NAuRA: Genomic Tool to Identify Staphylococcal Enterotoxins in *Staphylococcus aureus* Strains Responsible for FoodBorne Outbreaks

Déborah Merda[1*†], Arnaud Felten[1†], Noémie Vingadassalon[1], Sarah Denayer[2], Yacine Titouche[3], Lucia Decastelli[4], Bernadette Hickey[5], Christos Kourtis[6], Hristo Daskalov[7], Michel-Yves Mistou[1] and Jacques-Antoine Hennekinne[1]

[1] French Agency for Food, Environmental and Occupational Health & Safety (ANSES), University of Paris-Est, Maisons-Alfort, France, [2] Scientific Service of FoodBorne Pathogens, Sciensano, Brussels, Belgium, [3] Laboratory of Analytical Biochemistry and Biotechnology, University of Mouloud Mammeri, Tizi Ouzou, Algeria, [4] National Reference Laboratory for Coagulase-Positive Including Staphylococcus aureus, Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d'Aosta, Turin, Italy, [5] Department of Agriculture, Food and the Marine, Kildare, Ireland, [6] State General Laboratory, Food Microbiology Laboratory, Nicosia, Cyprus, [7] National Center of Food Safety, NDRVI, BFSA, Sofia, Bulgaria

Food contamination by staphylococcal enterotoxins (SEs) is responsible for many food poisoning outbreaks (FPOs) each year, and they represent the third leading cause of FPOs in Europe. SEs constitute a protein family with 27 proteins. However, enzyme immunoassays can only detect directly in food the five classical SEs (SEA-SEE). Thus, molecular characterization methods of strains found in food are now used for FPO investigations. Here, we describe the development and implementation of a genomic analysis tool called NAuRA (Nice automatic Research of alleles) that can detect the presence of 27 SEs genes in just one analysis- and create a database of allelic data and protein variants for harmonizing analyses. This tool uses genome assembly data and the 27 protein sequences of SEs. To include the different divergence levels between SE-coding genes, parameters of coverage and identity were generated from 10,000 simulations and a dataset of 244 assembled genomes from strains responsible for outbreaks in Europe as well as the RefSeq reference database. Based on phylogenetic inference performed using maximum-likelihood on the core genomes of the strains in this collection, we demonstrated that strains responsible for FPOs are distributed throughout the phylogenetic tree. Moreover, 71 toxin profiles were obtained using the NAuRA pipeline and these profiles do not follow the evolutionary history of strains. This study presents a pioneering method to investigate strains isolated from food at the genomic level and to analyze the diversity of all 27 SE-coding genes together.

Keywords: staphylococcal enterotoxins genes, gene detection, variant diversity, bioinformatics tool, genomic

INTRODUCTION

Staphylococcus aureus strains can be responsible for staphylococcal food poisoning by secreting enterotoxins in food matrices. The symptoms caused by the consumption of food contaminated by staphylococcal enterotoxins (SEs) include nausea, vomiting, abdominal cramping, and diarrhea. Symptoms appear quickly after ingestion, between 30 min and 8 h. SEs represent the third leading cause of food poisoning outbreaks (FPOs) in Europe, and they are the second leading FPO cause in France after Salmonella (Kérouanton et al., 2007). Since 2010, the number of FPOs caused by bacterial toxins have significantly increased in France (EFSA and ECDC, 2019). In 2018, bacterial toxins caused 799 FPOs in France and 935 in Europe, of which 114 FPOs were caused by SEs (EFSA and ECDC, 2019).

Currently, 27 SEs have been described (SEA to SEE and SEG to SElX, SElY, SElZ, SEl26, SEl27) and they compose a superfamily of secreted single-chain globular proteins whose molecular weights vary from 19 to 29 kD. After ingestion, SEs have two major toxic activities: (i) a neurotoxic activity that activates the vagus nerve and the emetic center of the brain, triggering vomiting reflexes, (ii) a superantigenic activity leading to the non-specific activation of T lymphocytes, causing a strong fever. Although all SEs have a superantigenic function, not all have emetic functions (Hennekinne et al., 2012). Tests for emetic functions have been performed on 18 SEs in *Suncus murinus* and primates (Omoe et al., 2013; Ono et al., 2017).

SEs are coded by genes of about 700–800 bp, localized on mobile genetic elements (MGEs), such as plasmids, prophages or pathogenicity islands (SaPIs) (Bania et al., 2006; Driebe et al., 2015). Within the same MGE, several SE-coding genes can be found, i.e., the plasmid pF5 harbors the genes coding for SER, SET, SES, and SElJ (Argudín et al., 2010). The genomic island carrying the enterotoxin gene cluster (*egc*) frequently harbors five genes coding for SEG, SEI, SEM, SEN, and SEO. Some variants of this genomic island have the gene coding for SElV but not the genes coding for SEI and SEM (Argudín et al., 2010).

All genes coding for enterotoxins may originate from the *egc*, which is considered as an enterotoxin nursery and can diverge from *egc* by duplication, transposition and mutation (Jarraud et al., 2001; Grumann et al., 2014). Thus, these evolutionary mechanisms have led to several genetic clusters in enterotoxin genes (three reported by Jarraud et al., 2001 and five by Ono et al. (2008), including genes belonging to the *egc* cluster. Furthermore, SEs present different levels of divergence (Ono et al., 2008). For example, SEA and SEE share 81% of homology with regard to their amino-acid sequences, whereas SEB and SEC share 67% of homology (Balaban and Rasooly, 2000).

To detect SE genes, it is necessary to consider the different divergence levels that can best define possible toxigenic profiles and thus help identify SEs implicated in an outbreak. In a FPO investigation, screening for virulence factors is very important. SE genes are currently detected in strains using PCR method developed by the European Reference Laboratory for Coagulase Positive Staphylococci (EURL CPS) and described in Roussel et al. (2015). This method has been used to investigate Italian and Belgian outbreaks (Bianchi et al., 2014; Carfora et al., 2015;

Denayer et al., 2017; Monistero et al., 2018). This method can only detect 11 SE genes. Therefore, no information is available on the 16 other genes for FPO investigations. Other PCR methods have been developed to detect 18 of the 27 SE genes (e.g., Zhang et al., 2018). However, the PCR approach has several limitations. First, it does not provide any information on nucleotide variability; second, the amplification of pseudogenes leads to false positives; and finally, mutations in the primer binding sites can produce false negatives.

Genomic approaches to detect virulence factors are under development. Advances in whole genome sequencing (WGS) make it possible to access the complete toxin repertoire of a given strain. WGS has been used for several years for monitoring purposes or for outbreak investigations on several pathogens, because this method offers an alternative to other molecular biology methods, such as pulsed-field gel electrophoresis (PFGE) or PCR (Oakeson et al., 2017). Bioinformatics pipelines have been set up by European Reference Centers, e.g., for *Neisseria meningitidis* (Bogaerts et al., 2019).

