HAL
open science

Niche specialization and spread of Staphylococcus capitis involved in neonatal sepsis

Thierry Wirth, Marine Bergot, Jean-Philippe Rasigade, Bruno Pichon, Maxime Barbier, Patricia Martins-Simoes, Laurent Jacob, Rachel Pike, Pierre Tissieres, Jean-Charles Picaud, et al.

To cite this version:

Thierry Wirth, Marine Bergot, Jean-Philippe Rasigade, Bruno Pichon, Maxime Barbier, et al.. Niche specialization and spread of Staphylococcus capitis involved in neonatal sepsis. Nature Microbiology, 2020, 5 (5), pp.735-745. 10.1038/s41564-020-0676-2 . hal-03056509

HAL Id: hal-03056509

https://hal.science/hal-03056509v1

Submitted on 11 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-020-0676-2

Niche specialization and spread of *Staphylococcus capitis* involved in neonatal sepsis

Thierry Wirth[1,2]✉, Marine Bergot[3], Jean-Philippe Rasigade[1,3,4], Bruno Pichon[5], Maxime Barbier[1], Patricia Martins-Simoes[3,4], Laurent Jacob[6], Rachel Pike[5], Pierre Tissieres[7,8], Jean-Charles Picaud[9], Angela Kearns[5], Philip Supply[10], Marine Butin[4,11,40]✉, Frédéric Laurent[3,4,40], the International Consortium for Staphylococcus capitis neonatal sepsis* and the ESGS Study Group of ESCMID*

The multidrug-resistant *Staphylococcus capitis* NRCS-A clone is responsible for sepsis in preterm infants in neonatal intensive care units (NICUs) worldwide. Here, to retrace the spread of this clone and to identify drivers of its specific success, we investigated a representative collection of 250 *S. capitis* isolates from adults and newborns. Bayesian analyses confirmed the spread of the NRCS-A clone and enabled us to date its emergence in the late 1960s and its expansion during the 1980s, coinciding with the establishment of NICUs and the increasing use of vancomycin in these units, respectively. This dynamic was accompanied by the acquisition of mutations in antimicrobial resistance- and bacteriocin-encoding genes. Furthermore, combined statistical tools and a genome-wide association study convergently point to vancomycin resistance as a major driver of NRCS-A success. We also identified another *S. capitis* subclade (alpha clade) that emerged independently, showing parallel evolution towards NICU specialization and non-susceptibility to vancomycin, indicating convergent evolution in NICU-associated pathogens. These findings illustrate how the broad use of antibiotics can repeatedly lead initially commensal drug-susceptible bacteria to evolve into multidrug-resistant clones that are able to successfully spread worldwide and become pathogenic for highly vulnerable patients.

Preterm birth affects 11% of pregnancies globally, and is the leading cause of death in children younger than 5 yr of age[1,2]. Despite technical and medical advances in NICUs in recent decades, complications related to preterm birth remain a major cause of neonatal death[1]. Among these complications, nosocomial late-onset sepsis (LOS, occurring 3 d after birth), mainly involving skin commensal coagulase-negative staphylococci (CoNS)[3], represents a frequent cause of morbidity and mortality in newborns[3–5]. Rates of such nosocomial infections are grouping in relation to the improved survival of increasing numbers of small and vulnerable newborns along with the need for use of invasive devices for these patients.

A single multidrug-resistant *Staphylococcus capitis* clone, named NRCS-A, has recently emerged as a major pathogen among newborns in NICUs. This clone has been isolated in 17 countries throughout the world, almost exclusively in NICUs, and with a high prevalence in some settings[6–9], indicating a global dissemination that is highly specific to the NICU environment. The extensive diffusion of the NRCS-A clone is a major concern given its multidrug resistance profile, including resistance to methicillin, aminoglycosides, fosfomycin and reduced susceptibility to vancomycin, which raises the issue of potential therapeutic failure in infected newborns. Moreover, the implication of *S. capitis* species in late-onset sepsis has been reported as an independent risk factor for severe morbidity in preterm infants[10]. This unique epidemiology represents the first example of worldwide diffusion of a CoNS clone, a phenomenon usually described for *Staphylococcus aureus* clones and recently reported for three globally disseminated multidrug-resistant *Staphylococcus epidermidis* clones[11].

Despite available epidemiological and phenotypical description of the NRCS-A clone, the spatiotemporal dynamics of its spread and the genetic basis that favoured its high specificity for newborns remain poorly understood. In this study, we performed whole-genome sequencing analysis of a unique set of 250 *S. capitis* clinical isolates, collected between 1994 and 2015—including the oldest and most recent isolates from newborns and adults available—from 22 countries in Europe, the Americas, southern Asia and Oceania. We combined phylogenomics and molecular clock analyses to reconstruct the evolutionary history and spatiotemporal emergence of the NRCS-A clone. We also explored genetic and phenotypic determinants that may underlie its epidemiological success and specificity for newborns.

Results

Overall population structure of *S. capitis*. To reconstruct the overall population structure and mode of evolution of *S. capitis*, we

[1]Institut Systématique Evolution Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, Université des Antilles, EPHE, Paris, France. [2]PSL University, EPHE, Paris, France. [3]Institut des Agents Infectieux, Département de Bactériologie, Centre National de Référence des Staphylocoques, Hospices Civils de Lyon, Lyon, France. [4]Centre International de recherche en Infectiologie, INSERM U1111 - CNRS UMR5308 - ENS Lyon - Université Lyon 1, Lyon, France. [5]Staphylococcus Reference Section, National Infection Service, Public Health England, London, UK. [6]Laboratoire Biométrie et Biologie Evolutive, CNRS UMR5558, Université Lyon 1, Lyon, France. [7]Institut de Biologie de la cellule (I2BC-UMR9198), CNRS, CEA, Univ. Paris Sud, Université Paris Saclay, Gif-sur-Yvette, France. [8]Service de Réanimation Néonatale, Hôpitaux Universitaires Paris Sud APHP, Le Kremlin-Bicêtre, Paris, France. [9]Service de Réanimation Néonatale, Hôpital Croix Rousse, Hospices Civils de Lyon, Lyon, France. [10]Université de Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Centre d'Infection et d'Immunité de Lille, Lille, France. [11]Service de Réanimation Néonatale, Hôpital Femme Mère Enfant, Hospices Civils de Lyon, Lyon, France. [40]These authors contributed equally: Marine Butin, Frédéric Laurent. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: wirth@mnhn.fr; marine.butin@chu-lyon.fr

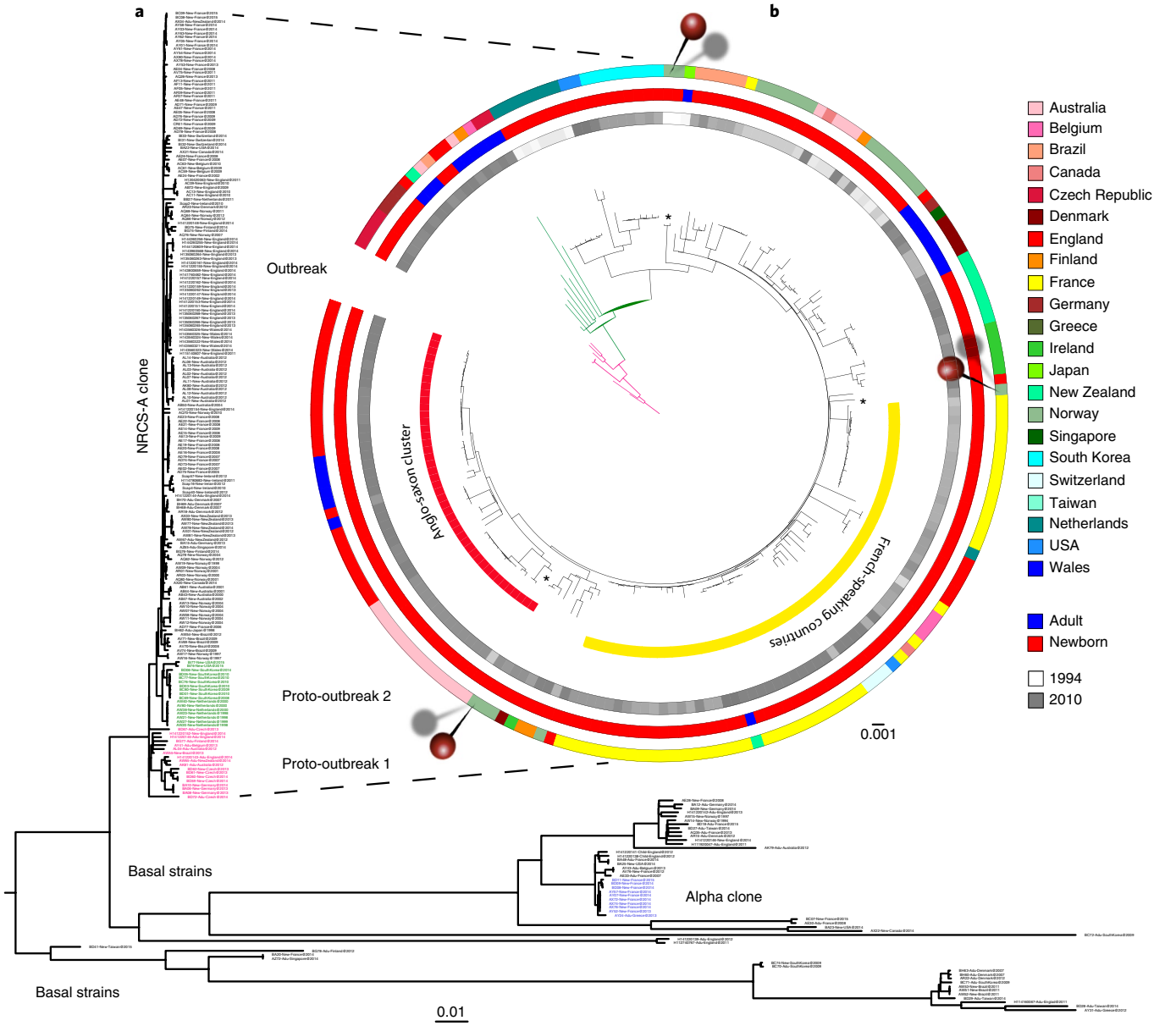


Fig. 1 | Phylogenetic inference of the global *S. capitis* strain collection and expansion of the multidrug-resistant NRCS-A clone. a, Maximum-likelihood tree based on the SNPs identified from 250 clinical isolates originating from 22 countries, using CR01 as the reference genome. Strain codes include patient type (New, newborns; Adu, adults) and country of isolation. NRCS-A isolates that are part of the proto-outbreak 1 and 2 lineages are shown in pink and green, respectively. Isolates of the alpha clone in the basal strain group are shown in blue. **b**, Circular maximum-likelihood tree based on the 197 NRCS-A isolates. Overlaid are the geographical source of the strains (outer ring), the type of human host (middle ring) and the year of isolation (inner ring). Colour codes used in the three rings are shown on the right. The three needles and stars pinpoint the putative role of Norway as the source of the successive propagation waves of the outbreak clone. The scale bars indicate the number of nucleotide substitutions per site in the maximum-likelihood tree. Proto-outbreak 1 and 2 lineages are shown in pink and green, respectively.

first aligned the genome sequences of the 250 isolates against the NRCS-A reference genome CR01, resulting in a total of 22,621 single nucleotide polymorphisms (SNPs). To quantify recombination, we used ClonalFrameML[12], which is specifically aimed at analysing whole-genome sequence data (see Supplementary Information). The results indicated that the impact of recombination (r) on the genome-wide substitution rate in *S. capitis* overall is almost equal to the impact of mutation (m), with $r/m = 0.85$. ClonalFrameML identified 190 recombination events in the global genealogy (Extended Data Fig. 1). The largest detected events (up to 26 kb) are probably products of horizontal gene transfer, some of which correspond to the insertion of pathogenicity islands.

Clonal specialization and geographical dispersion of NRCS-A. The reconstructed maximum-likelihood tree (Fig. 1a) enabled us to draw a clear distinction between NRCS-A isolates that harbour the previously described specific NRCS-A pulsed-field gel electrophoresis pattern[8] ($n = 197$) and all the other strains found in basal positions ($n = 53$; hereafter 'basal'). These reconstructions revealed that this NRCS-A population is composed of at least three sublineages, which we named in chronological order of divergence on the basis of the observed branching order in the tree: 'proto-outbreak 1' ($n = 18$), 'proto-outbreak 2' ($n = 17$) and 'outbreak' ($n = 162$) (Fig. 1a,b). These three clades are supported both by bootstrap values greater than 95% and by the trimodal distribution of the

pairwise SNP distances between all NRCS-A strains (Extended Data Fig. 2c).

The proportions of neonatal isolates differed markedly among the (sub)lineages, as visualized in a circular maximum-likelihood tree (Fig. 1b), a minimum-spanning tree (Extended Data Fig. 2a) and a principal component analysis (Extended Data Fig. 2b). While the neonatal isolates only comprised 43% (23/53) of the basal strains and 50% (9/18) of the proto-outbreak 1 isolates, they represented 100% (17/17) and 94% (153/162) of the proto-outbreak 2 and outbreak strains, respectively (Extended Data Fig. 2d), indicating a specialization of the most recent or outbreak branches. In addition, geographical structuring was also apparent in the outbreak part of the reconstructed NRCS-A phylogeny, comprising two large subgroups enriched in isolates originating from French-speaking countries (Canada, Belgium, France and Switzerland) and English-speaking countries (Australia, England, United States of America and Wales), respectively (Fig. 1b). Notably, Norwegian strains branched at the deepest nodes of each of these geographical subgroups. Furthermore, the two most basal strains in the entire outbreak part of the tree are also from Norway, suggesting a putative sourcing effect.

Moreover, a particular subgroup in the basal strain clade, which we called alpha, appears to follow a similar emergence pattern as the NRCS-A outbreak clone (Fig. 1a). The alpha clade shows extremely restricted genetic diversity and apparent enrichment in neonatal isolates, with nine recent isolates from a single NICU setting in France (in 2013 and 2014) (Supplementary Table 1).

