally transferred genomic islands in bacterial and archaeal genomes. *Database* **2015**, bav058, <https://doi.org/10.1093/database/bav058>
- Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., deBoy, R.T., Ravel, J. et al. (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J. Bacteriol.* **187**, 2426–2438, <https://doi.org/10.1128/JB.187.7.2426-2438.2005>
- Carver, T., Thomson, N., Bleasby, A., Berriman, M. and Parkhill, J. (2008) DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**, 119–120, <https://doi.org/10.1093/bioinformatics/btn578>

- 27 de la Cruz, F. and Davies, J. (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol.* **8**, 128–133, [https://doi.org/10.1016/S0966-842X\(00\)01703-0](https://doi.org/10.1016/S0966-842X(00)01703-0)
- 28 Novick, R.P. (2003) Mobile genetic elements and bacterial toxinoses: the superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid* **49**, 93–105, [https://doi.org/10.1016/S0147-619X\(02\)00157-9](https://doi.org/10.1016/S0147-619X(02)00157-9)
- 29 Contreras-Moreira, B. and Vinuesa, P. (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pan-genome analysis. *Appl. Environ. Microbiol.* **79**, 7696–7701, <https://doi.org/10.1128/AEM.02411-13>
- 30 Clarke, S.R. and Foster, S.J. (2006) Surface adhesins of *Staphylococcus aureus*. *Adv. Microb. Physiol.* **51**, 187–224, [https://doi.org/10.1016/S0065-2911\(06\)51004-5](https://doi.org/10.1016/S0065-2911(06)51004-5)
- 31 Herron-Olson, L., Fitzgerald, J.R., Musser, J.M. and Kapur, V. (2007) Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS ONE* **2**, e1120, <https://doi.org/10.1371/journal.pone.0001120>
- 32 Yoon, S.H., Park, Y.K. and Kim, J.F. (2015) PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. *Nucleic Acids Res.* **43**, D624–D630, <https://doi.org/10.1093/nar/gku985>
- 33 Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D. et al. (2009) Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *J. Bacteriol.* **191**, 3462–3468, <https://doi.org/10.1128/JB.01804-08>
- 34 Lindsay, J.A. and Holden, M.T. (2006) Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct. Integr. Genomics* **6**, 186–201, <https://doi.org/10.1007/s10142-005-0019-7>
- 35 Bastian, M., Heymann, S. and Jacomy, M. (2009) Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media, San Jose, California **8**, 361–362, 17/05/2009
- 36 Diep, B.A., Gill, S.R., Chang, R.F., Phan, T.H., Chen, J.H., Davidson, M.G. et al. (2006) Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet North Am. Ed.* **367**, 731–739, [https://doi.org/10.1016/S0140-6736\(06\)68231-7](https://doi.org/10.1016/S0140-6736(06)68231-7)
- 37 Hiramatsu, K., Ito, T., Tsubakishita, S., Sasaki, T., Takeuchi, F., Morimoto, Y. et al. (2013) Genomic basis for methicillin resistance in *Staphylococcus aureus*. *Infect. Chemother.* **45**, 117–136, <https://doi.org/10.3947/ic.2013.45.2.117>
- 38 Jani, M., Sengupta, S., Hu, K. and Azad, R.K. (2017) Deciphering pathogenicity and antibiotic resistance islands in methicillin-resistant *Staphylococcus aureus* genomes. *Open Biol.* **7**, 170094, <https://doi.org/10.1098/rsob.170094>
- 39 Moon, B.Y., Park, J.Y., Robinson, D.A., Thomas, J.C., Park, Y.H., Thornton, J.A. et al. (2016) Mobilization of genomic islands of *Staphylococcus aureus* by temperate bacteriophage. *PLoS ONE* **11**, e0151409, <https://doi.org/10.1371/journal.pone.0151409>
- 40 Burrus, V., Pavlovic, G., Decaris, B. and Guédon, G. (2002) Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.* **46**, 601–610, <https://doi.org/10.1046/j.1365-2958.2002.03191.x>
- 41 Omelchenko, M.V., Makarova, K.S., Wolf, Y.I., Rogozin, I.B. and Koonin, E.V. (2003) Evolution of mosaic operons by horizontal gene transfer and gene displacement *in situ*. *Genome Biol.* **4**, R55, <https://doi.org/10.1186/gb-2003-4-9-r55>
- 42 Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F. et al. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**, 1768–1770, <https://doi.org/10.1126/science.1122050>
- 43 Chen, N., Ou, H.Y., van Aartsen, J.J., Jiang, X., Li, M., Yang, Z. et al. (2010) The pheV phenylalanine tRNA gene in *Klebsiella pneumoniae* clinical isolates is an integration hotspot for possible niche-adaptation genomic islands. *Curr. Microbiol.* **60**, 210–216, <https://doi.org/10.1007/s00284-009-9526-4>
- 44 Imanishi, I., Nicolas, A., Caetano, A.B., de Paula Castro, T.L., Tartaglia, N.R., Mariutti, R. et al. (2019) Exfoliative toxin E, a new *Staphylococcus aureus* virulence factor with host-specific activity. *Sci. Rep.* **9**, 16336, <https://doi.org/10.1038/s41598-019-52777-3>
- 45 Wagner, P.L. and Waldor, M.K. (2002) Bacteriophage control of bacterial virulence. *Infect. Immun.* **70**, 3985–3993, <https://doi.org/10.1128/IAI.70.8.3985-3993.2002>
- 46 Sui, H.S.J., Fedynak, A., Hsiao, W., Langille, M. and Brinkman, F. (2009) The association of virulence factors with genomic islands. *PLoS ONE* **4**, e8094, <https://doi.org/10.1371/journal.pone.0008094>
- 47 Wan, P. and Che, D. (2014) A computational framework for tracing the origins of genomic islands in prokaryotes. *Int. Sch. Res. Notices* **2014**, <https://doi.org/10.1155/2014/732857>
- 48 Lin, W.S., Cunneen, T. and Lee, C.Y. (1994) Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J. Bacteriol.* **176**, 7005–7016, <https://doi.org/10.1128/JB.176.22.7005-7016.1994>
- 49 Dubey, G.P. and Ben-Yehuda, S. (2011) Intercellular nanotubes mediate bacterial communication. *Cell* **144**, 590–600, <https://doi.org/10.1016/j.cell.2011.01.015>
- 50 Dordet Frisoni, E., Marendá, M.S., Sagné, E., Nouvel, L.X., Guéillot, R., Glaser, P. et al. (2013) ICEA of *Mycoplasma agalactiae*: a new family of self-transmissible integrative elements that confers conjugative properties to the recipient strain. *Mol. Microbiol.* **89**, 1226–1239, <https://doi.org/10.1111/mmi.12341>
- 51 Ge, Y., Guo, G., Ge, B., Yin, H. and Yin, H. (2018) The spleen microbiota of small wild mammals reveals distinct patterns with tick-borne bacteria. *PLoS Negl. Trop. Dis.* **12**, e0006499, <https://doi.org/10.1371/journal.pntd.0006499>