A specific pipeline called Staphopia was developed for *Staphylococcus* genomic analysis, but no tool for enterotoxin detection has been implemented in it (Petit and Read, 2018). However, for FPO investigations, it is important to identify the enterotoxin-coding genes to establish the link between the SE found in the incriminated food and the strain, and to study other SEs potentially present in food for which immunoassay methods have not been developed. To detect virulence factors in genomic sequences two approaches can be used: (i) mapping reads to reference sequences, e.g., as implemented in ARIBA (Hunt et al., 2017) and (ii) local alignment using BLAST. A benchmark study based on *S. aureus* strains has shown that the choice of method (mapping or local alignment) has no impact on the results (Mason et al., 2018). When genome assemblies are available, the BLAST method is faster than mapping and, with public databases, assembled genomes are more accessible than raw data. Therefore, the objective of this work was to develop a bioinformatics pipeline called NAuRA (Nice Automatic Research of Alleles) to detect genes coding for SEs in *S. aureus* genomes. NAuRA can screen every new genome for its toxin content and toxin allele profile. Moreover, NAuRA can create an allelic database to harmonize protein variant nomenclature of SEs, and to facilitate information sharing between outbreak cases. NAuRA was implemented to study enterotoxin-coding gene families in *S. aureus* strains responsible for FPOs in Europe between 2005 and 2017.

MATERIALS AND METHODS

Bioinformatics Development

An in-house workflow NAuRA was developed in Python to automate and harmonize the screening for genes of interest in a collection of genomes (**Figure 1**). Based on BLAST, NAuRA detects the presence of genes/proteins from a user-defined query list in the user-defined collection of genomes and provides an allele identifier when a matching gene/protein is found. Every allele-type sequence having a new identifier is added to the database of queries. Furthermore, the workflow offers

the possibility to perform a quick clustering analysis using a neighbor-joining approach based on allele sequences. NAuRA is compatible with both nucleic and protein sequences and is available on github[1].

From genomes in GenBank format and a list of query genes of interest in FASTA format, NAuRA converts GenBank files to FASTA nucleic acid (fna) or FASTA amino acid (faa) files where each sequence represents an open reading frame (ORF) previously predicted by Prokka. Then, sequences of interest are aligned to the chosen genomes using BLAST + (Camacho et al., 2009) and the results are filtered with two thresholds: (1) the minimum percentage of similarity and (2) the minimum percentage of coverage. If a gene is considered as present, i.e., if the two percentages of similarity and coverage are greater than or equal to the two thresholds, the ORF is extracted and compared to all known alleles using BLAST. If a new allele is detected, i.e., if a mutation in an amino acid is detected, it is added to the corresponding query FASTA file and a new identifier is assigned. After the alignment step, NAuRA outputs for each genome a gene presence/absence matrix in which the allele identifier is specified. NAuRA offers the option to align all alleles of each query using Clustal Omega (Sievers et al., 2011) to construct a phylogenetic tree using the neighbor-joining method implemented in Phylip (Felsenstein, 2013). Node support values are calculated using the bootstrap method. The consensus tree is obtained using the sumtree.py script from the DendroPy Library (Sukumaran and Holder, 2010).

To analyze a gene family with an unknown level of divergence, a second Python script was developed, NAuRA_BPF (**Figure 1B**). To optimize the setting of threshold values, BLAST analysis was performed with a series of coverage and similarity thresholds randomly chosen in uniform distribution. The best analyses corresponding to specific threshold values were selected using a monophyly test, which assumes that all alleles belonging to a given gene family form a monophyletic branch on a phylogenetic tree.

Strain Collection and DNA Extraction

A collection of 143 genomes belonging to *S. aureus* and the cause of FPO in Europe from 2005 to 2017 were collected by the EURL CPS and conserved at −80°C. This collection represented the known diversity of strains isolated from food in Europe and was typed using PFGE; some were reference strains (**Supplementary Table S1**). For DNA extraction, strains were regenerated on Milk Plate Count Agar for 24 h at 37°C. To purify the culture, one colony was placed on brain heart infusion (BHI) agar for 16–20 h at 37°C. DNA extraction was performed using the Wizard® Genomic DNA Purification kit (Promega) from the colonies on BHI agar. A pre-lysis step was added with a treatment of 105 μl of lysozyme and 15 μl of lysostaphin for 30 min at 37°C. DNA quality was evaluated using a NanoDrop spectrophotometer and the concentration was quantified using a Qubit fluorometer. DNA integrity was evaluated by electrophoresis on 0.8% agarose gel.

Genome Sequencing, Assembly, and Annotation

Genomes were sequenced using Illumina NextSeq system (ICM institute, Paris, France) and the Nextera XT DNA Library Prep Kit was used to construct libraries. The sequencing data were presented as reads of 2*150 bp. Before assembly, reads were normalized using BBnorm[2] to reduce coverage to 100x and trimmed using Trimmomatic (Bolger et al., 2014) to remove bases having a Phred score of less than 30 and to remove reads smaller than 50 bp. Assembly was performed in three steps. First, the *de novo* assembly was processed using Spades with default parameters. Then, the nearest complete public genome of *S. aureus* was found using Mash (Ondov et al., 2016) and used to perform the scaffolding step in MeDuSa (Bosi et al., 2015). Finally, a gap-closing step was processed with GMcloser (Kosugi et al., 2015). Genome assemblies were evaluated using Quast (Gurevich et al., 2013) and scaffolds smaller than 200 bp were removed. Genome annotation was performed using Prokka (Seemann, 2014). Raw reads and assembly data are available in the European Nucleotide Archive (ENA) under the study number PRJEB36867.

Publicly available genomes were also added to the dataset: 101 assembly genomes available from the RefSeq database in NCBI. They were re-annotated using Prokka to ensure a homogeneous dataset. Multilocus sequence typing (MLST) was performed on each sample, the sequence type (ST) was assigned using the MLST program[3], and the database PubMLST[4] and the scheme of seven housekeeping genes (*arcC, aroE, glpF, gmk, pta, tpi,* and *yqiL*) developed for *S. aureus* (Enright et al., 2000).

Thus, our total collection included 101 genomes from public databases and 143 genomes sequenced in-house. For strains sequenced in-house, they were isolated in Europe; 103 were potentially responsible for an FPO, 14 were reference strains, 15 were isolated from self-testing and 10 had been used in a previous study (Titouche et al., 2019). The quality of sequencing, assembly and annotation was robust: the coverage of reads was on average 329x (**Supplementary Table S1**). Regarding the assembly, the expected size of *S. aureus* genomes was found (on average 2.71 Mb) and the number of contigs varied between 1 and 69. The dataset annotation was harmonized using the same software (i.e., Prokka), and the number of annotated genes varied between 2248 and 2751.

Screening for Genes of Interest

We screened for the previously described 26 enterotoxins- and TSST1-coding genes using protein sequences as queries on the 143 *S. aureus* genomes isolated from food matrices and on the 101 publicly available genomes (**Supplementary Table S1**). The accession numbers of these genes are given in **Supplementary Table S1**. Reviewed protein sequences available in public databases were chosen. For the other SEs for which no reviewed sequences were available, their genomic localization, and for two new SEls, *sel26,* and *sel27*, the protein sequences

---

[1]https://github.com/afelten-Anses/NAuRA

[2]https://jgi.doe.gov/data-and-tools/bbtools/
[3]https://github.com/tseemann/mlst
[4]https://pubmlst.org/saureus/