Temporal emergence and dispersion of NRCS-A. To date the emergence and reconstruct the temporal evolution of NRCS-A, we used a Bayesian approach. As a critical prerequisite, we checked whether the NRCS-A isolates represent a measurably evolving population. Critically, none of the runs with randomized molecular data had substitution rate estimates that overlap with the rates inferred with the real molecular data, supporting a conclusion of measurable genome evolution and robustness of our temporal inferences (Fig. 2a). The best-fitting evolutionary model for the NRCS-A population was obtained under a Bayesian skyline model with a relaxed clock, leading to a rate of 1.51×10^{-6} substitutions per position per year (Fig. 2c), or 1.7 mutations per genome per year. This mutation rate is close to the rate inferred from *S. aureus* genomic datasets, also in the range of 1×10^{-6} substitution per nucleotide per year[13,14]. According to the coalescence-based reconstructions of demographic variation, the NRCS-A clone population experienced a tenfold expansion in the late 1980s, followed by minor fluctuations and a sharp final decline that initiated in 2013 (Fig. 2d). The time of emergence of the most recent common ancestor (TMRCA) of the NRCS-A clone (Fig. 2b) was estimated to be 1969 (95% highest posterior density, 1952–1986) and that of the outbreak sublineage was estimated to be 1982 (1974–1993).

Genomic specificities of NRCS-A and alpha clades. To identify genomic specificities potentially underlying the global success of the NRCS-A clone, we first focused on 55 genes previously identified as specifically present in NRCS-A *S. capitis* strains[15]. Among these 55 genes, 28 were detected both in outbreak strains and in proto-outbreak 1 and proto-outbreak 2 strains, notably including the gene *nsr* encoding for nisin resistance and the additional cell-wall teichoic-acids-associated cluster (*tar*FIJL) (Supplementary Table 1). The other 27 genes were found only in the outbreak strains and were absent from the proto-outbreak strains. These 27 genes are carried by a specific composite staphylococcal chromosomic cassette (SCC), the SCC*mec*-SCC*cad/ars/cop* cassette[16], associating a type V SCC*mec* cassette and a SCC*cad/ars/cop* cassette carrying genes conferring resistance to heavy metals, with a total of nine genes associated with the type III-A clustered regularly interspaced short palindromic repeats (CRISPR) element (Fig. 3a).

Because of SCC*mec* cassette- and CRISPR-related genomic specificity of the outbreak group, we further searched for and characterized SCC cassettes and CRISPR elements in all 250 genomes (Fig. 3a). This revealed that this composite SCC locus with the type III-A CRISPR element is a unique genomic feature of the outbreak group. By contrast, the proto-outbreak 1 and proto-outbreak 2 strains harbour two unrelated types of SCC*mec* cassettes, type IV and type II, respectively, and no CRISPR element. Notably, alpha clade isolates all harbour a type IV SCC*mec* cassette, identical to the one found in proto-outbreak 1 isolates (Fig. 3a).

Positive selection signatures of NRCS-A and alpha clades. In order to detect potential signatures of positive selection of the outbreak strains, we performed a genome scan analysis using a Bayesian model[17,18] that detects the structure and clustering of individuals in a population. On the basis of the Bayesian principal component analysis, principal component 2 (PC2) segregated the outbreak strains from the rest of the NRCS-A strain collection (including the proto-outbreak 1 and proto-outbreak 2 isolates) (Extended Data Fig. 3). A total of 32 SNPs were associated with PC2 and displayed the highest Mahalanobis distances (reflecting correlations between SNPs most related to the genetic structure), including 18 non-synonymous SNPs potentially involved in diversifying selection. These genes are shown in Extended Data Fig. 4 and notably include *rpsJ*, involved in tigecyclin resistance[19], and *gln*Q, a glutamine ABC transporter that may be involved in vancomycin resistance[20], as well as a gene coding for the lipase LipA, which has recently been shown to be a suppressor of macrophage activation, rendering these cells inefficient at controlling infection by *S. aureus*[21,22] (Extended Data Fig. 5).

Likewise, 17 SNPs were identified that distinguish isolates of alpha clade from the other basal isolates (Extended Data Fig. 4), notably including one that results in a nonsense mutation (creating a stop codon) in the *murR* gene, which was previously described as related to peptidoglycan synthesis and vancomycin resistance[23], and a second that results in a non-synonymous mutation in the *pgc*A gene, which is involved in cell-wall properties[23] (Extended Data Fig. 5).

Phenotypic specificities including antimicrobial resistance. Most of the outbreak isolates harboured multidrug-resistant phenotypes, including resistance to oxacillin (96% of the isolates), aminoglycosides (99%) and resistance or heteroresistance to vancomycin (99%) (Extended Data Figs. 6 and 3b). This profile is broadly consistent with the preferential use of these antimicrobials in NICUs[24] (see Supplementary Information for further details). Multidrug resistance levels were also much higher in alpha clade isolates than in the basal group isolates (Extended Data Fig. 6). All alpha clade isolates exhibited resistance to vancomycin, oxacillin and aminoglycosides, consistent with the dominant use of these compounds in NICUs.

However, no statistically significant differences were observed between the outbreak strains and other groups for the additional tested phenotypes, including phagocytosis, cytotoxicity assay, tolerance to desiccation, kinetics of bacterial growth in standard conditions and under oxidative stress, and biofilm production (Extended Data Fig. 7 and Supplementary Information).

Genetic determinants of antimicrobial resistance. Genotypes of resistance were largely consistent with the observed resistance phenotypes, as well-established resistance determinants were identified in many strains, such as *mec*A in methicillin-resistant isolates, *aac*(6′)-*aph*(2″) gene in aminoglycoside-resistant isolates, and the *grlA-gyrA* mutation in ciprofloxacin-resistant proto-outbreak 2 strains (Extended Data Fig. 6 and Supplementary Table 1). To

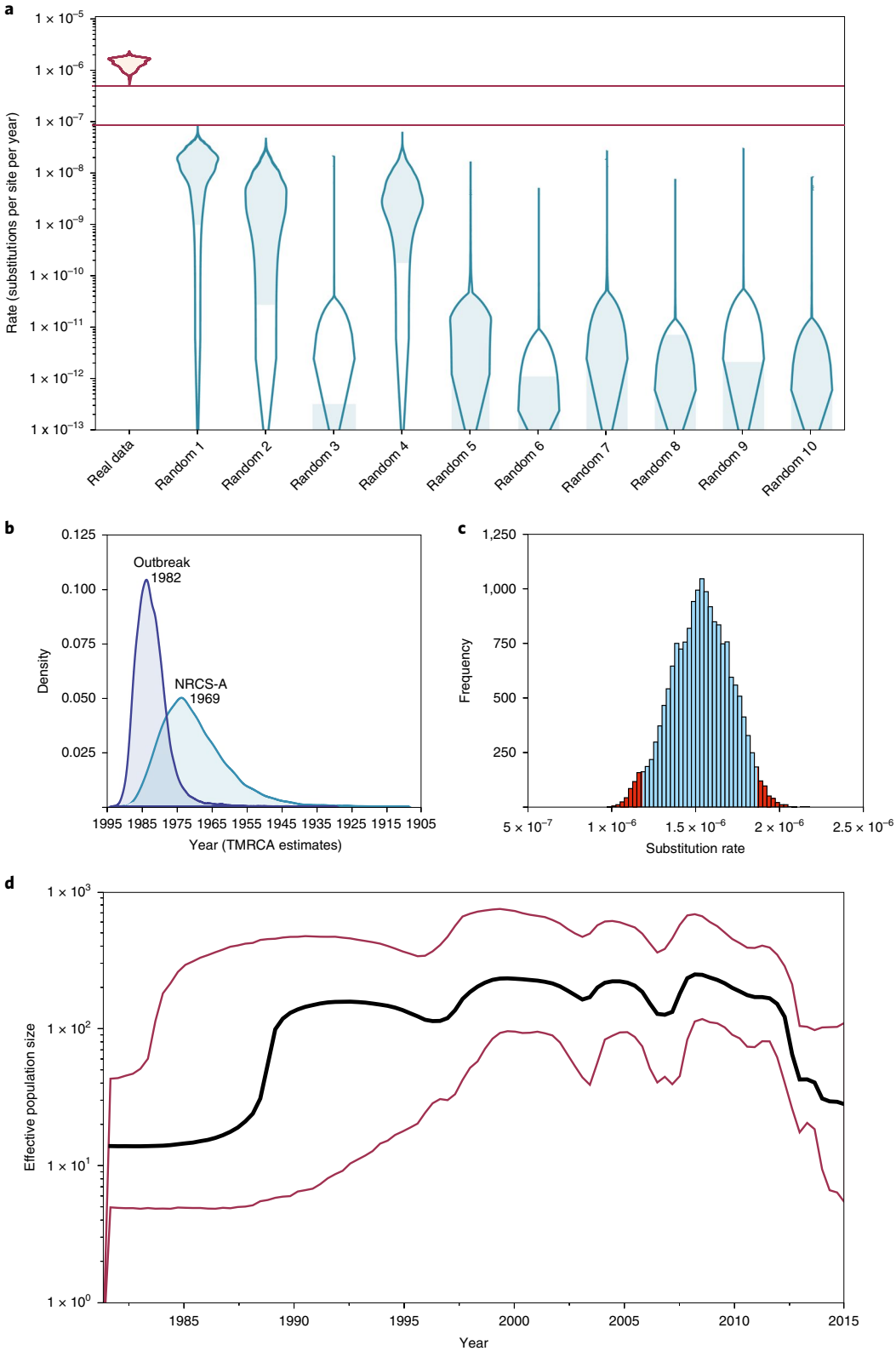


Fig. 2 | Demographic and temporal evolution parameters of the NRCS-A strain population obtained from Bayesian inferences. a, Tip-randomization test on molecular clock rates under a Bayesian skyline model and a strict clock. Calculated clock rates are represented using non-parametric violin plots that contain all data points. The violin plots include all the data and the box marker indicates the interquartile range. Note that there is no overlap between rates calculated with randomized datasets and the real dataset. **b**, Posterior probability densities of the time of emergence of the most recent common ancestor of the NRCS-A (blue, $n=197$) and NRCS-A outbreak (violet, $n=162$) clones. Inferred estimates correspond to the respective modal values. **c**, Posterior probability density of the nucleotide substitution rate estimate, with 95% confidence intervals represented in blue. **d**, Bayesian skyline plot indicating variations over time of the effective population size of the NRCS-A strain population ($n=197$). The estimated variations and the 95% confidence intervals are represented by black and red lines, respectively.

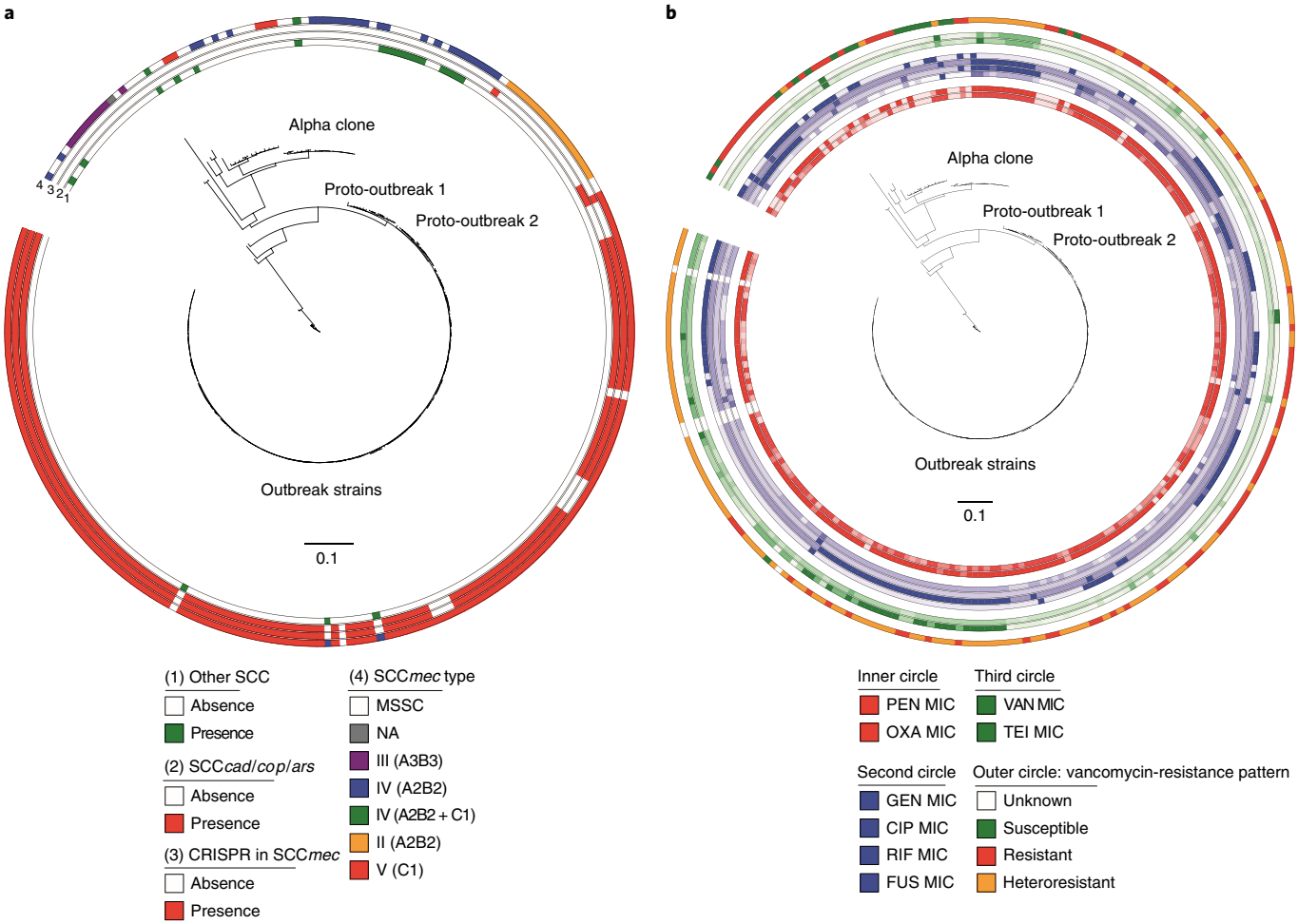


Fig. 3 | Distribution of SCC*mec* cassettes and antibiotic phenotypic resistance profiles of the *S. capitis* isolates. a, Distribution and architecture of the SCC*mec* cassettes among the main clades of the global *S. capitis* strain collection. Overlaid are the main detected SCC*mec* types (first ring from the exterior; type of recombinase in SCC*mec* cassette is indicated in brackets), presence or absence of CRISPR element in SCC*mec* (second ring inwards), of SCC*cad/cop/ars* (third ring), and of other SCC cassettes (fourth ring). Colour codes used in the rings are shown on the bottom left. MSSC, methicillin-susceptible *S. capitis*. **b**, Distribution of antibiotic phenotypic resistance profiles of isolates and their associated MIC. Overlaid are the vancomycin-resistance patterns (first ring from the exterior), teicoplanin and vancomycin MICs (second and third rings from the exterior, respectively), fusidic acid, rifampicin, ciprofloxacin and gentamicin (fourth to seventh rings, respectively), and oxacillin and penicillin (eighth and ninth rings, respectively). Colour intensities are proportional to the MIC values; darker colours indicate higher MICs. The scale bars represent the numbers of nucleotide substitutions per site in the maximum-likelihood trees. PEN, penicillin; OXA, oxacillin; GEN, gentamicin; CIP, ciprofloxacin; RIF, rifampicin; FUS, fusidic acid; VAN, vancomycin; TEI, teicoplanin.

search for additional potential mutation candidates, we performed a genome-wide association study (GWAS) search on the outbreak clade isolates, selected as the largest strain group to maximize statistical power, using vancomycin MICs as quantitative traits. Consistently, vancomycin minimum inhibitory concentration (MIC) was found to be significantly associated with several independent genomic variants (Fig. 4 and Extended Data Fig. 8), involving, for instance, the global regulator gene *sarA* (harmonic mean of *P* value (HMP) = 1.7×10^{-7}), the DNA mismatch repair system-associated gene *mutL* (HMP = 2.5×10^{-7}) and genes related to bacterial survival and growth such as *rsm*E (HMP = 7.6×10^{-9}) and *cit*B (HMP = 3.4×10^{-7}).

