Supporting Information

Strauß et al. 10.1073/pnas.1702472114

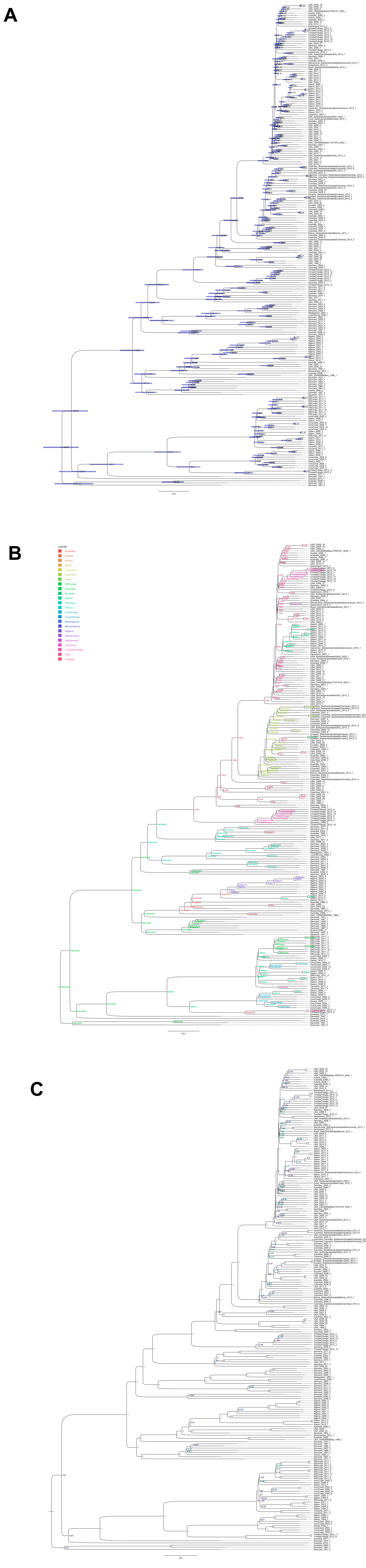

Fig. S1. Maximum clade credibility (mcc) tree representation of the BEAST analysis results of 224 *Staphylococcus aureus* ST8 isolates. The timescale indicates years in the past. The youngest isolate was collected in 2013; the most recent common ancestor (MRCA) of all isolates dates back to 1854 (HPD 95%, 1841 to 1865). (*A*) Node labels indicate divergence times of common ancestors of phylogenetic groups; the bars indicate the respective 95% highest posterior density intervals. (*B*) Node labels and branch colors indicate the isolates' origins and allow the reconstruction of possible spatial transmission routes. (C) Node labels indicate posterior support values; colors represent high (blue) or low (green) support (gradient).

Fig. S2. Maximum Likelihood phylogeny of single nucleotide polymorphisms of the core genomes of 224 *Staphylococcus aureus* ST8 isolates, rooted with ST1 NCBI Reference Genome MW2 (accession no. BA000033.1). Bootstrap values are indicated on nodes and highlighted in color gradient from violet (high values) to green (low values). (*A*) Inclusion of MW2 in presentation. (*B*) Exclusion of MW2 in presentation.

Dataset S1. Epidata, *spa*-type ENA accession numbers, and genotyping results of USA300 characteristic genes of all 224 ST8 isolates

Dataset S1

The numbers represent the unique allele name, which can also be found in the corresponding fasta file (Dataset S2). 'Failed' results indicate pseudogenes with unknown functionality.

Dataset S2. Fasta sequences used as reference templates in the BLAST search for the identification of USA300-specific genes in *Staphylococcus aureus* whole genome sequences

Dataset S2

Each tab contains the reference sequences for one gene. Association with genetic elements is color-coded: green, Staphylococcal cassette chromosome *mec* (SCC*mec*); yellow, Panton–Valentine leukocidin (PVL); blue, capsule polysaccharide; violet, arginine catabolic mobile element (ACME); red, copper and mercury resistance element (COMER). New alleles were identified within thresholds of ≥95% nucleotide identity and ≥99% alignment coverage.

Dataset S3. Nomenclature of staphylococcal cassette chromosome *mec* subtyping

Dataset S3

Dataset S4. FASTA file containing the concatenated core genomic single nucleotide polymorphisms (SNPs) of 224 *Staphylococcus aureus* ST8 isolates analyzed in this study

Dataset S4

This file was used as the basis for phylogenetic reconstructions. To use this file, please remove the .txt suffix.

Dataset S5. TXT file containing the .xml file generated by BEAUti 1.8.2 and used as input for BEAST

Dataset S5

This file contains all settings and information used for the final Bayesian analysis of our dataset. To use this file with BEAST, please exchange the .txt suffix with .xml.

Dataset S6. To use this file, please exchange the .txt suffix with .kml and open with Google Earth

Dataset S6

It visualizes the global spread of all 224 *Staphylococcus aureus* isolates analyzed in this study from 1854 to 2013 based on BEAST data. If Google Earth is not installed locally, please download it from https://www.google.com/earth/download/gep/agree.html.

Dataset S7. List of targets used for detection of USA300-specific yet unknown genetic markers

Dataset S7

USA300-NAE specific targets were created based on the genome of NCBI RefSeq USA300 genome FPR3757 (NC_007793.1), USA300-SAE specific targets were created based on the genome of strain M121 (CP007670.1), the African USA300 clade specific targets were created based on the contig sequences of strain Gabon_2009_1 (SAMEA4081496) (annotation was done using the RAST system under rast.nmpdr.org/ using default parameters), and the non-USA300 outgroup targets were created based on isolate C1768 (LAMT01000000).