FIGURE 1 | Analytical schematic of staphylococcal enterotoxin (SE) gene detection. **(A)** NAuRA (Nice Automatic Research of Alleles). **(B)** Determination of the best parameters to filter the BLAST results.
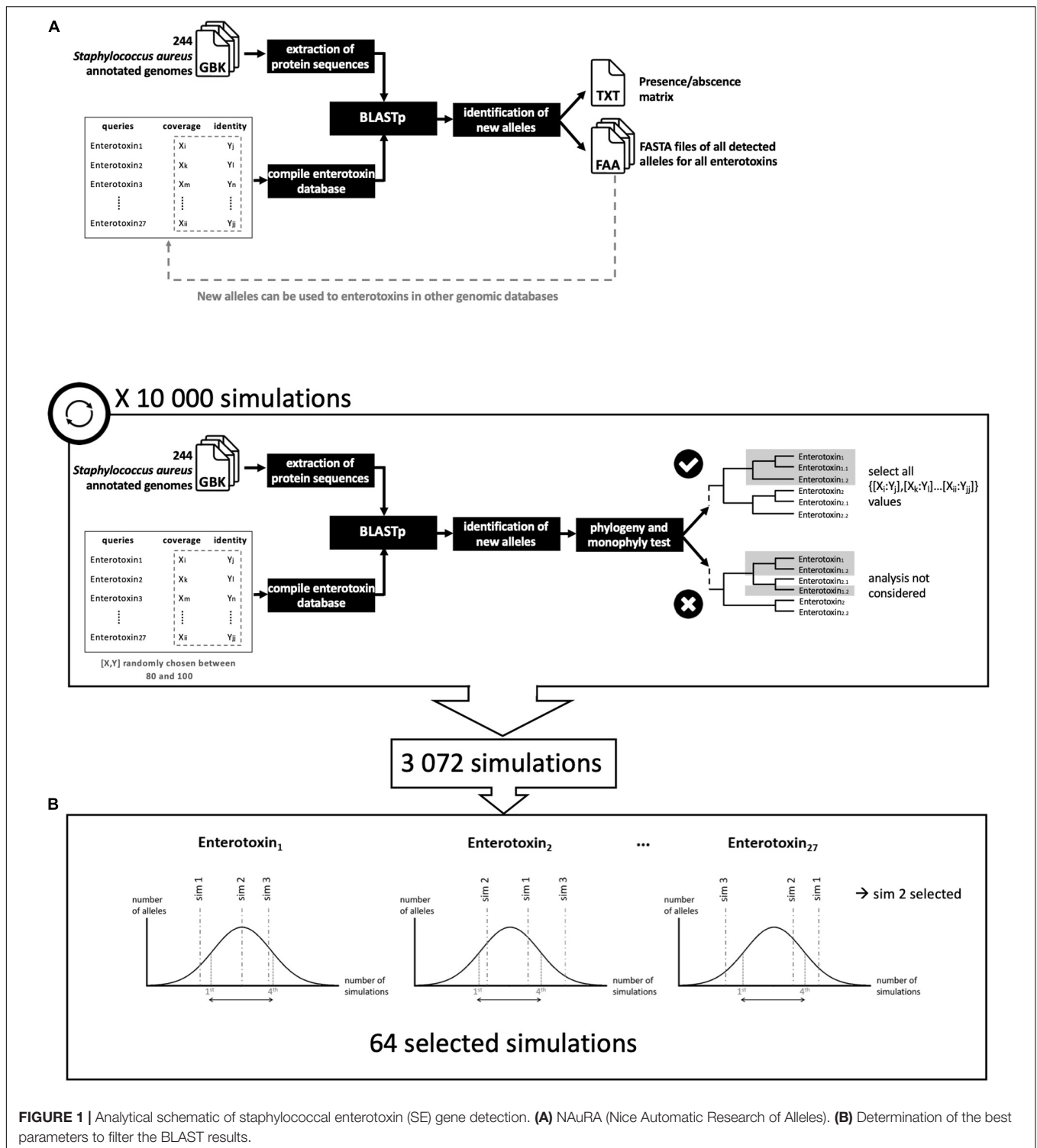
were taken from the literature (Zhang et al., 2018). Because these genes are genetically close, parameter adjustment was performed to discriminate all alleles of each gene using the Python script NAuRA_BPF. Coverage and similarity parameters were tested for values between 80 and 100 to perform 10,000 simulations. Among these simulations, monophyletic groups were found for each gene

in 3072 simulations, the other 6928 analyses were removed. The best analysis was selected to maximize allelic diversity for each gene. For each gene, a distribution of the number of different alleles was established using the 3072 selected simulations. A random choice among analyses providing a number of alleles between the first and fourth quartile of the distribution was

performed. This procedure conserves the diversity of alleles and reduces the stringency of the selection parameters. Nucleotide sequences were extracted from GenBank files, and using BLAST best thresholds result obtained by NAuRA for the best analysis.

Determination of Thresholds for Detecting Staphylococcal Enterotoxins

NAuRA is a Python pipeline to detect sequences of interest in genomes with a BLAST approach. It was developed to mainly detect SE genes. With the different divergence levels, simulations were performed to optimize filter parameters of identity and coverage. Accordingly, among the 10,000 simulations, 64 were conserved after the monophyly test (**Figure 1**). The values of the parameters used for the 64 simulations covered the whole uniform distribution between 80 and 100 for identity and coverage parameters, except for the four following genes: *sea*, *see*, *sep,* and *selv*. This means that the 64 selected simulations were based mainly on these four genes; for this reason, for the final analysis, the lowest threshold to maintain the allele monophyly was 80% of identity and coverage for the 23 SE genes, 90% of identity for *selv* and 84% for identity for *sea*, *sep,* and *see*. These values confirmed the low divergence levels between *sem* and *selv* and between *sea*, *sep,* and *see.* On the **Supplementary Figure S1**, three phylogenies based on protein variants were presented and corresponded to three situations: (i) released parameters, (ii) stringent parameters, and (iii) selected parameters. The released parameters were percentage of identity inferior to 84% for all genes (80% of identity). The stringent parameters were percentage of identity superior to 84% for all genes (between 95 and 100%). And the selected parameters were parameters described before. With released parameters, no differentiation between *sea*, *see,* and *sep* variants was observed (**Supplementary Figure S1A**). With stringent parameters, some genes, such as *seo* and *selu*, exhibit few variants, with the consequence to increase the false-negative results (**Supplementary Figure S1B**). The phylogeny based on variants obtained with selected parameters described below allows to maximize diversity while maintaining monophyly (**Supplementary Figure S1C**).

Comparisons With PCR Method and Toxin Detection

Profiles obtained by NAuRA were compared to PCR-detected genes according to the methods described in Roussel et al. (2015), This method allows the detection of 11 SE-genes (*sea*, *seb*, *sec*, *sed*, *see*, *seg*, *seh*, *sei*, *selj*, *sep*, *ser*). PCR results were available for each strain of the genome collection sequenced for this study., Toxin detection in food was performed according the standard ISO 19020 for the "classical" toxins (SEA to SEE), and results were available for 54/143 strains. These results were also compared to profiles obtained by NAuRA.

Core Genome Definition and Phylogenetic Analyses

The core genome of the 244 strains, which is defined by all genes shared by all genomes in only one copy, was determined using

Roary (Page et al., 2015) applied on the GFF3 files obtained in Prokka. Based on the pBLASTn approach, Roary extracts the pan-genome and aligns all its ORF sequences using MAFFT (Katoh et al., 2002). To remove the misaligned regions, the Gblock program (Castresana, 2000) was used on this alignment. To obtain a robust phylogenetic tree with the maximum-likelihood method, iQtree (Chernomor et al., 2016) was run. The node support values were calculated using the bootstrap method. Bootstrap values were calculated with 1000 replicates and a bootstrap convergence test was carried out to check if this number of replicates was sufficient. The representation of this tree was obtained using the iTOL web viewer. The enterotoxin presence/absence coding genes and FPO information was added on the phylogenetic tree.