Specific resistance patterns correlate with *S. capitis* epidemic success in newborns. Acquisition of increased drug resistance is an expected key adaptive advantage for the spread of nosocomial pathogens, especially for those such as *S. capitis*, that are exposed to the high drug selection pressure that prevails in NICUs. Thus, the specific resistance pattern of outbreak and alpha strains (Fig. 3b) led us to hypothesize that resistance to drugs widely used in NICUs, especially vancomycin, might drive their epidemic success. To test this, we searched for genomic signatures of epidemic success in our dataset and analysed their correlation with resistance patterns. Epidemic success was inferred using the recently described time-scaled haplotypic density (THD) approach[25], which assigns a relative index of success over time to each isolate in a population on the basis of the distribution of genetic distances.

The highest THD success indices were found in isolates of the outbreak clade, followed by proto-outbreak 2, proto-outbreak 1, alpha and basal groups (Fig. 5b and Supplementary Table 1). In bivariate analysis, the success index correlated with higher MICs of β-lactams, gentamicin and glycopeptides, and lower MICs of ciprofloxacin (Fig. 5a, bivariate model), reflecting the specific resistance profile of the successful outbreak clade (Fig. 3b). After controlling for population structure by introducing the isolate's clade as an additional covariate (Fig. 5a, controlled model), the associations

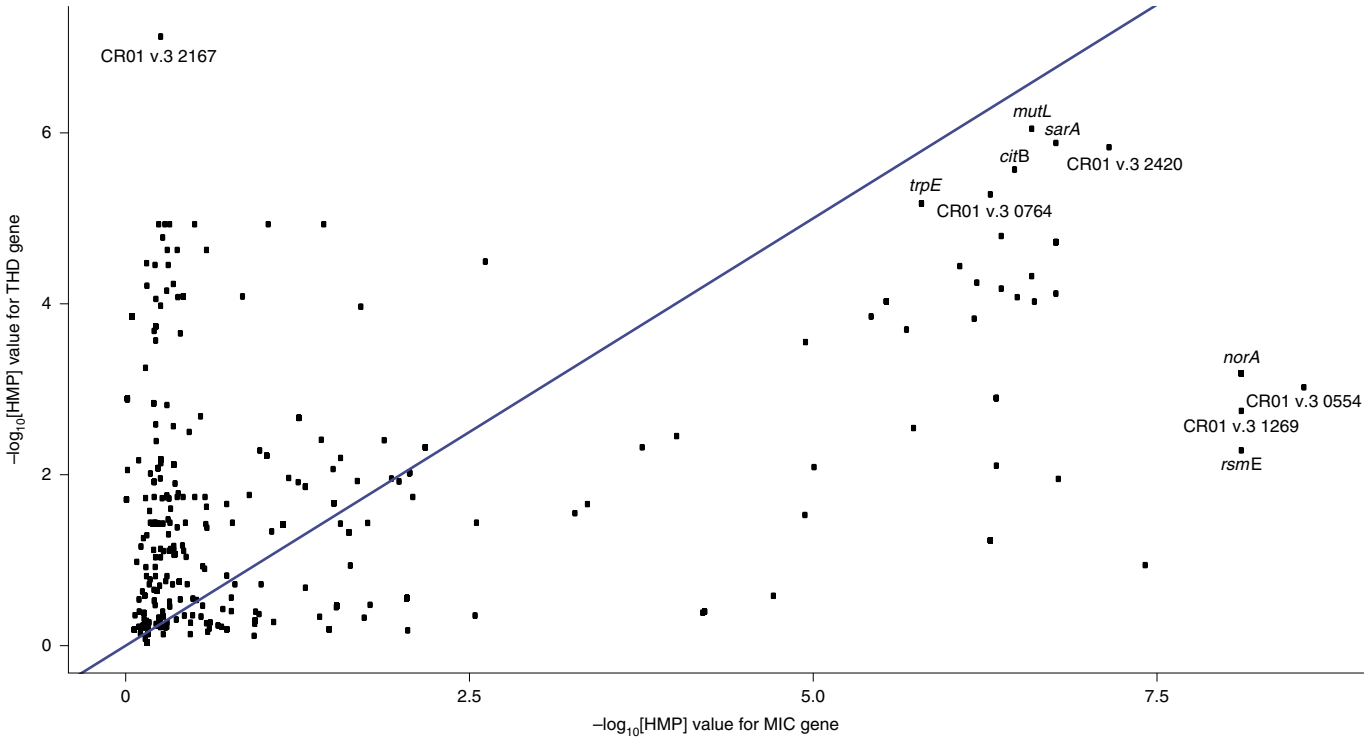


Fig. 4 | GWAS scatter plot. A GWAS was performed on outbreak isolates ($n = 162$). All *k*-mers connected to the same gene were agglomerated via their HMP using the R package HarmonicmeanP (see Methods). Shown here is the scatter plot of common genes ($n = 261$) detected using DBGWAS and associated with vancomycin MIC and/or THD success index. The *x* axis represents the $-\log_{10}$[HMP] value for vancomycin MIC and the *y* axis represents the $-\log_{10}$[HMP] THD index for epidemic success. Genes with a $-\log_{10}$[HMP] of < 7.5 on either axis and/or > 5 on both axes, were deemed significant and named on the graph. These genes are also listed in Extended Data Fig. 8.

between the success index and MICs were nearly abolished for β-lactams, gentamicin and ciprofloxacin, indicating that resistance to these drugs correlated with success variations between clades but not within them. Likewise, the strong association of success index with high rifampicin MICs (Fig. 5a) was probably driven by a very dense and specific cluster of rifampicin-resistant isolates within the outbreak clade only (Fig. 3b).

Interestingly, the association between success and glycopeptide MICs remained highly significant after controlling for population structure (Fig. 5a), indicating that higher glycopeptide MICs predicted variations of strain success both between and within clades (Fig. 5a). Indeed, in a subgroup analysis of each clade and group, higher vancomycin MICs correlated with success among isolates of the NICU-adapted outbreak and alpha clades, whereas these MICs negatively correlated with success within proto-outbreak 1 and basal strain clades (Fig. 5b). This suggests that vancomycin resistance is only beneficial in the NICU setting. As an independent approach to further assess the role of vancomycin resistance in the epidemic success, we performed a second GWAS screen of the outbreak clade isolates using the THD success index as a quantitative trait (Fig. 4 and Extended Data Fig. 8). Variants in *sar*A and *mut*L, previously suspected to influence vancomycin resistance[26,27], were found to be significantly associated with epidemic success, consistently supporting a pivotal role of vancomycin resistance acquisition in the epidemic expansion of NRCS-A. Of note, the small number of alpha clade isolates prevented us using a parallel GWAS analysis to detect a convergent evolution signal. In addition, the success-associated substitutions found in the outbreak clade were not present in alpha clade isolates based on protein sequence alignment. This suggests that the observed phenotype convergence between these clades, in terms of vancomycin resistance and NICU-associated success, involved different evolutionary pathways rather than genetic homoplasy.

Further in line with the hypothesis of the selective benefit of vancomycin resistance in NICUs, higher glycopeptide MICs correlated with a neonatal origin of the isolates, both in bivariate analysis and after controlling for population structure (Fig. 5c), indicating that vancomycin resistance and a NICU setting co-occur in the same strains, irrespective of their lineage. Collectively, these findings point to vancomycin selective pressure as a major driving force of the NRCS-A epidemic in NICUs, and to a triangular interaction between elevated glycopeptide MICs, adaptation to NICU setting and epidemic success.

Discussion

Using the largest genome set of *S. capitis* from multiple countries ever investigated, we unravelled the evolutionary history and main drivers of the global spread of the *S. capitis* NRCS-A clone, which is highly associated with neonatal sepsis. We estimated that the common ancestor of this NICU-associated clone emerged around 1969. Despite the relatively large confidence interval around this estimation, this predicted date is consistent with the establishment of the first NICUs in North America and Europe in the 1960s (first in Yale New-Haven, USA in 1960, then in Paris, France in 1966 and Lausanne, Switzerland in 1967). This emergence was followed by a sharp increase in effective population size of the NRCS-A lineage in the mid 1980s. The results of THD analysis point to vancomycin resistance as a main driver of this expansion, as they reveal a triangular correlation between epidemic success of *S. capitis* strains, level of vancomycin resistance and neonatal infection. The driving role of vancomycin resistance in the success of the clone is further supported by results of the GWAS identifying vancomycin

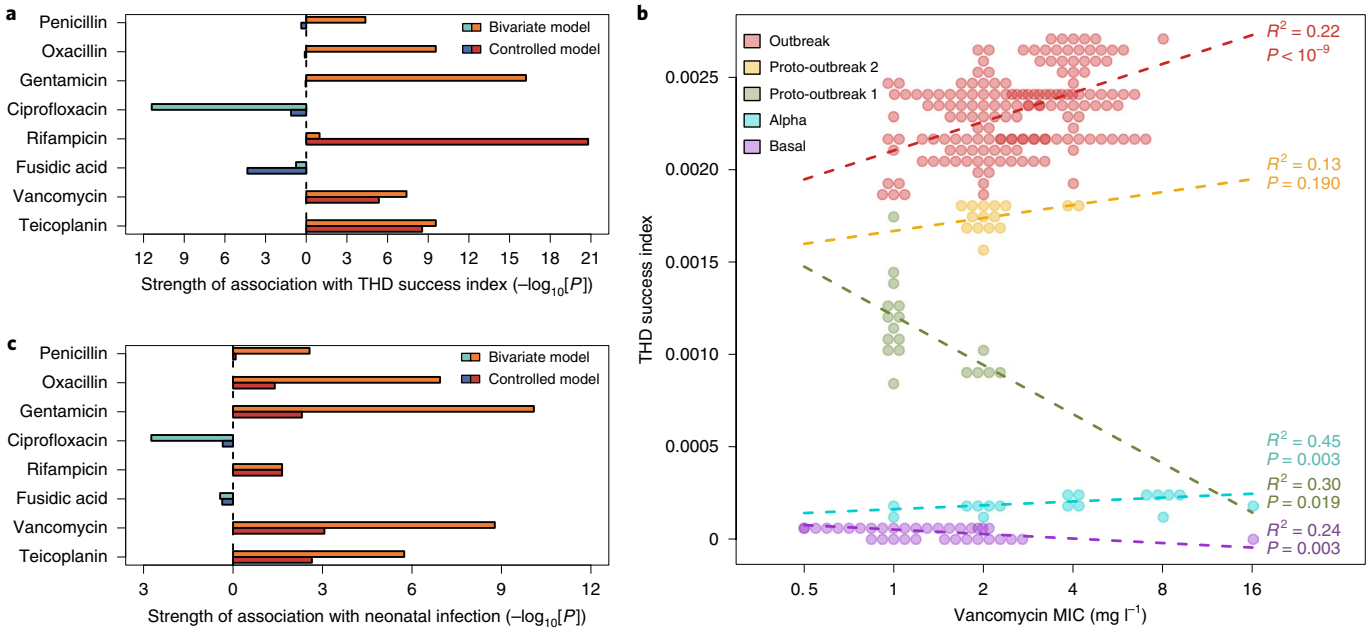


Fig. 5 | Associations of antibiotic resistance profiles with epidemic success and neonatal infection in *S. capitis* isolates. a, Strength of association of antibiotic MICs with THD success index in linear regression models controlled for population structure (controlled models) or not (bivariate models). Bars represent coefficient significance; bar direction denotes coefficient sign. **b**, Correlation of vancomycin MIC with the THD success index in different phylogenetic groups of *S. capitis*. Horizontal jitter was added to markers to reduce overlap. Dashed lines denote regression slopes, shown with R^2 coefficient and Pearson's correlation test *P* value. **c**, Strength of association of antibiotic MICs with a neonatal infection (as opposed to infection in adults) in logistic regression models controlled or not for population structure. $n = 250$ isolates.

resistance-related genes as associated with the success index of outbreak strains (such as *sarA*)[26], and results from Bayesian principal component analysis revealing an outbreak clade-specific SNP in the *glnQ* gene, which is also associated with vancomycin resistance[20].

These findings and conclusions are consistent with (1) the accelerated use of vancomycin in the 1980s[24,28,29], which may have favoured the wide dissemination of the vancomycin non-susceptible NRCS-A clone; (2) the extensive use of vancomycin, particularly in NICU settings[30], since cases of late-onset sepsis are frequent in hospitalized newborns and have the specificity to mainly involve methicillin-resistant CoNS[31]; (3) previous in vitro data demonstrating that NRCS-A is able to rapidly adapt under vancomycin selective pressure[32].

Strikingly, these patterns of epidemic success in a neonatal context associated with glycopeptide resistance emerge both in the outbreak clade and the independent alpha clade. In addition, alpha clade isolates also show high glycopeptide MICs and a specific non-synonymous SNP in a gene related to glycopeptide resistance (*murR*)[23]. This exceptional example of parallel phenotypic evolution in independent clades evolving in similar environments further supports the scenario that the epidemic success of NICU-adapted *S. capitis* lineages is mainly driven by vancomycin selection pressure. Moreover, the alpha clade strains acquired the same SCC*mec* type IV cassette as that found in the proto-outbreak 1 strains.

While vancomycin use appears to be the major determinant of success among NRCS-A isolates within a clade, the high levels of (additional) resistance to other antistaphylococcal drugs of frequent use in NICUs[30] (such as β-lactams and aminoglycosides) were also associated with differential epidemic success between *S. capitis* main clades as revealed by THD analysis. Furthermore, some country-specific patterns of NRCS-A strain success were consistent with specific local drug-selective pressure related to local practices of antimicrobial agent use. For instance, strong associations of THD success index with high rifampicin MIC was driven by a compact cluster of rifampicin-resistant isolates in the outbreak clade, mostly isolated in 2013–2014 in a NICU in Lyon, France. This specific pattern is consistent with high consumption of rifampicin from the 2010s in this NICU (F.L., personal communication). Similarly, fusidic acid resistance was specifically identified in NICU isolates from New Zealand, reflecting a local antibiotic selective pressure[33], as recently reported by Carter et al.[9]. In sum, the above observations represent strong indications that the successful specialization of NRCS-A for NICU environments results from its remarkable adaptation to antimicrobial selection pressures specific to NICUs. This conclusion is reminiscent of recent findings indicating that the emergence and global spread of nosocomial multidrug-resistant lineages of *S. epidermidis* has been driven by positive selection due to exposure to hospital environments and practices, such as single-antibiotic-impregnated medical devices[11].

Other genes and SNPs were identified as conserved in and unique to NRCS-A isolates, notably *nsr*, which encodes resistance to nisin. This bacteriocin is produced by lactococci and other bacteria involved in the early colonization of the gut, and is usually active against a wide range of gram-positive bacteria, including staphylococci and enterococci[34]. The establishment of NRCS-A in the gut microflora of newborns, which was shown to affect one third of newborns in a single-NICU prospective study[35], may therefore be favoured by nisin resistance. Another prominent candidate is the *tarJ* gene, involved in the biosynthesis of wall teichoic acids[36], which may therefore be involved in the ability of NRCS-A to colonize environmental surfaces in NICU settings and escape eradication by disinfection procedures. Furthermore, the exclusive presence of a composite SCC cassette (SCC*mec* type V and SCC*cad*/*ars*/*cop*) including a CRISPR element in the outbreak strains suggests a potential role of its acquisition in enhancing the success of this lineage. Further studies are needed to explore these hypotheses.

The origin of the NRCS-A clone and the paths of its worldwide dissemination remain unknown. A scenario of continuous and

repeated global seeding from the same, single source of contamination of globally used equipment or products as previously reported[37] had been initially evoked. This hypothesis is unsupported by our findings, as it would have induced a different tree topology and would have been expected to abolish the temporal signal in the evolutionary clock calculations. Nevertheless, the detection of some phylogeographical patterns based on the present genome analyses raises intriguing questions. Although the power of the present study is probably limited in terms of providing a definitive answer, the recurrent presence of isolates originating from Norway in the deepest nodes of the entire outbreak group as well as its two major subclades, dominated by French-speaking and Anglo-Saxon countries, respectively, points towards a putative source and plausible bifurcating directions of European dissemination.

Our Bayesian skyline analyses suggest some decline in the effective population size of NRCS-A from 2011 onwards. Because NRCS-A has been shown to colonize the NICU environment, especially in incubators[9], we hypothesize that this decline could be the consequence of recent general reinforcement of standard hygiene precautions and improvement in the routine disinfection of incubators[38].

In conclusion, our global framework provides the scientific community with a robust basis for further study of the evolutionary processes leading to the switch from poorly-virulent commensal and highly susceptible bacteria to multidrug-resistant clones that are able to infect vulnerable patients and to disseminate worldwide. This worrisome situation calls for urgent upscaling and sustained maintenance in the worldwide standards of antimicrobial stewardship, to prevent additional, repeated emergence of other multidrug-resistant clones.

Methods

Sample collection. The study collection was designed to be representative of the temporal, geographical and clinical diversity of *S. capitis* species. Therefore, a panel of clinical microbiology laboratories worldwide were contacted to send a sample of their own collection of clinical *S. capitis* isolates collected from blood culture to the French National Reference Center for Staphylococci, Hospices Civils de Lyon, France. In some laboratories, the absence of identification at the species level for CoNS strains, collected in blood cultures from newborns, did not allow for sampling of *S. capitis* in their collection, somewhat limiting the potential extent of the study sample. Each laboratory was asked to include their oldest and most recent strains still available. The French National Reference Center for Staphylococci subsequently selected representative isolates from each laboratory, including isolates from adults and newborns (when both were available) and covering the largest possible period of isolation. Of note, isolates from the NICU of Lyon, France, were relatively overrepresented to enhanced the search for possible local signals of evolution.

The species identification of each isolate was confirmed using matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS). After genomic analysis (see below), several isolates were excluded because of a misidentification ($n = 3$) or because of poor sequence data quality ($n = 4$). Two-hundred-and-fifty isolates were included in the final analysis, collected from blood cultures of newborns and adult patients, from 22 countries throughout the world between 1994 and 2015 (collection composition detailed in Supplementary Table 1).

DNA extraction, library preparation and sequencing. DNA extraction was carried out using a modified protocol for the QIAsymphony DSP DNA mini kit (Qiagen). In brief, a single colony of each test isolate was cultured on blood agar incubated overnight at 37 °C in aerobic conditions. As a pre-lysis step, 1 µl of bacterial growth was suspended in 230 µl lysis buffer comprising 203.6 µl of P1 buffer (Qiagen), 10 µl of 100 mg ml^{-1} lysozyme (Sigma), 10 µl of 1 mg ml^{-1} lysostaphin (Sigma), 2.4 µl of Triton X-100 (Sigma) and 4 µl RNAse A (Qiagen). The suspension was incubated for 30 min at 37 °C followed by digestion with 20 µl of proteinase K (from DSP DNA mini kit, Qiagen) and incubation at 56 °C for 2 h. Extraction of DNA from the treated cells was performed on the QIAsymphony SP platform (Qiagen) using the DSP DNA mini kit and eluted in 100 µl of ATE buffer (Qiagen).

Paired-end libraries of 100 bp read length were prepared using the Nextera XT DNA Sample Preparation kit (Illumina) and sequenced on an Illumina HiSeq 2500 platform under rapid run mode according to the manufacturer's instructions.

Mapping and genome analysis. Quality of Illumina paired-end reads was assessed with FastQC v.0.11.7. Potential contamination during the library preparation and/or sequencing was assessed using Kraken 2[39] to assign taxonomic labels to each read and check whether they were mainly (>90%) attributable to *S. capitis*. Kraken 2 was used with MiniKraken DB_8GB database. Three additional genomes were removed from the study because they were not identified as *S. capitis*.

Illumina paired-end reads were de novo assembled using the Unicycler v.0.4.5 pipeline[40] with default parameters that performed reads correction and trimming, assembly, scaffolding, removing overlap and bridging. Assemblies quality was assessed with QUAST v.4.6.3[41]. Four additional genomes were further excluded from our study due to low coverage. Genome annotation was performed with Prokka v.1.13. A previous study has highlighted 59 genes specifically present in *S. capitis* strains belonging to the NRCS-A clone[15]. The presence or absence of these genes in all the 250 sequenced strains was determined using SRST2[42] against a database made up of these 59 NRCS-A-specific genes.

A 90% threshold was retained to define the presence or the absence of a gene from a group of strains (that is proto-outbreak 1, proto-outbreak 2, outbeak and others). Thus, a gene was considered to be present in a group if it was found in more than 90% of strains in the group. In the same way, a gene was considered as absent if was not found in at least 90% of strains in the group. The type of SCCmec elements was determined using SCCmecFinder[43] or by blast analyses targeting *mec* element and *ccr* genes of the *ccr* cassettes from assembled genomes (SPAdes v.3.8.0). To confirm the presence status (complete, incomplete or absent) of the composite cassette previously described in the reference NRCS-A strain CR01[16], we used Bowtie2[44] to map the reads from all 250 sequenced strains against the full composite SCC*mec*-SCC*cad*/*cop*/*ars* element of strain CR01. The presence and typing of CRISPR elements was performed using CRISPRcasFinder[45].

Phylogenetics. The phylogenetic signal of the dataset was investigated using the likelihood-mapping method implemented in Tree-Puzzle 7.1[46] by analysing 10,000 random quartets. This method proceeds by evaluating, using maximum-likelihood groups of four randomly chosen sequences (quartets). The three possible unrooted tree topologies for each quartet are weighted, and the posterior weights are then plotted using triangular coordinates, such that each corner represents a fully resolved tree topology. Thus, the resulting distribution of the points shows whether the data are suitable for a phylogenetic reconstruction or not.

Phylogenetic relationships were reconstructed using the maximum-likelihood approach implemented in Phyml 3.412[47]. The robustness of the maximum-likelihood tree topology was assessed with bootstrapping analyses of 1,000 pseudoreplicated datasets. Phylogenies were rooted with the midpoint rooting option using FigTree software v.1.4. The profiles of drug resistances for each strain and cassette element were plotted on the maximum-likelihood tree using Itol[48]. Linear regression analysis of the root-to-tip distances against sampling time was performed using TempEst1.5[49]. To assess the robustness of our root-to-tip regression, we performed a permutation test of 5,000 replicates using the lmPerm package[50] in R, which was confirmed by tip randomization (see below). For the coalescent-based analyses, evolutionary rates and tree topologies were analysed using the general time-reversible and Hasegawa–Kishino–Yano substitution models with gamma distributed among-site rate variation with four rate categories (Γ4). The substitution rate was estimated under different demographic and clock models using Beast v.2.3.2[51], taking advantage of a sampling timeframe from 1994 to 2015. We tested both a strict molecular clock (which assumes the same evolutionary rates for all branches in the tree) and a relaxed clock that allows different rates among branches. To assess the robustness of the temporal inference, we performed ten additional runs after randomization of the sampling dates. Constant-sized and Bayesian skyline plot models, based on a general, non-parametric prior that enforces no particular demographic history, were used. For each model, two independent chains were conducted for 200 million generations and convergence was assessed by checking effective sample size values for key parameters using Tracer v.1.6. We used Tracer 1.6 to calculate the $\log_{10}$ Bayes factors to compare the models after a burning of 10% of the chain. Bayes factors represent the ratio of the marginal likelihood of the models being compared. Approximate marginal likelihoods for each coalescent model were calculated via importance sampling (1,000 bootstraps) using the harmonic mean of the sampled likelihoods. A ratio between 3 and 10 indicates moderate support that one model better fits the data than another, whereas values greater than 10 indicate strong support.

Recombination detection. We first initiated a visual inspection with the algorithm Splitstree3[52], followed by the Phi test. In a second step, ClonalFrame1.2[53] was run for 50,000 iterations on the whole-genome nucleotide alignment; the initial half was discarded as Markov chain Monte Carlo burn-in. Runs of ClonalFrame were performed three independent times and showed high congruence in their reconstructed phylogenies and recombination events. Numbers of mutation and recombination events were computed for each reconstructed branch substitution event introduced by either mutation or recombination. We then estimated the relative effect of recombination and mutation on genetic change (r/m) and the relative rate of mutation and recombination (ρ/θ, respectively) estimated by ClonalFrame, as previously described[54].

Selection signatures. Additionally, to capture SNPs that may explain the outbreak success, we used the R version of the software Pcadapt[17] to perform a genome scan on the basis of a Bayesian factor model. We chose $K=4$ factors and selected SNPs with the highest Mahalanobis distance and associated with the principal component 2, separating basal strains (proto-outbreak 1 and 2) from the NRCS-A proper outbreak. The factor analysis was performed on the centred genotype matrix that was not scaled. The Markov chain Monte Carlo algorithm was initialized using singular value decomposition, and the total number of steps was equal to 400 with a burn-in of 200 steps.

Protein structure. The mutated protein structures were predicted by Phyre2 software online (http://www.sbg.bio.ic.ac.uk/phyre2)[55] and visualized with UCSF Chimera[56].

Genotypes and phenotypes of antibiotic resistance. The of antibiotics (penicillin, oxacillin, ciprofloxacin, moxifloxacin, erythromycin, clindamycin, tetracycline, daptomycin, fusidic acid, gentamicin, linezolid, quinupristin-dalfopristin, rifampicin, teicoplanin and vancomycin) were determined by agar dilution[57] and interpreted in accordance with European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines (http://www.eucast.org). Susceptibility to cefoxitin was assessed by Etest (Biomérieux) and the results were interpreted using EUCAST criteria. Heteroresistance to vancomycin was assessed as described previously[58].

Biomarkers of antibiotic resistance were detected using GeneFinder (M. Doumith, unpublished). In brief, sequence reads were mapped against a set of reference sequences using Bowtie2. The criteria for gene detection were (1) read coverage 100% of reference length, (2) >90% homology to reference sequence and (3) mean depth of sequencing > 10% of mean depth of chromosomal region controls (as a measure of possible contamination). Non-synonymous chromosomal mutations were reported as amino acid position except for 23S rRNA, where they were reported as nucleotide position.

Next, THD analysis was performed. THD success index was computed as described[25] on the basis of the matrix of genetic distances between isolates (SNP counts). The THD approach converts genetic distances to a density measure in the space of haplotypes, which captures the number of putative between-host transmission events in the isolate's ancestry during a user-specified time window or timescale. This approach was previously shown to accurately capture the influence of well-established drivers of epidemic transmission of *Mycobacterium tuberculosis* and to detect novel potential drivers[25,59]. User-defined parameters were a mutation rate of 1.51×10^{-6} substitutions per position per year, as defined under the best-fitting evolutionary model (Fig. 2c); a genome size of 2.52×10^{6} bp as defined previously in the *S. capitis* NRCS-A prototype strain CR01[15], and a timescale of 20 yr, reflecting the approximate duration of the 1990–2010 period of maximal effective population size in the sampled collection (Fig. 2d). The resulting THD success index was then used as a response variable in linear regression models including antibiotic MICs as predictors and, where specified, the phylogenetic group as a covariate to control for population structure. All analyses were conducted using R v.3.3.2.