Cophenetic and Robinson Foulds Metric

Comparison of phylogenetic tree based on core genome and dendrogram based on enterotoxin profiles was performed using the cophenetic index. This index could vary between −1 for perfect negative correlation and 1 for perfect positive correlation. A cophenetic index of 0 indicates an absence of correlation between topologies. Then, this index allows evaluating the correlation between the core genome and the toxigenic profile. The Robinson Foulds (RF) metric was also calculated, and this metric indicates the number of distances between topologies. The cophenetic index and the RF metric were calculated with the R packages *phangorn* and *dendextend* (Cardona et al., 2013).

RESULTS

Description and Justification of the Dataset

Based on the core genome, the dataset was representative of the known diversity of *S. aureus* species. Furthermore, MLST analysis was performed on all sequences and 62 different STs were found in the collection, although it was not possible to determine the ST for 22 strains. The clonal complex was estimated for these strains. Among the 62 different STs, the 10 most frequent STs found in 40,000 genomes of *S. aureus* (Petit and Read, 2018) were the following: ST22, ST8, ST5, ST239, ST398, ST30, ST45, ST15, ST36, and ST105. However, some STs frequent in food, which were not detected in Petit and Read (2018) study were also identified, e.g., ST1 and ST389.

Strains involved in FPOs formed a paraphyletic group (starred in **Figure 2**). In the maximum likelihood tree, strains responsible for FPOs were distributed throughout the tree and were mixed with strains isolated from humans or environment (**Figure 2**). For example, strains occurring in France in 2017 (17SBCL13STA, 17SBCL08STA, 17SBCL09STA) constituted a monophyletic group with genomes from the RefSeq database and also with strains responsible for FPOs in Bulgaria in 2017 (17SBCL580, 17SBCL585, and 17SBLC586) and in Belgium in 2013 (13SBCL319STA and 13SBCL320). This pattern suggests that strains are not structured according the country of isolation or year of isolation. For

instance, strains isolated in 1997, 2011, 2013, and 2015 belonging to ST8, are monophyletic with genomes from the RefSeq database.

Toxin Profiles of Strains

Toxin profiles obtained by NAuRA are highly in accordance at 93.7% with profiles obtained by PCR approach. Indeed, only nine different profiles out of 143 were observed between the two methods. Of the nine profiles, eight corresponded to a gene detected by PCR but not by NAuRA, four of them implicated *seg* gene, two of them implicated *sed* gene, one implicated *sec* gene, and the last implicated two genes: *sea* and *seh*. The 9th different profile corresponded to no detection of *sep* gene by PCR. Concerning the comparison of NAuRA results and toxin detection in food, eight results were discordant from 54 available results. Then, 85.2% of results were in accordance. Among these eight discordances, six corresponded to no toxin detection whereas *se* gene was detected by NAuRA and PCR methods, and the other two corresponded to SEC detection, whereas *sec* gene was not detected in the strain, suggesting presence of other strains in the food.

The gene distribution did not follow the phylogeny (**Figure 2**) as expected for genes found on mobile genetic elements (MGE); subject to horizontal gene transfer. The cophenetic index comparing the phylogeny based on the core genome and the dendrogram based on toxin profiles was 0.25. The closer this index is to 0, the more the two compared trees are statistically different. The symmetric difference, or the Robinson Foulds metric, was 436, suggesting that 436 branches differed between the phylogeny and the dendrogram. This index does not take branch length into account, and only compares the topologies. The higher the number of this metric, the more the topologies differ.

The *egc* cluster (in red in **Figure 2**) was present in several groups of strains: 36% of strains had the *egc* cluster and it is absent in the two monophyletic groups, suggesting the loss of this cluster during the evolutionary history of *S. aureus*. Regarding genes located on plasmids, several toxin profiles were found within the same strain group. For example, the strains 11CEB284STA and 47A clustered together, but only one of them had the *sed*, *selj*, and *ser* genes and the other had *seb*, *sek*, and *seq*.

No difference in toxin profiles was observed between strains responsible for FPO and the other strains. Strains from the RefSeq database had the same toxin profile as FPO strains. For example, 15SBCL1507STA and GCA_00262835 both showed the *egc* cluster and the SElX-coding gene; the 13CEB176STA and GCA_001350955 strains had the genes coding for SEB, SEK, and SEQ. As no toxin profile can be shared by environmental strains and FPO, this could suggest that food contamination occurs via environmental strains.

A large number of toxin profiles were found (71 profiles, of which 34 were found only once) (**Figure 3**). The *selx* gene was found in the majority of profiles (53/71), as well as the *egc* cluster (39/71). Some genes were never found together.

Indeed, the *ser*, *sed*, and *selj* genes localized on plasmid were never found with the *seq*, *sek*, and *seb* genes localized on SaPI. This non-association suggests that strains cannot carry both MGEs simultaneously. Moreover, the *sea*, *see*, and *sep* genes which are genetically close, were never found together within the same strain.