GWAS. DBGWAS[60] was used to identify coding-sequence variations associated with: (1) vancomycin MIC and (2) the THD epidemic success index. DBGWAS is a recently introduced GWAS method that tests the association between the phenotype and the presence of fixed-length subsequences (termed *k*-mers) in the genome. *k*-Mer presence can denote genetic variants ranging from SNPs to large mobile genetic elements, making *k*-mer based GWAS a flexible and hypothesis-free approach. DBGWAS agglomerates sets of overlapping *k*-mers as single components using a De Bruijn graph and annotates the *k*-mers to enable gene-level interpretation.

Due to the strong population structure of our complete collection and the dominance of the outbreak isolates, a GWAS analysis of the entire dataset would have identified genes canonical to the outbreak clade rather than genes truly associated with the outcome. Thus, GWAS analyses were restricted to the outbreak clade isolates to better identify within-clade determinants of success and resistance.

We ran DBGWAS 0.5.4 with default settings and no *P* value-based selection threshold to preserve all the potential *k*-mers. All *k*-mers connected to the same gene were agglomerated via their HMP using the R package Harmonicmeanp. Clusters of orthologous groups (COG) annotation for each gene was performed using MaGe[61]. A box plot of the $-\log_{10}$[HMP] significance values was constructed with COG categories as groups to identify associations between COG categories and the model outcome. To identify gene variants associated with vancomycin MIC, epidemic success or both, we constructed a scatter plot of each gene where the X and Y axis represented the $-\log_{10}$[HMP] in the vancomycin MIC and epidemic success GWAS, respectively. Genes with a $-\log_{10}$[HMP] > 7.5 on either axis, or >5 on both axes, were deemed significant.

Phenotypic assays. To identify the phenotypic features of the outbreak clade that have led to its success in NICUs worldwide, several phenotypic assays were performed on a subset of representative isolates. Three to eight isolates (depending on the experiments) were randomly selected from each of the four subgroups identified by the phylogeographical analysis (outbreak group, proto-outbreak 1 group, proto-outbreak 2 group and 'other isolates' group). First, because a previous study has reported the higher morbidity of *S. capitis* NRCS-A-related neonatal sepsis when compared to other CoNS[10], we investigated the virulence of 12 representative *S. capitis* isolates using the culture supernatants cytotoxicity assay, as described elsewhere, using THP1 cells[62]. Second, *S. capitis* has been shown to persist in the inert environment of NICUs[9], which is why the tolerance to desiccation of 12 isolates was investigated using a previously described method[63]. The kinetics of bacterial growth of 24 representative isolates were also tested in standard conditions[64] and under oxidative stress using an ethanol-supplemented medium to a final concentration of 6.5%, as previously described[65].

In addition, several phenotypic tests were targeted on the basis of the identified discriminant SNPs. Since a specific SNP was identified in the gene *fru*K, previously shown to be essential for biofilm formation[66], the biofilm production was evaluated on 24 representative isolates using crystal violet method in standard conditions and in 1% glucose supplemented media[67]. Finally, the phagocytosis by leukocytes (monocytes, granulocytes, $CD16^+$ neutrophils and $CD16^+CD11b^+$ neutrophils) from whole cord blood was investigated for a subset of five representative isolates of outbreak and basal strain groups, as described elsewhere[68].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets supporting the results of this article are available from the Sequence Read Archive under accession no. PRJNA493527. Additional data on the 250 strains are available in Supplementary Table 1.

Received: 19 October 2019; Accepted: 28 January 2020;
Published online: 27 April 2020

References

1. Howson, C. P., Kinney, M. V., McDougall, L. & Lawn, J. E., Born too soon preterm birth action group. Born too soon: preterm birth matters. *Reprod. Health* **10**, S1 (2013).
2. Liu, L. et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the sustainable development goals. *Lancet* **388**, 3027–3035 (2016).
3. Boghossian, N. S. et al. Late-onset sepsis in very low birth weight infants from singleton and multiple-gestation births. *J. Pediatr.* **162**, 1120–1124 (2013).
4. Stoll, B. J. et al. Late-onset sepsis in very low birth weight neonates: the experience of the NICHD neonatal research network. *Pediatrics* **110**, 285–291 (2002).
5. Cohen-Wolkowiez, M. et al. Early and late onset sepsis in late preterm infants. *Pediatr. Infect. Dis. J.* **28**, 1052–1056 (2009).
6. Rasigade, J. P. et al. Methicillin-resistant *Staphylococcus capitis* with reduced vancomycin susceptibility causes late-onset sepsis in intensive care neonates. *PLoS ONE* **7**, e31548 (2012).
7. Butin, M. et al. Wide geographical dissemination of the multiresistant *Staphylococcus capitis* NRCS-A clone in neonatal intensive-care units. *Clin. Microbiol. Infect.* **22**, 46–52 (2016).
8. Butin, M., Martins-Simoes, P., Rasigade, J. P., Picaud, J. C. & Laurent, F. Worldwide endemicity of a multidrug-resistant *Staphylococcus capitis* clone involved in neonatal sepsis. *Emerg. Infect. Dis.* **23**, 538–539 (2017).
9. Carter, G. P. et al. Genomic analysis of multi-resistant *Staphylococcus capitis* associated with neonatal sepsis. *Antimicrob. Agents Chemother.* **62**, e00898–18 (2018).
10. Ben Said, M. et al. Late-onset sepsis due to *Staphylococcus capitis* 'neonatalis' in low-birthweight infants: a new entity? *J. Hosp. Infect.* **94**, 95–98 (2016).
11. Lee, J. Y. H. et al. Global spread of three multidrug-resistant lineages of *Staphylococcus epidermidis*. *Nat. Microbiol.* **3**, 1175–1185 (2018).
12. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).
13. Stegger, M. et al. Origin and evolution of European community-acquired methicillin-resistant *Staphylococcus aureus*. *mBio* **5**, e01044–14 (2014).
14. Nubel, U. et al. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog.* **6**, e1000855 (2010).
15. Simoes, P. M. et al. Single-molecule sequencing (PacBio) of the *Staphylococcus capitis* NRCS-A clone reveals the basis of multidrug resistance and adaptation to the neonatal intensive care unit environment. *Front. Microbiol.* **7**, 1991 (2016).
16. Martins Simoes, P. et al. Characterization of a novel composite staphylococcal cassette chromosome *mec* (SCC*mec*-SCC*cad/ars/cop*) in the neonatal sepsis-associated *Staphylococcus capitis* pulsotype NRCS-A. *Antimicrob. Agents Chemother.* **57**, 6354–6357 (2013).

17. Duforet-Frebourg, N., Bazin, E. & Blum, M. G. B. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Mol. Biol. Evol.* **31**, 1–13 (2014).
18. Merker, M. et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* **47**, 242–249 (2015).
19. Beabout, K. et al. The ribosomal S10 protein is a general target for decreased tigecycline susceptibility. *Antimicrob. Agents Chemother.* **59**, 5561–5566 (2015).
20. Kuroda, M., Kuwahara-Arai, K. & Hiramatsu, K. Identification of the up- and down-regulated genes in vancomycin-resistant *Staphylococcus aureus* strains Mu3 and Mu50 by cDNA differential hybridization method. *Biochem. Biophys. Res. Commun.* **269**, 485–490 (2000).
21. Grayczyk, J. P., Harvey, C. J., Laczkovich, I. & Alonzo, F. 3rd A lipoylated metabolic protein released by *Staphylococcus aureus* suppresses macrophage activation. *Cell Host Microbe* **22**, 678–687 (2017).
22. Zorzoli, A., Grayczyk, J. P. & Alonzo, F. *III. Staphylococcus aureus* tissue infection during sepsis is supported by differential use of bacterial or host-derived lipoic acid. *PLoS Pathog.* **12**, e1005933 (2016).
23. Hu, Q., Peng, H. & Rao, X. Molecular events for promotion of vancomycin resistance in vancomycin intermediate *Staphylococcus aureus*. *Front. Microbiol.* **7**, 1601 (2016).
24. Krzyzaniak, N., Pawlowska, I. & Bajorek, B. Review of drug utilization patterns in NICUs worldwide. *J. Clin. Pharm. Ther.* **41**, 612–620 (2016).
25. Rasigade, J.-P. et al. Strain-specific estimation of epidemic success provides insights into the transmission dynamics of tuberculosis. *Sci. Rep.* **7**, 45326 (2017).
26. Lamichhane-Khadka, R. et al. sarA inactivation reduces vancomycin-intermediate and ciprofloxacin resistance expression by *Staphylococcus aureus*. *Int. J. Antimicrob. Agents* **34**, 136–141 (2009).
27. Schaaff, F., Reipert, A. & Bierbaum, G. An elevated mutation frequency favors development of vancomycin resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **46**, 3540–3548 (2002).
28. Jacqz-Aigrain, E., Zhao, W., Sharland, M. & van den Anker, J. N. Use of antibacterial agents in the neonate: 50 years of experience with vancomycin administration. *Semin. Fetal Neonatal Med.* **18**, 28–34 (2013).
29. Levine, D. P. Vancomycin: a history. *Clin. Infect. Dis.* **42**, S5–S12 (2006).
30. Mukhopadhyay, S., Sengupta, S. & Puopolo, K. M. Challenges and opportunities for antibiotic stewardship among preterm infants. *Arch. Dis. Child Fetal Neonatal Ed.* **104**, F327–F332 (2019).
31. Cailes, B. et al. Antimicrobial resistance in UK neonatal units: neonIN infection surveillance network. *Arch. Dis. Child Fetal Neonatal Ed.* **103**, F474–F478 (2018).
32. Butin, M. et al. Adaptation to vancomycin pressure of multiresistant *Staphylococcus capitis* NRCS-A involved in neonatal sepsis. *J. Antimicrob. Chemother.* **70**, 3027–3031 (2015).
33. Williamson, D. A. et al. High usage of topical fusidic acid and rapid clonal expansion of fusidic acid-resistant *Staphylococcus aureus*: a cautionary tale. *Clin. Infect. Dis.* **59**, 1451–1454 (2014).
34. Millette, M. et al. Capacity of human nisin- and pediocin-producing lactic acid bacteria to reduce intestinal colonization by vancomycin-resistant enterococci. *Appl. Environ. Microbiol.* **74**, 1997–2003 (2008).
35. Butin, M. et al. Vancomycin treatment is a risk factor for vancomycin-nonsusceptible *Staphylococcus capitis* sepsis in preterm neonates. *Clin. Microbiol. Infect.* **23**, 839–844 (2017).
36. Brown, S., Santa Maria, J. P. Jr & Walker, S. Wall teichoic acids of gram-positive bacteria. *Annu. Rev. Microbiol.* **67**, 313–336 (2013).
37. Nasser, R. M. et al. Outbreak of *Burkholderia cepacia* bacteremia traced to contaminated hospital water used for dilution of an alcohol skin antiseptic. *Infect. Control Hosp. Epidemiol.* **25**, 231–239 (2004).
38. Ory, J. et al. Successful implementation of infection control measure in a neonatal intensive care unit to combat the spread of pathogenic multidrug resistant *Staphylococcus capitis*. *Antimicrob. Resist. Infect. Control* **8**, 57 (2019).
39. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
40. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595 (2017).
41. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
42. Inouye, M. et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6**, 90 (2014).
43. Kaya, H. et al. SCCmecFinder, a web-based tool for typing of staphylococcal cassette chromosome *mec* in *Staphylococcus aureus* using whole-genome sequence data. *mSphere* **3**, e00612–17 (2018).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
45. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
46. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
47. Guindon, S. et al. New algorithms and mehtods to estimate maximum-likelihood phylogenies: asessing the performance of PhyML 2.0. *Syst. Biol.* **59**, 307–321 (2010).
48. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nuc. Acids Res.* **44**, W242–W245 (2016).
49. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
50. Anderson, M. J. & Robinson, J. Permutation tests for linear models. *Aust. NZ J. Stat.* **43**, 75–88 (2001).
51. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
52. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
53. Didelot, X., Lawson, D., Darling, A. & Falush, D. Inference of homologous recombination in bacteria using whole genome sequences. *Genetics* **186**, 1435–1449 (2010).
54. Joseph, S. J., Didelot, X., Gandhi, K., Dean, D. & Read, T. D. Interplay of recombination and selection in the genomes of *Chlamydia trachomatis*. *Biol. Direct* **6**, 28 (2011).
55. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
56. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Andrews, J. M. Determination of minimum inhibitory concentrations. *J. Antimicrob. Chemother.* **48**, 5–16 (2001).
58. Satola, S. W., Farley, M. M., Anderson, K. F. & Patel, J. B. Comparison of detection methods for heteroresistant vancomycin-intermediate *Staphylococcus aureus*, with the population analysis profile method as the reference method. *J. Clin. Microbiol.* **49**, 177–183 (2011).
59. Barbier, M. et al. Changing patterns of human migrations shaped the global population structure of *Mycobacterium tuberculosis* in France. *Sci. Rep.* **8**, 5855 (2018).
60. Jaillard, M. et al. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet.* **14**, e1007758 (2018).
61. Vallenet, D. et al. MaGe: a microbial genome annotation system supported by synteny results. *Nuc. Acids Res.* **34**, 53–65 (2006).
62. Maali, Y. et al. Understanding the virulence of *Staphylococcus pseudintermedius*: a major role of pore-forming toxins. *Front. Cell Infect. Microbiol.* **8**, 221 (2018).
63. Loftus, R. W., Dexter, F., Robinson, A. D. M. & Horswill, A. R. Desiccation tolerance is associated with *Staphylococcus aureus* hypertransmissibility, resistance and infection development in the operating room. *J. Hosp. Infect.* **100**, 299–308 (2018).
64. Karauzum, H. et al. Comparison of adhesion and virulence of two predominant hospital-acquired methicillin-resistant *Staphylococcus aureus* clones and clonal methicillin-susceptible *S. aureus* isolates. *Infect. Immun.* **76**, 5133–5138 (2008).
65. Streker, K., Freiberg, C., Labischinski, H., Hacker, J. & Ohlsen, K. *Staphylococcus aureus* NfrA (SA0367) is a flavin mononucleotide-dependent NADPH oxidase involved in oxidative stress response. *J. Bacteriol.* **187**, 2249–2256 (2005).
66. Loo, C., Mitrakul, K., Voss, I., Hughes, C. & Ganeshkumar, N. Involvement of an inducible fructose phosphotransferase operon in *Streptococcus gordonii* biofilm formation. *J. Bacteriol.* **185**, 6241–6254 (2003).
67. Valour, F. et al. *Staphylococcus epidermidis* in orthopedic device infections: the role of bacterial internalization in human osteoblasts and biofilm formation. *PLoS ONE* **8**, e67240 (2013).
68. Tissieres, P. et al. Innate immune deficiency of extremely premature neonates can be reversed by interferon-γ. *PLoS ONE* **7**, e32863 (2012).

Acknowledgements

We thank M. Stegger and his team for insightful exchanges during the manuscript drafting and C. Allix-Béguec, C. Gaudin, M. Mairey and S. Duthoy for their help in genome sequencing. This project was supported by the European Society of Clinical Microbiology and Infectious Diseases study group (Project P307-14), the Fondation pour la Recherche Médicale (project ING20160435683) and the European Union Patho-Ngen-Trace (project FP7-278864).

Author contributions

M.Butin, T.W., J.-C.P. and F.L. conceived the project. M.Butin and F.L. established and analysed clinical and reference isolate datasets. B.P., A.K. and R.P. performed DNA extractions. P.S. performed DNA sequencing. B.P., A.K. and R.P. performed antimicrobial susceptibility testing. P.T. performed phagocytosis assays. M.Butin performed all additional phenotypic assays. T.W., M.Barbier, P.M.-S. and M.Bergot analysed genomic

data. J.-P.R. participated in genomic analyses and performed THD analysis. M.Bergot and L.J. performed GWAS analysis. T.W., M.Butin, P.S. and F.L. drafted the manuscript. All authors reviewed and contributed to the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41564-020-0676-2.

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-0676-2.

Correspondence and requests for materials should be addressed to T.W. or M.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

The International Consortium for Staphylococcus capitis neonatal sepsis

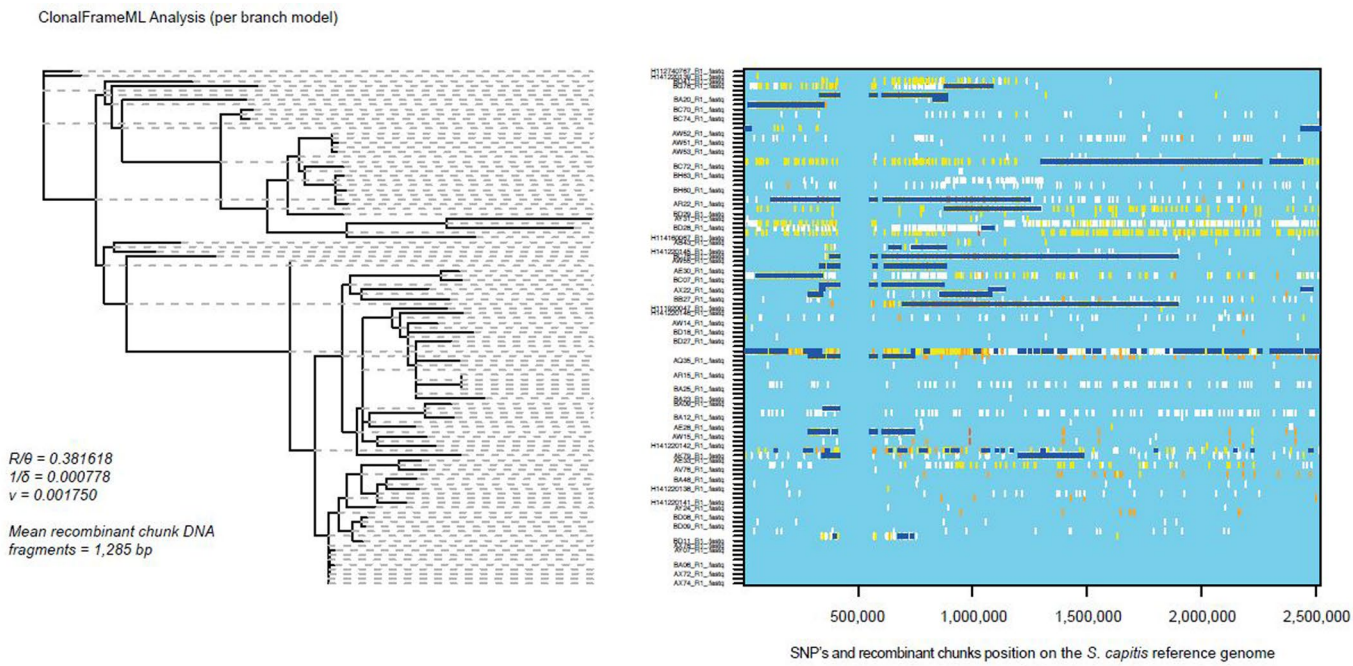
Vaclava Adamkova[12], Timothy Barkham[13], Karsten Becker[14], Desiree Bennett[15], Olivier Claris[11], Clarence Buddy Creech[16], Herminia De Lencastre[17], Margaret Deighton[18], Olivier Denis[19], John Ferguson[20], Yhu-Chering Huang[21], Claus Klingenberg[22], Andre Ingebretsen[23], Celine Laferrière[24], Katia Regina Netto dos Santos[25], Jacques Schrenzel[26], Iris Spiliopoulou[27], Stefania Stefani[28], Kim TaekSoo[29], Eveliina Tarkka[30], Alex Friedrich[31], Christina Vandenbroucke-Grauls[32], James Ussher[33], Francois Vandenesch[4] and Lars Westblade[34]

[12]Institute of Medical Biochemistry and Laboratory Diagnostics, General University Hospital, Prague, Czech Republic. [13]Tan Tock Seng Hospital, Singapore, Singapore. [14]University Hospital Münster, Münster, Germany. [15]Irish Meningitis and Sepsis Reference Laboratory, Temple Street Children's University Hospital, Dublin, Ireland. [16]Vanderbilt University School of Medicine, Nashville, USA. [17]Instituto de Tecnologia Quimica e Biologica, Oeiras, Portugal. [18]School of Applied Sciences, RMIT University, Bundoora, Australia. [19]Hôpital Erasme – ULB, Bruxelles, Belgique. [20]University of New Castle, Callaghan, Australia. [21]Chang Gung Children's Hospital, Taoyuan, Taiwan. [22]University Hospital of North Norway, Tromsø, Norway. [23]Oslo University Hospital Rikshospitalet, Oslo, Norway. [24]CHU Sainte-Justine, Montréal, Canada. [25]Rio de Janeiro Federal University, Rio de Janeiro, Brazil. [26]Laboratory of Bacteriology and the Genome Research Lab, Geneva University Hospital, Geneva, Switzerland. [27]University of Patras, Patras, Greece. [28]University of Catania, Catania, Italy. [29]Seoul National University Hospital, Seoul, Korea. [30]Helsinki University Central Hospital laboratory HUSLAB, Helsinki, Finland. [31]Medical Microbiology and Infection Prevention, University Medical Center, Groningen, Netherlands. [32]Medical Microbiology & Infection Control, Amsterdam, Netherlands. [33]University of Otago, Otago, New Zealand. [34]Children Healthcare of Atlanta, Atlanta, USA.

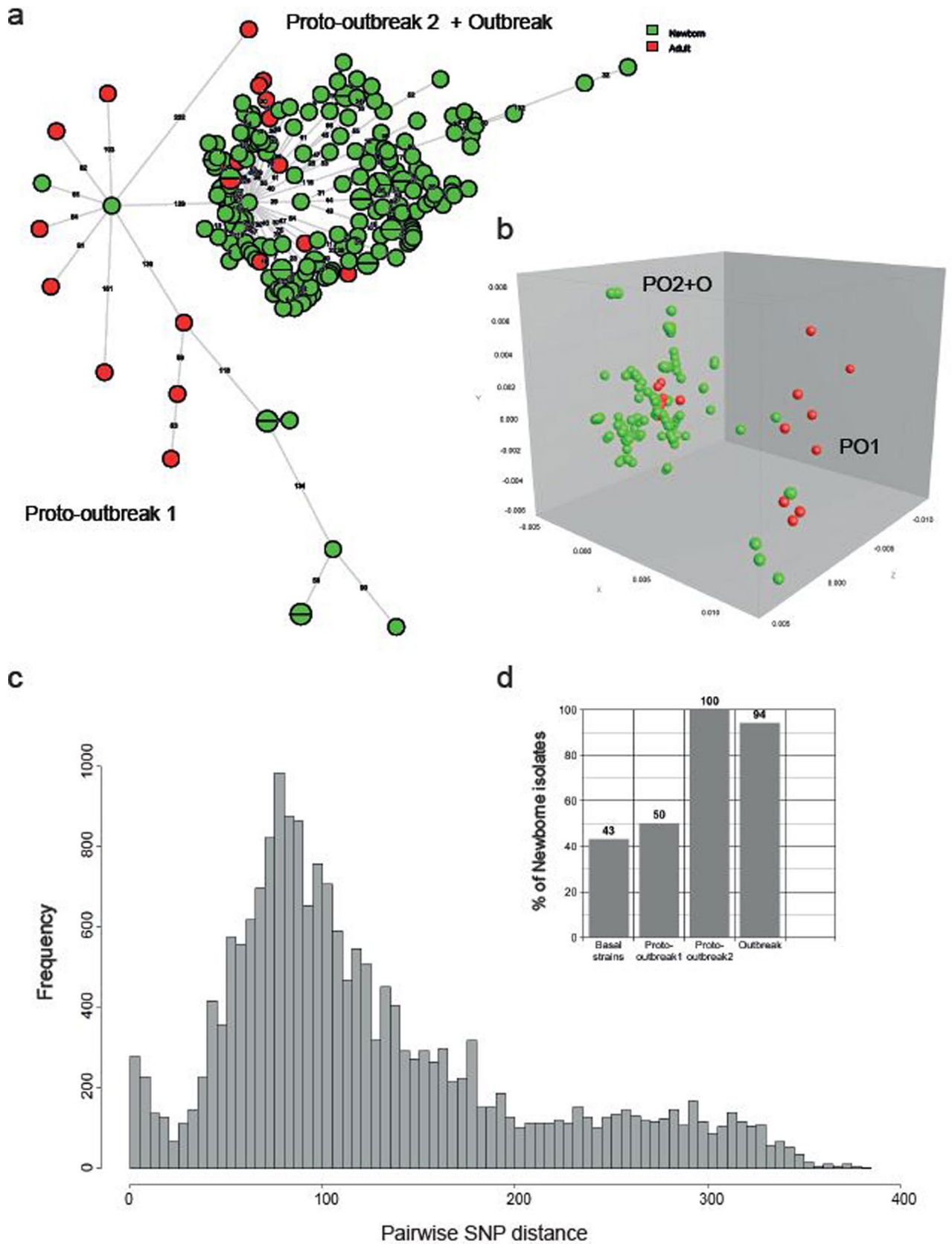
The ESGS Study Group of ESCMID

Jodi Lindsay[35], Francois Vandenesch[4], Anders Rhod Larsen[36], Philipp Zanger[37], Barbara C. Kahl[38] and Cristina Prat Aymerich[39]

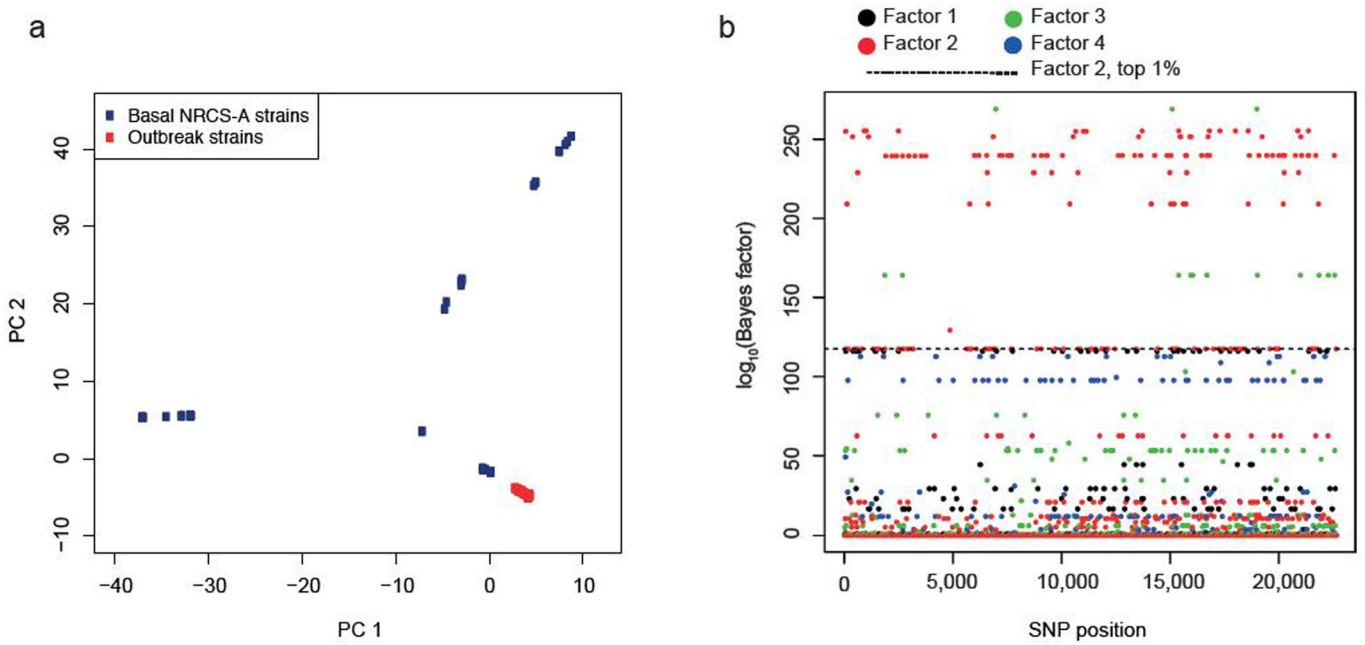
[35]Institute for Infection and Immunity, St George's, University of London, London, United Kingdom. [36]Statens Serum Institut, Microbiology and Infection Control, Reference Laboratory for Antimicrobial Resistance and Staphylococci, Copenhagen, Denmark. [37]Institute of Global Health, Epidemiology & Biostatistics, Ruprecht Karls University, Heidelberg, Germany. [38]Institut für Med. Mikrobiologie Universitätsklinikum Münster, Münster, Germany. [39]Hospital Universitari Germans Trias i Pujol, Microbiology, Badalona, Spain.