The number of enterotoxins for each strain varied between 0 and 11 (**Figure 2** and **Supplementary Table S2**). Overall, 15.9% (39/244) of strains had only one enterotoxin corresponding mainly to SElX, 21% strains had at least five enterotoxins corresponding to the *egc* cluster and the SElX coding gene. Gene frequency in the collection varied among the different SE-coding genes. For example, 86% (210/244) of strains had the SElX-coding gene and 30% (74/244) of strains carried the gene coding for SEA. Concerning genes of *egc* cluster, 34% carried the gene coding for SEG (83/244), 34.8% (85/244) carried the gene coding for SEM, 36.4% (89/244) carried the gene coding for SEN and 36.4% (89/244) carried the gene coding for SEO. But genes coding for SES and SET were only present in three and two strains respectively and the gene coding for SElV was absent in the collection.
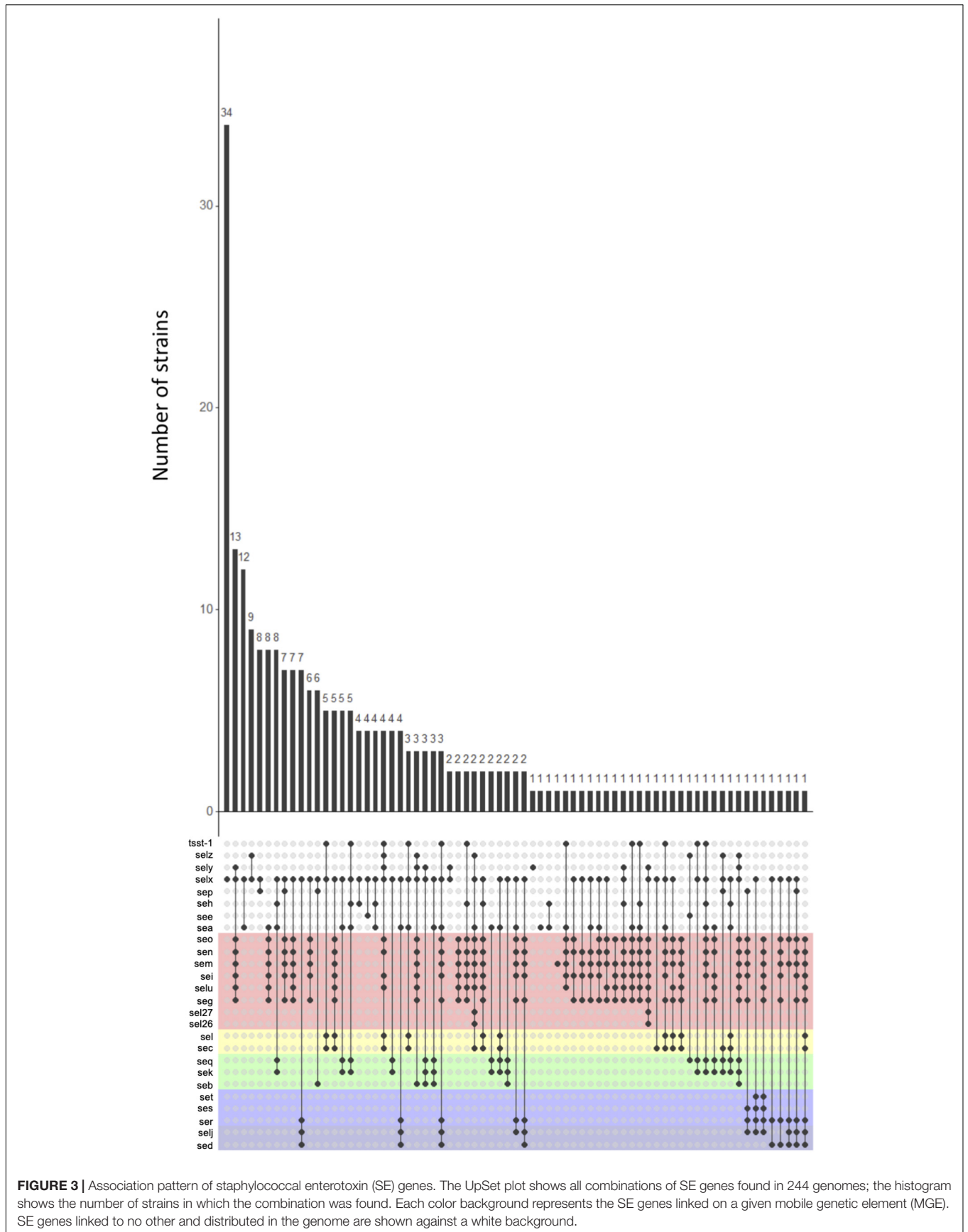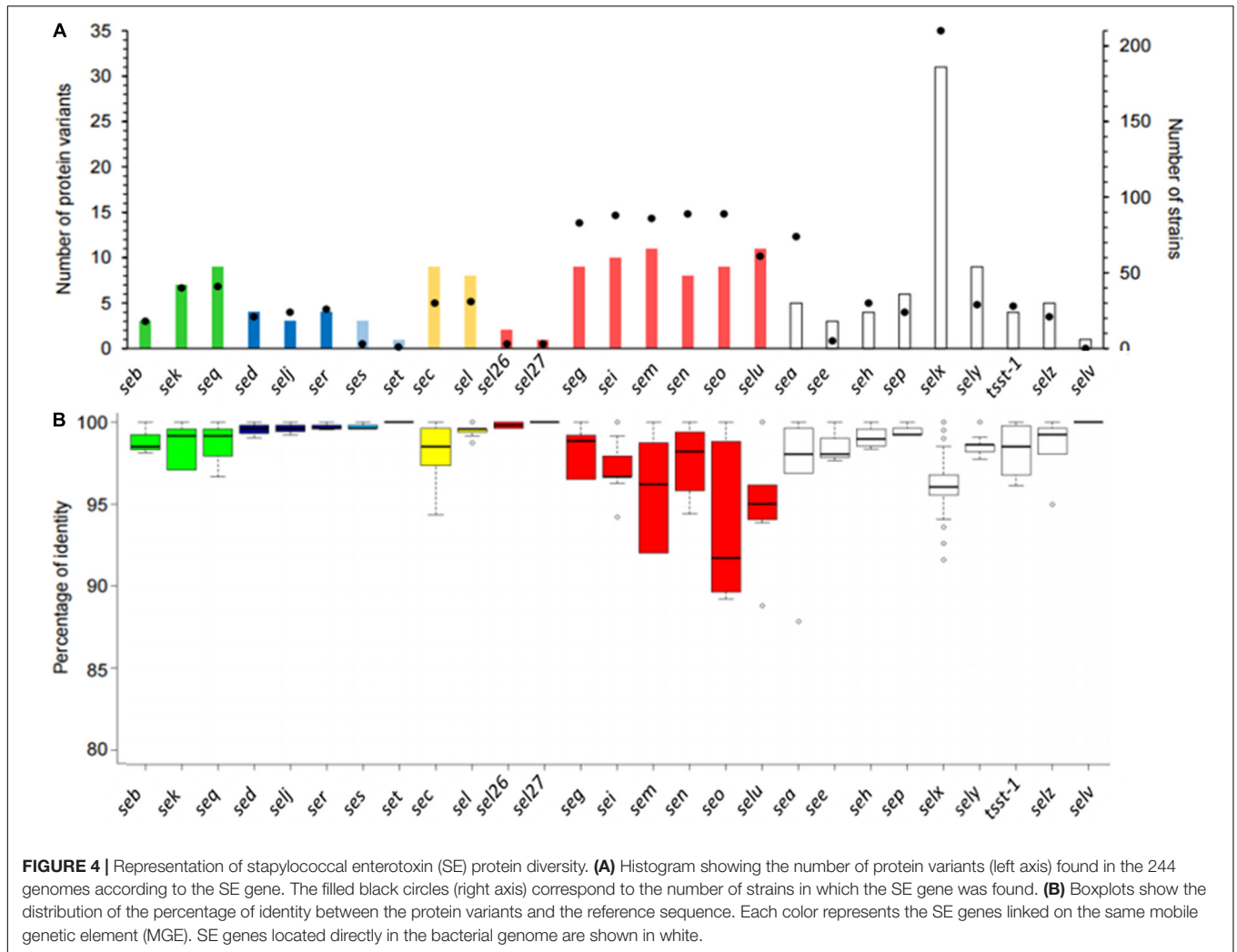
Enterotoxin Sequence Analyses

The number of variants differed according to the SEs (**Figure 4**). The number of protein variants varied between 1 and 11, except for the SElX-coding gene for which 31 protein variants were found (**Figure 4**). For the *egc* cluster comprising *seg*, *sei*, *sem*, *sen*, and *seo*, the same magnitude of protein. A similar number of variants in SED/SElJ pair localized on pIB485, and the SEC/SEL pair localized on SaPIbov was observed. It is also possible for genes localized on the same plasmids to have a different number of variants. For example, one protein variant was found in the dataset for SET, whereas four variants were found for the other toxins localized on the same plasmid (SElJ, SER and SES).

In all, 16 SE genes (*sea*, *seb*, *see*, *seg*, *seh*, *sei*, *selj*, *sem*, *sen*, *seo*, *sep*, *ser*, *set*, *selz*, *sel26*, and *sel27*) had one dominant protein variant in the dataset (at least 50% of strains), whereas no dominant protein variant was observed for the 11 other SE genes (*sec*, *sed*, *sek*, *see*, *seq*, *ses*, *selu*, *selx*, *sely*). Therefore, the predominance of some alleles may facilitate the choice of reference sequences for detection methods in food.

DISCUSSION

SEs are responsible for FPOs and their presence in food is regulated by European legislation (Ostyn et al., 2010). If, for example, SEs are found in a 25 g sample of cheese, this contaminated food cannot be marketed, potentially representing high economic losses for the food company. Therefore, it is important to set up a robust and sensitive methods to detect enterotoxins. Currently, only five enterotoxins are directly detectable in food. To help outbreak investigation and establish a link between toxins and strains, molecular biology methods have been developed to detect SE-coding genes in strains, mainly using PCR. These methods are costly in terms of development and laboratory time. On

Tree scale: 0.1

FIGURE 2 | Phylogeny of *Staphylococcus aureus* and distribution of staphylococcal enterotoxin (SE) genes. The phylogenetic inference based on the core genome was obtained using Roary on 244 of *S. aureus* genomes. The stars indicate strains involved in food poisoning outbreaks (FPOs). Each filled circle indicates the presence of SE genes; each color indicates SE genes linked on the same mobile genetic element (MGE). SE genes located on plasmid pIB485 are shown in blue; SE genes located on the same pathogenicity island in yellow or green; SE genes located on the *egc* cluster in red and all other genes not located on an MGE in black.

FIGURE 3 | Association pattern of staphylococcal enterotoxin (SE) genes. The UpSet plot shows all combinations of SE genes found in 244 genomes; the histogram shows the number of strains in which the combination was found. Each color background represents the SE genes linked on a given mobile genetic element (MGE). SE genes linked to no other and distributed in the genome are shown against a white background.

FIGURE 4 | Representation of stapylococcal enterotoxin (SE) protein diversity. **(A)** Histogram showing the number of protein variants (left axis) found in the 244 genomes according to the SE gene. The filled black circles (right axis) correspond to the number of strains in which the SE gene was found. **(B)** Boxplots show the distribution of the percentage of identity between the protein variants and the reference sequence. Each color represents the SE genes linked on the same mobile genetic element (MGE). SE genes located directly in the bacterial genome are shown in white.

the other hand, in this genomics era, both nucleotide and protein sequences are available for a specific strain. In contrast to toxin detection, molecular biology methods and genomic require strain cultivation and DNA extraction before analysis. Genomic and molecular biology methods were strongly in accordance, and differences were observed in particular for *seg* gene. This cluster is submitted to recombination events, and *se* gene fragments were found, suggesting pseudogeneization, false positive results in PCR. More differences were observed between gene content and toxin detection in food. This could be explained by the fact that not all the required conditions for toxin secretion were present. Indeed, SEs would not all secrete at the same time during the growth (Derzelle et al., 2009).

NAuRA is the first pipeline that can detect all 27 currently described SE genes and the simulation parameter adjustment can establish specific parameters for genetically close genes. The Staphopia pipeline, specifically developed for *S. aureus*, detects only nine SE genes available in the VFDB 2016 database (Chen et al., 2015) (*sea, seb, sec, sed, see, seh, sek, seq, tsst-1*), because this pipeline was mainly developed to analyze antibiotic

resistance (Petit and Read, 2018). SeqSphere software was mainly developed to analyze core-genome MLSTs and can be also used to detect 18 enterotoxins (Strauß et al., 2017). However, the task template for *S. aureus* and based on sequences of a microarray system does not distinguish sea from sep genes, thus considered as the same gene.

It is very important to detect all SE-coding genes and to describe all existing SEs to assess the prevalence of toxigenic strains. An increase in the number of toxins detected may mean an increase in the percentage of toxigenic strains (Le Loir et al., 2003; Bania et al., 2006; Aydin et al., 2011). For example, the number of toxigenic strains identified in our study (97.5% of strains) is higher than that reported in precedent studies: although based on the genes coding for the five classical SE genes (SEA to SEE), the toxigenic strain frequency was around 60% in several countries, such as Turkey, Portugal and Italy (62.6, 60.1, and 59.8%, respectively). In Iran and Louisiana (United States), the detection of nine SEs showed that 77.6 and 84.9%, respectively, of strains were toxigenic in these regions (Pu et al., 2010; Mashouf et al., 2015). In another study, the SElX-coding gene is present in 95% of

strains, and it was localized in the core genome (Wilson et al., 2011). Thus, screening for SElX-coding genes in the dataset can reveal a high percentage of toxigenic strains. Currently, only the five classical SEs are detected directly in food for FPO investigations, and the toxigenic power of non-detected strains is not known.

Among the five genes coding for the classical SEs, the genes encoding the SEA and SEC toxins were the most frequent. This distribution is similar to that in a Jordanian study where these genes were detected by PCR (Naffa et al., 2006). However, the prevalence of *sea* is higher than *sec* in our study. The SEA-coding gene may indicate a human origin for the strain, whereas the SEC-coding gene may indicate a bovine origin for the strain. Therefore, strains responsible for FPOs in Europe likely have a human origin, as expected for FPOs (Le Loir et al., 2003; Ortega et al., 2010), because food contamination by *S. aureus* can often be retraced to hygiene problems.

For FPO investigations, it is important to detect all SE-coding genes in the *egc* cluster so as to avoid underestimating its prevalence. There are several profiles of *egc* in strains responsible for FPOs. The obtained profiles for the *egc* cluster corresponded to the profiles reported in Song et al. (2016), with only four additional profiles being found just once. This diversity confirms that *seg* and *sei* genes are not always linked, as shown in Adwan et al. (2013). Our frequencies of *egc1* and *egc2*, as defined by Song et al. (2016), were similar to those obtained in their study with 28 and 55%, respectively. These values correspond to the frequencies obtained for strains isolated from food and corroborate our data, because a large part of strains were also isolated from food. Likewise, the *egc* cluster profile shows higher diversity in strains isolated from food than other environments (Song et al., 2016). NAuRA provides access to the complete nucleic diversity of a specific dataset. This can be useful for the development of new, more efficient PCR methods, for example, by designing primers preferentially in the more conserved region of allele sequences. The study of nucleic diversity can also help to choose the reference protein sequence to use for immunoassays methods, for example. In the UniProt database, only nine SEs are annotated and can serve as a standard for method development (The UniProt Consortium, 2018). Moreover, to harmonize analyses, NAuRA can create a database of newly identified variants to include them in subsequent analyses.

The complete *egc2* cluster is the most frequent *egc* variant and is distributed throughout the phylogenetic tree based on the core genome, as revealed in trees based on MLVA and PFGE. In contrast, the *egc1* cluster formed a monophyletic group on the core-genome tree, but was scattered in Song et al. (2016) study, based on MLVA and PFGE on SEs. Several evolutionary scenarios may be responsible for this distribution pattern. The first scenario involves an *egc* cluster without the *selu* gene and several acquisitions at different phylogenetic levels. The second scenario involves a complete ancestral *egc* cluster and subsequent loss of genes during evolution. The latter scenario is the most parsimonious hypothesis, because it involves only one evolutionary event. This scenario also provides support for the

hypothesis according to which the *egc* cluster is an SE nursery (Jarraud et al., 2001; Grumann et al., 2014) and a highly dynamic region subject to duplication and transposition events.

The number of toxin profiles obtained was lower than the number reported in China (71 versus 120, respectively) (Chao et al., 2015). This difference can be explained by the fact that strain diversity in China is higher than the diversity in our genome collection or by PCR limitations, as PCR detects only a small gene portion whereas NAuRA detection is performed on the predicted proteome by structural annotation However, the presence of a gene in a genome does not necessarily imply its expression. In the case of FPOs, bacterial strains make food unfit for consumption by directly secreting the toxin into the food (Fisher et al., 2018). Therefore, linking a strain to a toxin requires transcriptomic methods to verify enterotoxin gene expression.