Extended Data Fig. 1 | CLONALFRAMEML analysis of recombination in *S. capitis*. Analysis was based on 55 genomes: all non-NRCS-A strains were included, however the clone NRCS-A was undersampled to avoid a statistical bias in favor of mutational changes. Dark blue horizontal bars indicate recombination events detected by the analysis.



Extended Data Fig. 2 | NRCS-A host types and genetic structure. a, NRCS-A isolates within an MSTREE based on the whole genome sequencing data. Each strain is represented by a circle or a fraction of a circle, colors correspond to different host types. Numbers indicate the mutational steps between the strains. **b**, Same data as above but represented in an MDS plot. **c**, Within NRCS-A diversity as assessed by mean pairwise SNP distances (N=197). **d**, Graphical chart representing the fraction of strains obtained from newborns in the basal, Proto-outbreak 1 and 2 and Outbreak strains.

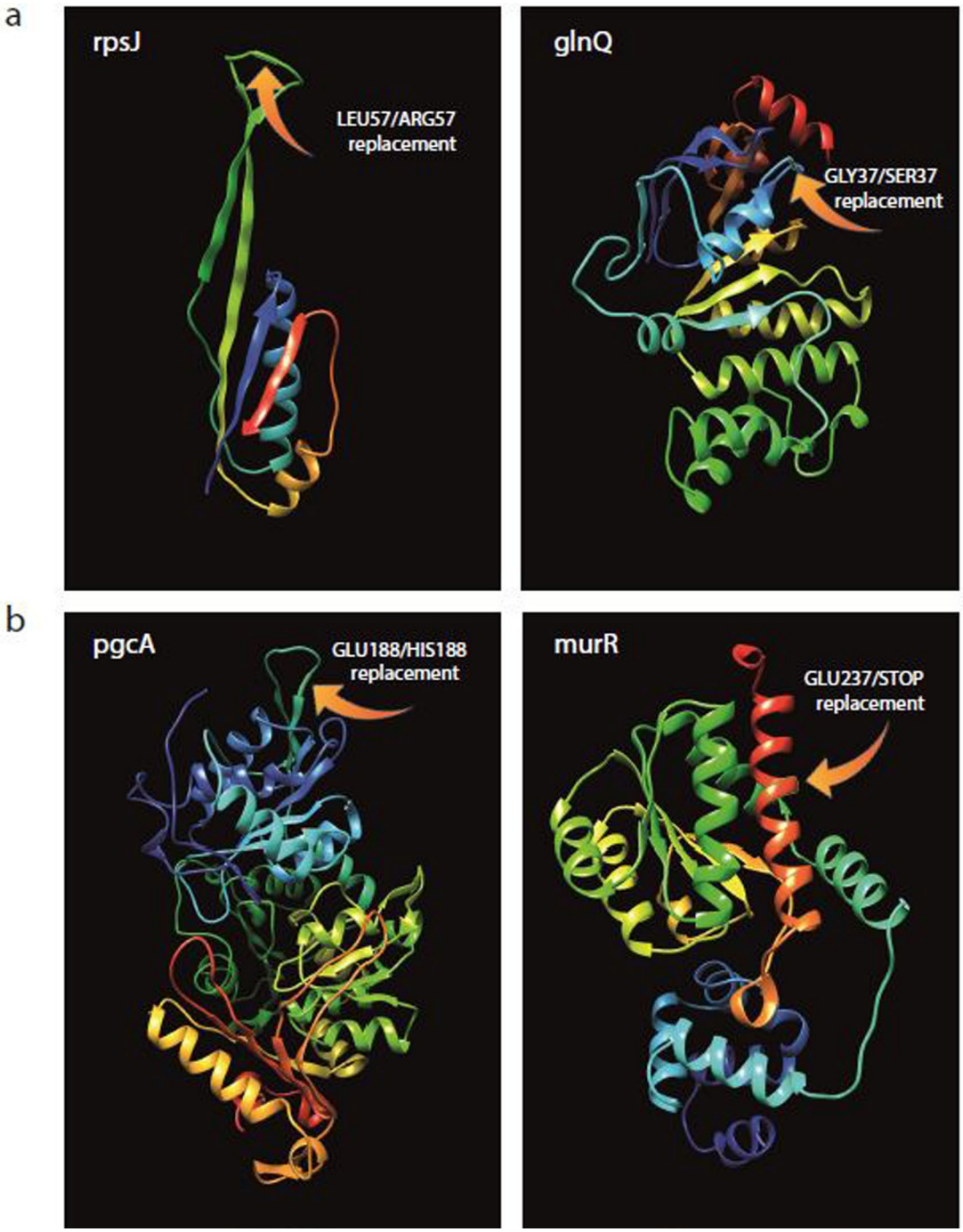


Extended Data Fig. 3 | Genome scan analysis of NRCS-A strains for detecting SNPs involved in local adaptation. a, Plot of the first 2 principal components (PC). The 197 NRCS-A strains are represented by points and colorized according to their phylogenic origin (Proto-outbreak 1 and 2 in blue, and Outbreak in red). PC 2 is the one separating the basal proto-outbreak 1 and 2 strains from the outbreak strains. **b**, Manhattan plot representing the 3,658 SNPs and values obtained after performing Mahalanobis distances. The SNPs are colorized according to the PC to which they correlate most (PC1 = black, PC2 = red, PC3 = green and PC4 = blue).

SNP position	nucleotide change	variant	aminoacid change	ORF	gene	product	COG_class*
Outbreak							
11548	C --> A	non synonymous	G219C	CR01_v3_0013	fruK	fructose-1-phosphate kinase	G
53011	A --> C	non synonymous	F140L	CR01_v3_0057	-	Lipase LipA	I
131775	A --> G	inter-genic	-	-	-	-	-
140199	G --> A	synonymous	-	CR01_v3_0161	yojF	conserved hypothetical protein	-
150113	A --> C	synonymous	-	CR01_v3_0172	-	putative phosphoglycolate phosphatase	R
315740	A --> G	synonymous	-	CR01_v3_0321	nfrA	NADPH-dependent oxidoreductase	C
803679	C --> T	synonymous	-	CR01_v3_0782	-	conserved protein of unknown function	S
1086981	G --> A	non synonymous	V377M	CR01_v3_1054	yjgC	putative molybdoenzyme	R
1098488	T --> A	synonymous	-	CR01_v3_1066	nhaC	Na+/H+ antiporter NhaC	C
1131838	C --> G	non synonymous	A302G	CR01_v3_1106	swrC	Swarming motility protein SwrC	V
1142227	A --> T	synonymous	-	CR01_v3_1114	pbuG	hypoxanthine/guanine permease	R
1143612	A/T --> G	non synonymous	K/M57R	CR01_v3_1116	rpsJ	ribosomal protein S10 (BS13)	J
1377186	G --> T	non synonymous	K139N	CR01_v3_1353	gatA	glutamyl-tRNA(Gln) amidotransferase (subunit A)	J
1403377	G --> A	non synonymous	D197N	CR01_v3_1382	-	ABC transporter, ATP-binding protein	GM
1426162	G --> A	non synonymous	G37S	CR01_v3_1399	glnQ	glutamine ABC transporter (ATP-binding protein)	E
1485698	T --> A	non synonymous	S581T	CR01_v3_1460	-	Cell wall surface anchor family protein	UDLW
1572754	G --> A	synonymous	-	CR01_v3_1525	polA	DNA polymerase I	L
1579236	T --> A	non synonymous	N453K	CR01_v3_1530	-	Replication initiation and membrane attachment protein, DnaB/DnaD family	L
1580294	A --> G	inter-genic	-	-	-	-	-
1634026	G --> A	non synonymous	M170I	CR01_v3_1581	tcdA	tRNA threonylcarbamoyladenosine dehydratase (t(6)A37 dehydratase)	H
1640526	T --> A	synonymous	-	CR01_v3_1588	-	Tetratricopeptide repeat protein	R
1724574	A --> C	non synonymous	K313Q	CR01_v3_1679	xseA	exodeoxyribonuclease VII (large subunit)	L
1730310	G --> A	synonymous	-	CR01_v3_1685	bkdAA	branched-chain alpha-keto acid dehydrogenase E1 subunit	C
1776544	G --> A	non synonymous	D103N	CR01_v3_1729	-	conserved protein of unknown function	S
1843408	G --> T	non synonymous	G69V	CR01_v3_1769	yteP	putative permease	G
1901658	A --> G	non synonymous	L27A	CR01_v3_1825	trpG	Anthranilate synthase component II	EH
1901659	A --> G	non synonymous	L27A	CR01_v3_1825	trpG	Anthranilate synthase component II	EH
1994747	A --> T	synonymous	-	CR01_v3_1916	rny	endoribonuclease Y	R
2211882	A --> G	non synonymous	K398E	CR01_v3_2115	-	Cation ABC transporter	P
2213806	A --> G	synonymous	-	CR01_v3_2117	purD	phosphoribosylglycinamide synthetase	F
2231951	G --> T	non synonymous	E139*	CR01_v3_2135	rihC	fragment of ribonucleoside hydrolase 3 (part 1); Uridine nucleosidase	F
2334298	T --> A	inter-genic	-	-	-	-	-
Clade alpha							
864389	A --> G	synonymous	K217K	CR01_v3_0841	capA	CapA protein	M
1007789	T --> G	non synonymous	Y116D	CR01_v3_0976	ycnI-like	Reeler-like domain of YcnI and similar proteins	S
1428895	T --> C	synonymous	F60F	CR01_v3_1402	nagB-like	Glucosamine-6-phosphate isomerase family protein	G
1487555	A --> G	non synonymous	K1200E	CR01_v3_1460	-	YSIRK signal domain/LPXTG anchor domain surface protein	S
1070078	A --> C	non synonymous	*237E	CR01_v3_1038	murR/rpiR	NADPH-dependent oxidoreductase	K
1372996	T --> A	non synonymous	S582T	CR01_v3_1349	ligA	NAD-dependant DNA ligase	L
1532960	T --> C	non synonymous	C17R	CR01_v3_1493	-	HAD family hydrolase	R
1578547	A --> C	non synonymous	K224Q	CR01_v3_1530	dnaB	helicase	L
1622733	A --> C	inter-genic	-	-	-	-	-
1632787	A --> G	non synonymous	M542V	CR01_v3_1579	aspS	aspartyl-tRNA synthetase	J
1648403	T --> G	non synonymous	L168V	CR01_v3_1594	-	O-methyltransferase family protein	R
1732483	A --> G	synonymous	G77G	CR01_v3_1687	bkdB	2-oxo acid dehydrogenase subunit E2	C
1922251	G --> A	synonymous	T58T	CR01_v3_1842	yneP	Putative acyl-CoA thioesterase YneP	R
2115224	A --> G	inter-genic	-	-	-	-	-
2185080	T --> C	non synonymous	N45S	CR01_v3_2088	-	conserved membrane protein of unknown function	J
2234931	G --> T	non synonymous	H188Q	CR01_v3_2138	pgcA	Phosphoglucomutase	G
2338628	T --> G	non synonymous	M96R	CR01_v3_2263	-	DsbA family protein	Q

*COG class (Clusters of Orthologous Groups of proteins) has been determined as described elsewhere (Nucleic Acids Res. 2000 Jan 1; 28(1): 33–36)

Extended Data Fig. 4 | Specific SNPs in Outbreak and Alpha isolates. Respectively 32 and 17 SNPs were specifically identified in Outbreak strains among NRCS-A strains (n=197) or in clade Alpha strains among Basal strains (n=53). Those SNPs were identified using PCADAPT.

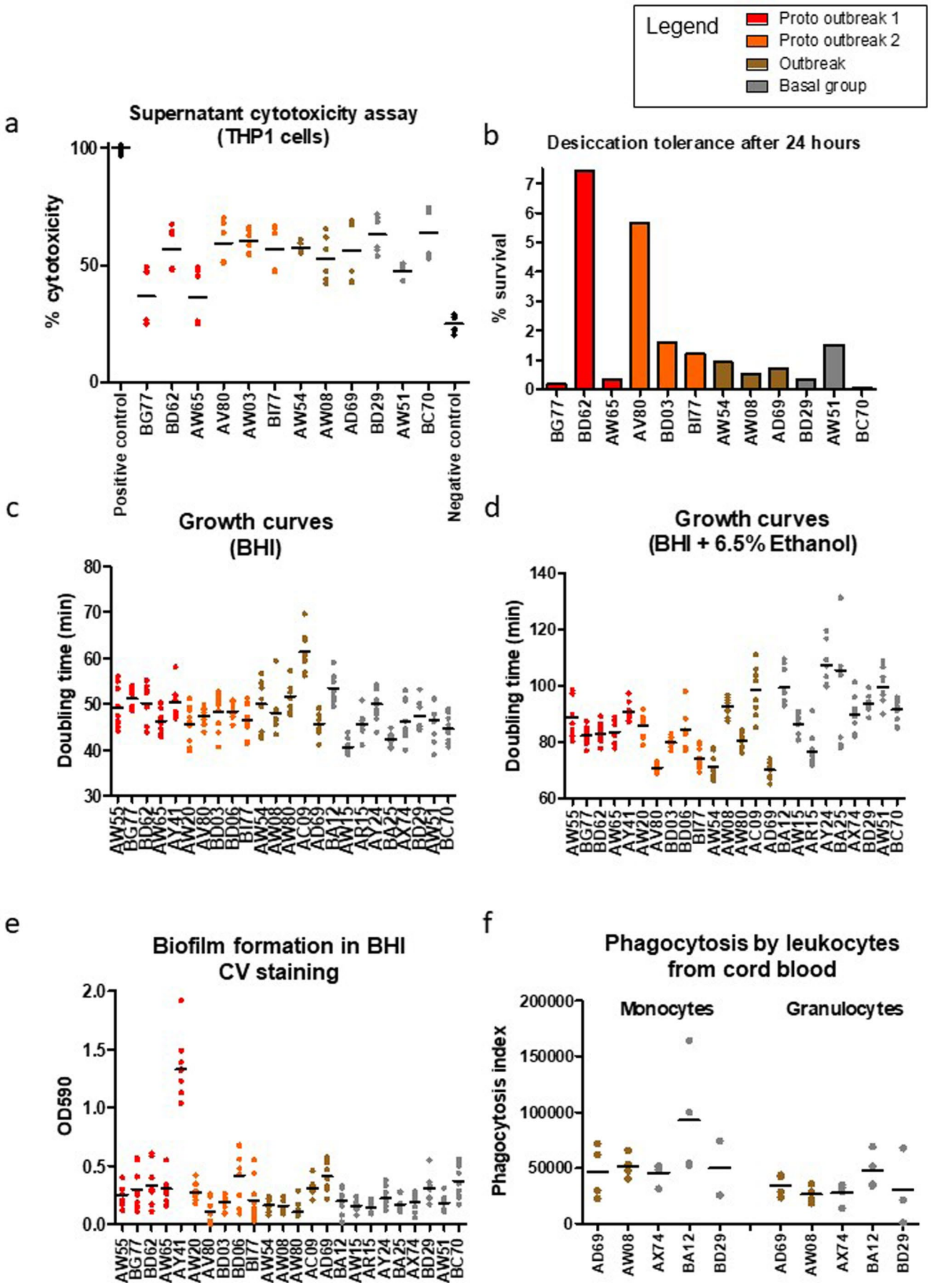


Extended Data Fig. 5 | Tertiary protein structures. a, Positions on the tertiary protein structure of outbreak specific non-synonymous mutations detected via PCADAPT and involved in antibiotic resistance (tigecycline and vancomycin). **b**, Positions on the tertiary protein structure of alpha-clone specific non-synonymous mutations for a set of two genes involved in cell wall synthesis. Visualization and predictions were executed by PHYRE2 software (http://www.sbg.bio.ic.ac.uk/phyre2).