Finally, NAuRA may be sensitive to sequence assembly biases due to a large number of contigs (fragmented assembly). Thus, poor-quality assemblies can reduce the number of detected genes if they are located at the ends of contigs. NAuRA is also limited to detecting only genes annotated in Prokka, ignoring any non-annotated genes actually present in the strain. This assembly anomaly can occur in repeat sequence regions such as insertion sequences (Is) or VNTR regions (Hawkey et al., 2015). In the case of poor assembly quality or Is/VNTR-coding gene searches, it would be more reasonable to use a mapping approach such as GeneFinder (Mason et al., 2018).

CONCLUSION

In conclusion, SE gene detection using bioinformatics analyses highlighted SE gene prevalence in European strains of *S. aureus*. These results can be used to improve fast detection methods, such as RT-PCR, microarray analyses or digital PCR, for SE genes. They can be used to determine the new target for the development of molecular detection methods, and improve outbreak investigation in Europe.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the European Nucleotide Archive (ENA) under accession number PRJEB36867 (https://www.ebi.ac.uk/ena/data/view/PRJEB36867).

AUTHOR CONTRIBUTIONS

DM and AF designed the project, developed the pipeline, analyzed the data, and wrote the manuscript. NV participated in strain choice and performed the DNA extraction and preparation for sequencing. SD, YT, LD, BH, CK, and HD provided strains for sequencing. M-YM helped to draft the manuscript and participated in the coordination of the study. J-AH designed the study, helped to draft the manuscript, and coordinated the study. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01483/full#supplementary-material

REFERENCES

Adwan, G., Adwan, K., Jarrar, N., Salamar, Y., and Barakat, A. (2013). Prevalence of *seg*, *seh* and *sei* Genes among Clinical and Nasal *Staphylococcus aureus* Isolates in Palestine. *Br. Microbiol. Res. J.* 3, 139–149. doi: 10.9734/bmrj/2013/2913

Argudín, M. Á, Mendoza, M. C., and Rodicio, M. R. (2010). Food poisoning and *Staphylococcus aureus* enterotoxins. *Toxins* 2, 1751–1773. doi: 10.3390/toxins2071751

Aydin, A., Sudagidan, M., and Muratoglu, K. (2011). Prevalence of staphylococcal enterotoxins, toxin genes and genetic-relatedness of foodborne *Staphylococcus aureus* strains isolated in the Marmara Region of Turkey. *Int. J. Food Microbiol.* 148, 99–106. doi: 10.1016/j.ijfoodmicro.2011.05.007

Balaban, N., and Rasooly, A. (2000). *Staphylococcal enterotoxins. Int. J. Food Microbiol.* 61, 1–10. doi: 10.1016/S0168-1605(00)00377-9

Bania, J., Dabrowska, A., Bystron, J., Korzekwa, K., Chrzanowska, J., and Molenda, J. (2006). Distribution of newly described enterotoxin-like genes in *Staphylococcus aureus* from food. *Int. J. Food Microbiol.* 108, 36–41. doi: 10.1016/j.ijfoodmicro.2005.10.013

Bianchi, D. M., Gallina, S., Bellio, A., Chiesa, F., Civera, T., and Decastelli, L. (2014). Enterotoxin gene profiles of *Staphylococcus aureus* isolated from milk and dairy products in Italy. *Lett. Appl. Microbiol.* 58, 190–196. doi: 10.1111/lam.12182

Bogaerts, B., Winand, R., Fu, Q., Van Braekel, J., Ceyssens, P.-J., Mattheus, W., et al. (2019). Validation of a bioinformatics workflow for routine analysis of whole-genome sequencing data and related challenges for pathogen typing in a european national reference center: *Neisseria meningitidis* as a proof-of-concept. *Front. Microbiol.* 10:362. doi: 10.3389/fmicb.2019.00362

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., et al. (2015). MeDuSa: a multi-draft based scaffolder. *Bioinformatics* 31, 2443–2451. doi: 10.1093/bioinformatics/btv171

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421–421. doi: 10.1186/1471-2105-10-421

Cardona, G., Mir, A., Rossello, F., Rotger, L., and Sanchez, D. (2013). Cophenetic metrics for phylogenetic trees, after Sokal and Rohlf. *BMC Bioinformatics* 14:3. doi: 10.1186/1471-2105-14-3

Carfora, V., Caprioli, A., Marri, N., Sagrafoli, D., Boselli, C., Giacinti, G., et al. (2015). Enterotoxin genes, enterotoxin production, and methicillin resistance in *Staphylococcus aureus* isolated from milk and dairy products in Central Italy. *Int. Dairy J.* 42, 12–15. doi: 10.1016/j.idairyj.2014.10.009

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chao, G., Bao, G., Cao, Y., Yan, W., Wang, Y., Zhang, X., et al. (2015). Prevalence and diversity of enterotoxin genes with genetic background of *Staphylococcus aureus* isolates from different origins in China. *Int. J. Food Microbiol.* 211, 142–147. doi: 10.1016/j.ijfoodmicro.2015.07.018

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2015). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239

Chernomor, O., von Haeseler, A., and Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65, 997–1008. doi: 10.1093/sysbio/syw037

Denayer, S., Delbrassinne, L., Nia, Y., and Botteldoorn, N. (2017). Food-borne outbreak investigation and molecular typing: high diversity of *Staphylococcus aureus* strains and importance of toxin detection. *Toxins* 9:407. doi: 10.3390/toxins9120407

Derzelle, S., Dilasser, F., Duquenne, M., and Deperrois, V. (2009). Differential temporal expression of the staphylococcal enterotoxins genes during cell growth. *Food. Microbiol.* 26, 896–904. doi: 10.1016/j.fm.2009.06.007

Driebe, E. M., Sahl, J. W., Roe, C., Bowers, J. R., Schupp, J. M., Gillece, J. D., et al. (2015). Using whole genome analysis to examine recombination across diverse sequence types of *Staphylococcus aureus*. *PLoS ONE* 10:e0130955. doi: 10.1371/journal.pone.0130955

EFSA, and ECDC (2019). The European Union One Health 2018 zoonoses report. *EFSA J.* 17:e05926.

Enright, M. C., Day, N. P., Davies, C. E., Peacock, S. J., and Spratt, B. G. (2000). Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J. Clin. Microbiol.* 38, 1008–1015. doi: 10.1128/jcm.38.3.1008-1015.2000

Felsenstein, J. (2013). *PHYLIP (Phylogeny Inference Package) version 3.695.* Distributed by the author. Department of Genomes Sciences, University of Washington, Seatle, WA.

Fisher, E. L., Otto, M., and Cheung, G. Y. C. (2018). Basis of virulence in enterotoxin-mediated staphylococcal food poisoning. *Front. Microbiol.* 9:436–436. doi: 10.3389/fmicb.2018.00436

Grumann, D., Nübel, U., and Bröker, B. M. (2014). *Staphylococcus aureus* toxins – Their functions and genetics. *Infect. Genet. Evol.* 21, 583–592. doi: 10.1016/j.meegid.2013.03.013

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hawkey, J., Hamidian, M., Wick, R. R., Edwards, D. J., Billman-Jacobe, H., Hall, R. M., et al. (2015). ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC Genomics* 16:667. doi: 10.1186/s12864-015-1860-2

Hennekinne, J.-A., De Buyser, M.-L., and Dragacci, S. (2012). *Staphylococcus aureus* and its food poisoning toxins: characterization and outbreak investigation. *FEMS Microbiol. Rev.* 36, 815–836. doi: 10.1111/j.1574-6976.2011.00311.x

Hunt, M., Mather, A. E., Sánchez-Busó, L., Page, A. J., Parkhill, J., Keane, J. A., et al. (2017). ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genomics* 3:e000131. doi: 10.1099/mgen.0.000131

Jarraud, S., Peyrat, M. A., Lim, A., Tristan, A., Bes, M., Mougel, C., et al. (2001). egc, a highly prevalent operon of enterotoxin gene, forms a putative nursery of superantigens in *Staphylococcus aureus*. *J. Immunol.* 166:669. doi: 10.4049/jimmunol.166.1.669

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kérouanton, A., Hennekinne, J. A., Letertre, C., Petit, L., Chesneau, O., Brisabois, A., et al. (2007). Characterization of *Staphylococcus aureus* strains associated with food poisoning outbreaks in France. *Int. J. Food Microbiol.* 115, 369–375. doi: 10.1016/j.ijfoodmicro.2006.10.050

Kosugi, S., Hirakawa, H., and Tabata, S. (2015). GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. *Bioinformatics* 31, 3733–3741. doi: 10.1093/bioinformatics/btv465

Le Loir, Y., Baron, F., and Gautier, M. (2003). *Staphylococcus aureus* and food poisoning. *Genet. Mol. Res.* 2, 63–76.

Mashouf, R. Y., Hosseini, S. M., Mousavi, S. M., and Arabestani, M. R. (2015). Prevalence of enterotoxin genes and antibacterial susceptibility pattern of *Staphylococcus aureus* strains isolated from animal originated foods in west of iran. *Oman Med. J.* 30, 283–290. doi: 10.5001/omj.2015.56

Mason, A., Foster, D., Bradley, P., Golubchik, T., Doumith, M., Gordon, N. C., et al. (2018). Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from *Staphylococcus aureus* whole-genome sequences. *J. Clin. Microbiol.* 56, e1815–e1817. doi: 10.1128/JCM.01815-17

Monistero, V., Graber, H. U., Pollera, C., Cremonesi, P., Castiglioni, B., Bottini, E., et al. (2018). *Staphylococcus aureus* isolates from bovine mastitis in eight countries: genotypes, detection of genes encoding different toxins and other virulence genes. *Toxins* 10, 247. doi: 10.3390/toxins10060247

Naffa, R. G., Bdour, S. M., Migdadi, H. M., and Shehabi, A. A. (2006). Enterotoxicity and genetic variation among clinical *Staphylococcus aureus* isolates in Jordan. *J. Med. Microbiol.* 55, 183–187. doi: 10.1099/jmm.0.46183-0

Oakeson, K. F., Wagner, J. M., Mendenhall, M., Rohrwasser, A., and Atkinson-Dunn, R. (2017). Bioinformatic analyses of whole-genome sequence data in a public health laboratory. *Emerg. Infect. Dis.* 23, 1441–1445. doi: 10.3201/eid2309.170416

Omoe, K., Hu, D.-L., Ono, H. K., Shimizu, S., Takahashi-Omoe, H., Nakane, A., et al. (2013). Emetic potentials of newly identified staphylococcal enterotoxin-like toxins. *Infect. Immun.* 81, 3627–3631. doi: 10.1128/IAI.00550-13

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17, 132. doi: 10.1186/s13059-016-0997-x

Ono, H. K., Hirose, S., Naito, I., Sato'o, Y., Asano, K., Hu, D.-L., et al. (2017). The emetic activity of *Staphylococcal enterotoxins*, SEK, SEL, SEM, SEN and SEO in a small emetic animal model, the house musk shrew. *Microbiol. Immunol.* 61, 12–16. doi: 10.1111/1348-0421.12460

Ono, H. K., Omoe, K., Imanishi, K., Iwakabe, Y., Hu, D.-L., Kato, H., et al. (2008). Identification and characterization of two novel *Staphylococcal enterotoxins*, Types S and T. *Infect. Immun.* 76:4999. doi: 10.1128/IAI.00045-08

Ortega, E., Abriouel, H., Lucas, R., and Gálvez, A. (2010). Multiple roles of *Staphylococcus aureus* enterotoxins: pathogenicity, superantigenic activity, and correlation to antibiotic resistance. *Toxins* 2, 2117–2131. doi: 10.3390/toxins2082117

Ostyn, A., De Buyser, M. L., Guillier, F., Groult, J., Félix, B., Salah, S., et al. (2010). First evidence of a food poisoning outbreak due to staphylococcal enterotoxin type E, France, 2009. *Eurosurveillance* 15:19528. doi: 10.2807/ese.15.13.19528-en

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Petit, R. A. III, and Read, T. D. (2018). *Staphylococcus aureus* viewed from the perspective of 40,000+ genomes. *PeerJ* 6:e5261. doi: 10.7717/peerj.5261

Pu, S., Wang, F., and Ge, B. (2010). Characterization of Toxin genes and antimicrobial susceptibility of *Staphylococcus aureus* isolates from louisiana retail meats. *Foodborne Pathog. Dis.* 8, 299–306. doi: 10.1089/fpd.2010.0679

Roussel, S., Felix, B., Vingadassalon, N., Grout, J., Hennekinne, J.-A., Guillier, L., et al. (2015). *Staphylococcus aureus* strains associated with food poisoning outbreaks in France: comparison of different molecular typing methods, including MLVA. *Front. Microbiol.* 6:882. doi: 10.3389/fmicb.2015.00882

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539–539. doi: 10.1038/msb.2011.75

Song, M., Shi, C., Xu, X., and Shi, X. (2016). Molecular typing and virulence gene profiles of enterotoxin gene cluster (*egc*)-positive *Staphylococcus aureus* isolates obtained from various food and clinical specimens. *Foodborne Pathog. Dis.* 13, 592–601. doi: 10.1089/fpd.2016.2162

Strauß, L., Stegger, M., Akpaka, P. E., Alabi, A., Breurec, S., Coombs, G., et al. (2017). Origin, evolution, and global transmission of community-acquired *Staphylococcus aureus* ST8. *Proc. Natl. Acad. Sci. U.S.A.* 114, E10596–E10604. doi: 10.1073/pnas.1702472114

Sukumaran, J., and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571. doi: 10.1093/bioinformatics/btq228

The UniProt Consortium (2018). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Titouche, Y., Hakem, A., Houali, K., Meheut, T., Vingadassalon, N., Ruiz-Ripa, L., et al. (2019). Emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) ST8 in raw milk and traditional dairy products in the Tizi Ouzou area of Algeria. *J. Dairy Sci.* 102, 6876–6884. doi: 10.3168/jds.2018-16208

Wilson, G. J., Seo, K. S., Cartwright, R. A., Connelley, T., Chuang-Smith, O. N., Merriman, J. A., et al. (2011). A Novel Core Genome-encoded superantigen contributes to lethality of community-associated MRSA necrotizing pneumonia. *PLoS Pathog.* 7:e1002271. doi: 10.1371/journal.ppat.1002271

Zhang, D.-F., Yang, X.-Y., Zhang, J., Qin, X., Huang, X., Cui, Y., et al. (2018). Identification and characterization of two novel superantigens among *Staphylococcus aureus* complex. *Int. J. Med. Microbiol.* 308, 438–446. doi: 10.1016/j.ijmm.2018.03.002