	Outbreak strains	Proto outbreak 1 strains	Proto outbreak 2 strains	Basal strains	clade alpha strains	Comparison between outbreak and Basal strains*	Comparison between outbreak and Clade alpha*	Comparison between Basal strains and Clade alpha*
Total number of isolates, n	162	18	17	53	10	-	-	-
Phenotypic patterns, % (n resistant isolates/n tested isolates)								
Oxacillin resistance	96% (153/159)	67% (12/18)	100% (17/17)	66% (35/53)	100% (10/10)	$1.88.10^{-07}$	1	0.032
Penicillin resistance	98% (156/159)	94% (17/18)	100% (17/17)	81% (43/53)	100% (10/10)	0.0015	1	0.21
Gentamicin resistance	99% (157/159)	61% (11/18)	100% (17/17)	58% (31/53)	100% (10/10)	$2.55.10^{-12}$	1	0.0093
Ciprofloxacin resistance	6% (9/159)	22% (4/18)	59% (10/17)	45% (24/53)	100% (10/10)	$2.14.10^{-12}$	$2.2.10^{-16}$	0.00045
Linezolid resistance	0% (0/159)	0% (0/18)	0% (0/17)	2% (1/53)	10% (1/10)	0.76	0.06	0.42
Daptomycin resistance	60% (95/159)	38% (6/16)	65% (11/17)	10% (5/50)	10% (1/10)	$1.78.10^{-06}$	0.006	1
Clindamycin resistance	1% (2/148)	44% (8/18)	82% (14/17)	32% (17/53)	30% (3/10)	$2.2.10^{-16}$	$4.59.10^{-05}$	1
Erythromycin resistance	9% (14/159)	56% (10/18)	100% (17/17)	55% (29/53)	70% (7/10)	$2.2.10^{-16}$	$2.04.10^{-07}$	0.47
Tetracyclin resistance	4% (6/157)	6% (1/18)	6% (1/17)	9% (5/53)	0% (0/10)	0.28	1	0.59
Vancomycin resistance	64% (102/159)	11% (2/18)	76% (13/17)	25% (13/53)	100% (10/10)	$2.13.10^{-06}$	0.048	$8.91.10^{-09}$
Vancomycin heteroresistance	35% (56/159)	78% (14/18)	24% (4/17)	47% (25/53)	0% (0/10)	0.05	0.05	0.0030
Teicoplanin resistance	48% (76/159)	11% (2/18)	24% (4/17)	21% (11/53)	80% (8/10)	$1.81.10^{-05}$	0.10	$2.65.10^{-06}$
Fusidic acid resistance	30% (45/150)	17% (3/18)	41% (7/17)	43% (23/53)	10% (1/10)	0.30	0.32	0.044
Rifampicin resistance	21% (34/159)	0% (0/18)	0% (0/17)	25% (15/53)	100% (10/10)	0.51	$2.99.10^{-07}$	$2.01.10^{-07}$
Genotypic patterns, % (n isolates harboring resistance gene or mutation/n tested isolates)								
*mec*A gene	98% (158/162)	65% (11/17)	100% (17/17)	64% (34/53)	100% (10/10)	$2.63.10^{-09}$	1	0.024
*bla*Z gene	90% (145/162)	100% (17/17)	100% (17/17)	83% (44/53)	100% (10/10)	1	0.59	0.26
aac(6')-*aph*(2") genes	99% (160/161)	59% (10/17)	100% (17/17)	57% (30/53)	100% (10/10)	$6.70.10^{-14}$	1	0.0065
*grl*A_SC(80:S-F;84:D-N);*gyr*A_SC(84:S-F)	0% (0/162)	0% (0/17)	0% (0/17)	11% (6/53)	0% (0/10)	0.0032	NA	0.48
*grl*A_SC(80:S-Y);*gyr*A_SC(84:S-F)	4% (7/162)	0% (0/17)	0% (0/17)	0% (0/53)	0% (0/10)	0.12	1	NA
*grl*A_SC(84:D-N or D-Y);*gyr*A_SC(84:S-F)	0% (0/162)	11% (2/17)	59% (10/17)	30% (16/53)	100% (10/10)	$9.05.10^{-14}$	$2.2.10^{-16}$	$7.18.10^{-07}$
other mutations *grl*A/*gyr*A	0% (0/162)	6% (1/17)	0% (0/17)	2% (1/53)	0% (0/10)	0.23	NA	1
23S mutation(2603:G-T)	0% (0/162)	0% (0/17)	0% (0/17)	2% (1/53)	10% (1/10)	0.76	0.06	0.42
*erm*A	0% (0/162)	0% (0/17)	100% (17/17)	21% (11/53)	0% (0/10)	$9.05.10^{-14}$	NA	0.17
*erm*C	7% (11/162)	53% (9/17)	0% (0/17)	26% (14/53)	80% (8/10)	$3.94.10^{-05}$	$2.98.10^{-11}$	0.00011
*msr*A	2% (4/162)	6% (1/17)	0% (0/17)	6% (3/53)	0% (0/10)	0.59	1	0.92
*tet*K/*tet*M genes	2% (4/162)	0% (0/17)	6% (1/17)	8% (4/53)	0% (0/10)	0.33	1	0.74
*fus*B gene	33% (53/162)	12% (2/17)	41% (7/17)	42% (22/53)	10% (1/10)	0.75	0.25	0.06
*mup*A gene	1% (1/162)	6% (1/17)	18% (3/17)	6% (3/53)	0% (0/10)	0.005	1	0.92
*fos*B gene	1% (1/162)	0% (0/17)	6% (1/17)	0% (0/52)	0% (0/10)	1	1	NA

*Presented data are p-values. Comparison was performed using two-sided Fisher exact test.

Extended Data Fig. 6 | Phenotypic and genotypic resistance patterns of *S. capitis* isolates. Phenotypic data of *S. capitis* isolates (n=250) were obtained from agar dilution and biomarkers of antibiotic resistance were detected using GENEFINDER. Comparison between groups of isolates was performed using two-sided Fisher exact test.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Phenotypic assays comparing a subset of representative isolates of each of the four subgroups identified by the phylogeographical analysis (Outbreak, Proto-outbreak 1, Proto-outbreak 2 and 'other isolates'). In all 6 graphs, center values represent means. **a**, Culture supernatants cytotoxicity assay using THP1 cells, adjusted on a positive control (Triton) of 12 representative *S. capitis* isolates (two independent experiments in triplicate for each strains). **b**, Survival of strains (n=12) after 24 hours of persistence in desiccation conditions (two independent experiments in triplicate for each strains). **c**, Comparison of the doubling time of bacterial growth during the exponential phase in standard conditions (BHI) of 24 representative *S. capitis* isolates (three independent experiments in triplicate for each strains) and **d**, Under oxidative stress (ethanol-supplemented medium to a final concentration of 6.5%) (n=24 strains, three independent experiments in triplicate for each strains). **e**, Quantification of biofilm production of 24 representative *S. capitis* isolates using crystal violet method (expressed as optic densitometry at 590nm) (three independent experiments in triplicate for each strains). **f**, Phagocytosis index of monocytes and granulocytes from cord blood for a subset of 5 representative isolates of "Outbreak" and "Basal" isolates (four independent experiments). Of note, results of phagocytosis of neutrophils and activated neutrophils are not represented here because they were similar to those with granulocytes.

gene_name_CR01_ref	HMP for THD index	HMP for vancomycin MICs	Type of substitution	Substitution	strand	begin	end	name	function	COG id	COG class	COG class signification
Association with THD index												
CR01_v3_2167	0.000000075	-	very variable	-	-2	2266734	2268197	-	unknown	COG5632	M	Cell wall/membrane/envelope biogenesis
Association with vancomycin MICs												
CR01_v3_0017	-	0.0000000076	synonymous substitution	-	-3	14537	15706	norA	Quinolone resistance protein NorA	COG0477	G	Carbohydrate transport and metabolism
CR01_v3_0554	-	0.0000000027	synonymous substitution	-	-3	548900	549937	-	putative perfringolysin O regulator protein	COG1299	G	Carbohydrate transport and metabolism
CR01_v3_1269	-	0.0000000076	non synonymous substitution	G422E	3	1281969	1283468	-	conserved membrane protein of unknown function	COG3428	S	Function unknown
CR01_v3_1626	-	0.0000000076	non synonymous substitution	G88A	1	1677763	1678515	rsmE	methylase of U1498 in 16S rRNA	COG1385	S	Function unknown
Association with both THD index and vancomycin MICs												
CR01_v3_1903	0.00000089	0.00000025	synonymous substitution	-	-1	1981288	1983255	mutL	DNA mismatch repair factor	COG0323	L	Replication, recombination and repair
CR01_v3_0096	0.0000013	0.00000017	non synonymous substitution	E120D ; T122N	3	91080	91454	sarA	Transcriptional regulator SarA	COG1846	K	Transcription
CR01_v3_2420	0.0000015	0.000000069	non synonymous substitution	I71K	-1	2492851	2493741	-	putative lipoprotein	COG0419	L	Replication, recombination and repair
CR01_v3_1843	0.0000027	0.00000034	synonymous substitution	-	-1	1922560	1925265	citB/AcnA	aconitate hydratase (aconitase)	COG1048	C	Energy production and conversion
CR01_v3_0764	0.0000052	0.00000051	non synonymous substitution	L182Q ; H225P	3	785034	786371	-	Aminotransferase class-III	COG0160	E	Amino acid transport and metabolism
CR01_v3_1826	0.0000067	0.0000016	non synonymous substitution	I389L	-3	1901735	1903141	trpE	Anthranilate synthase component 1	COG0147	H	Coenzyme transport and metabolism

THD time-scale haplotypic density ; HMP harmonic mean of p-value ; MIC minimal inhibitory concentration ; COG cluster of orthologous groups

Extended Data Fig. 8 | Genes associated with vancomycin MIC and/or THD success index using DBGWAS. Here are represented genes with a -log10 (HMP) > 7.5 on either axis, and/or > 5 on both axes, thus considered significant.

nature research

Corresponding author(s): Dr Marine BUTIN
Pr Thierry WIRTH

Last updated by author(s): Jun 15, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The following software have been used to collect data, the references are cited in the manuscript: FastQC v 0.11.7 Kraken 2 Unicycler v0.4.5 QUAST v 4.6.3 Prokka v 1.13
Data analysis	The following software have been used to perform analyses, the references are cited in the manuscript: ClonalFrame ML SRST2 SCCmecFinder SPAdes v3.8.0 Bowtie2 CRISPRcasFinder TREE-PUZZLE 7.1 PHYML 3.412 FIGTREE software v1.4 ITOL v3 TEMPEST1.5 LMPERM Package BEAST v2.3.2 TRACER V1.6 SPLITSTREE 3 ClonalFrame 1.2 PCADAPT

PHYRE2 software online (http://www.sbg.bio.ic.ac.uk/phyre2)
UCSF Chimera
GeneFinder
R version 3.3.2
DBGWAS 0.5.4
R package HARMONICMEANP
MaGe
NeighborNet

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data sets supporting the results of this article are available from the Sequence Read Archive (SRA) under BioProject no PRJNA493527

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We studied genomes and phenotypical features of a collection of 250 S. capitis isolates from adults, children and neonates from 22 countries, including multiple representatives of the endemic NRCS-A clone, in order to retrace its spread and identify drivers of its specific success in NICUs.
Research sample	A collection of 250 Staphylococcus capitis strains has been built. This collection aimed to represent the diversity within S. capitis species among strains found in clinical samples.
Sampling strategy	Microbiological laboratories worldwide (thanks to International consortium for Staphylococcus capitis neonatal sepsis) were asked to send a sample of their own collection of clinical S. capitis isolates to the French National Reference Center for Staphylococci, Hospices Civils de Lyon, France. All strains collected were included in the dataset.
Data collection	Collection was obtained in the National Reference Center for Staphylococci, Hospices Civils de Lyon, France. DNA extractions were performed in England (Public HEalth England, London, UK). Genome sequencing was performed in Lille (Institut PAsteur, Lille, France). Sequence and dataset analysis were performed both in National Museum of Natural History, Paris, France and in National Reference Center for Staphylococci, Lyon, France. Phenotypical tests were performed in Public HEalth England, London, UK, in he National Reference Center for Staphylococci, Hospices Civils de Lyon, France, and in Institut de Biologie de la cellule (I2BC-UMR9198), Paris, France.
Timing and spatial scale	Collection step was completed in 2015 that is why the more recent strain was isolated in 2015. Moreover, the oldest strain available in all laboratories contacted for the study was isolated in 1994. The spatial scale of the collection was as large as possible but was limited by the absence of Staphylococcus identification at the species scale in some countries (especially African countries, some laboratories in USA, some Asian countries).
Data exclusions	After sequencing, several isolates were excluded from the analysis because of a misidentification of bacterial species (n=3) or because of a lack of sequence data quality (n=4). These exclusion criteria were pre-established to ensure high quality of the database.
Reproducibility	not relevant in our study
Randomization	not relevant in our study design
Blinding	not relevant in our study design

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging