bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

On-person adaptive evolution of *Staphylococcus aureus* during atopic dermatitis increases disease severity

Felix M. Key[1,2,+], Veda D. Khadka[1,2], Carolina Romo-González[3], Kimbria J. Blake[4], Liwen Deng[4], Tucker C. Lynn[1,2], Jean C. Lee[5], Isaac M. Chiu[4], Maria Teresa García-Romero[6,#], Tami D. Lieberman[1,2,7,8,#]

Bacteria acquire adaptive mutations during infections and within healthy microbiomes [1-4], but the potential of bacterial mutations to impact disease is not well understood. The inflamed skin of people with atopic dermatitis (AD) is heavily colonized with *Staphylococcus aureus,* an opportunistic pathogen associated with both asymptomatic colonization of nasal passages and invasive disease[5,6]. While host genetic risk is critical to AD initiation [7,8], *S. aureus* worsens disease severity by inducing skin damage[9]. Here, we longitudinally track *S. aureus* evolution on 25 children with AD over 9 months —sequencing the genomes of 1,330 colonies— and identify common adaptive *de novo* mutations that exacerbate skin disease *in vivo*. Novel *S. aureus* genotypes replace their ancestors across the body within months, with signatures of adaptive, rather than neutral, forces. Most strikingly, the capsule synthesis gene *capD* obtained four parallel mutations within one patient and is involved in mutational sweeps in multiple patients. Despite the known role of capsule in phagocytic evasion[10], we find that an acapsular Δ*capD* strain colonizes better and produces worse disease severity on mouse skin than its encapsulated parental strain. Moreover, re-analysis of publicly available *S. aureus* genomes from 276 people confirms that CapD truncations are significantly more common among strains isolated from AD patients relative to other contexts. Together, these results suggest that targeting capsule-negative strains may be a potential avenue for decreasing *S. aureus* skin colonization and highlight the importance of single-mutation resolution for characterizing microbe-disease associations.

1. Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA
2. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
3. Experimental Bacteriology Laboratory, National Institute for Pediatrics, Mexico City, Mexico
4. Department of Immunology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA
5. Division of Infectious Disease, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
6. Department of Dermatology, National Institute for Pediatrics, Mexico City, Mexico.
7. Broad Institute, Massachusetts Institute of Technology, Cambridge, MA, USA
8. Ragon Institute, Massachusetts Institute of Technology, Cambridge, MA, USA

+Present address: Max Planck Institute for Infection Biology, 10117 Berlin, Germany (FMK)
#Co-corresponding authors: mgarciar@pediatria.gob.mx, tami@mit.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Main

During colonization of human microbiomes, bacteria acquire adaptive mutations that enhance their ability to survive in the human environment, resist antibiotics, and outcompete other strains [1-3,11]. While these *de novo* mutations rise in frequency due to the survival advantage they provide to the bacteria, their emergence may impact host metabolism, immune homeostasis, or microbiome dynamics. Understanding the tempo and consequences of variations across bacterial genomes is of particular importance for complex inflammatory diseases like atopic dermatitis (AD) and inflammatory bowel diseases, for which the causative role of the microbiome has been hard to pin down [12,13]. While recent studies have identified bacterial strains associated with inflammatory states [14-16], classic metagenomic approaches do not provide the resolution to robustly identify individual mutations emerging in disease states. As a result, the potential impact of *de novo* microbiome mutations on complex diseases is poorly understood.

Atopic dermatitis (AD) is one such chronic inflammatory skin disease with strong microbial associations and a complex etiology. AD transiently affects up to 20% of people during their lifetime [17] and is particularly prominent among children, who develop itchy patches of inflamed skin, typically located on the cubital and popliteal fossae (inside of elbows and backs of knees) [18]. Genetic and environmental defects in barrier function have been associated with AD, but are insufficient to explain the variation in disease development and response to treatment [19]. Notably, symptomatic AD skin of children and adults is usually colonized by the opportunistic pathogen *Staphylococcus aureus,* with abundance proportional to disease severity [6,9,20,21]; this species is otherwise rarely found on healthy skin [22]. Its native reservoir is thought to be the nares, where it asymptomatically colonizes 30% of healthy individuals [22]. However, *S. aureus* also causes a variety of human infections of the skin, bloodstream, lung, and bone. *S. aureus* strains vary in virulence potential, in the antibiotic resistance cassettes they carry, and the disease contexts in which they are found [5,23]. No strains have been robustly associated with AD, though a recent study has shown an essentiality of a key quorum sensing pathway during the formation of AD in young children [24]. While *de novo* mutations occurring in *S. aureus* in young children with AD have been observed [25], the fate of these mutations over time and their consequences have not been characterized. Here, we use longitudinal sampling and culture-based whole-genome sequencing to identify mutations acquired by *S. aureus* that emerge under natural selection on individual people and which are capable of directly impacting disease severity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

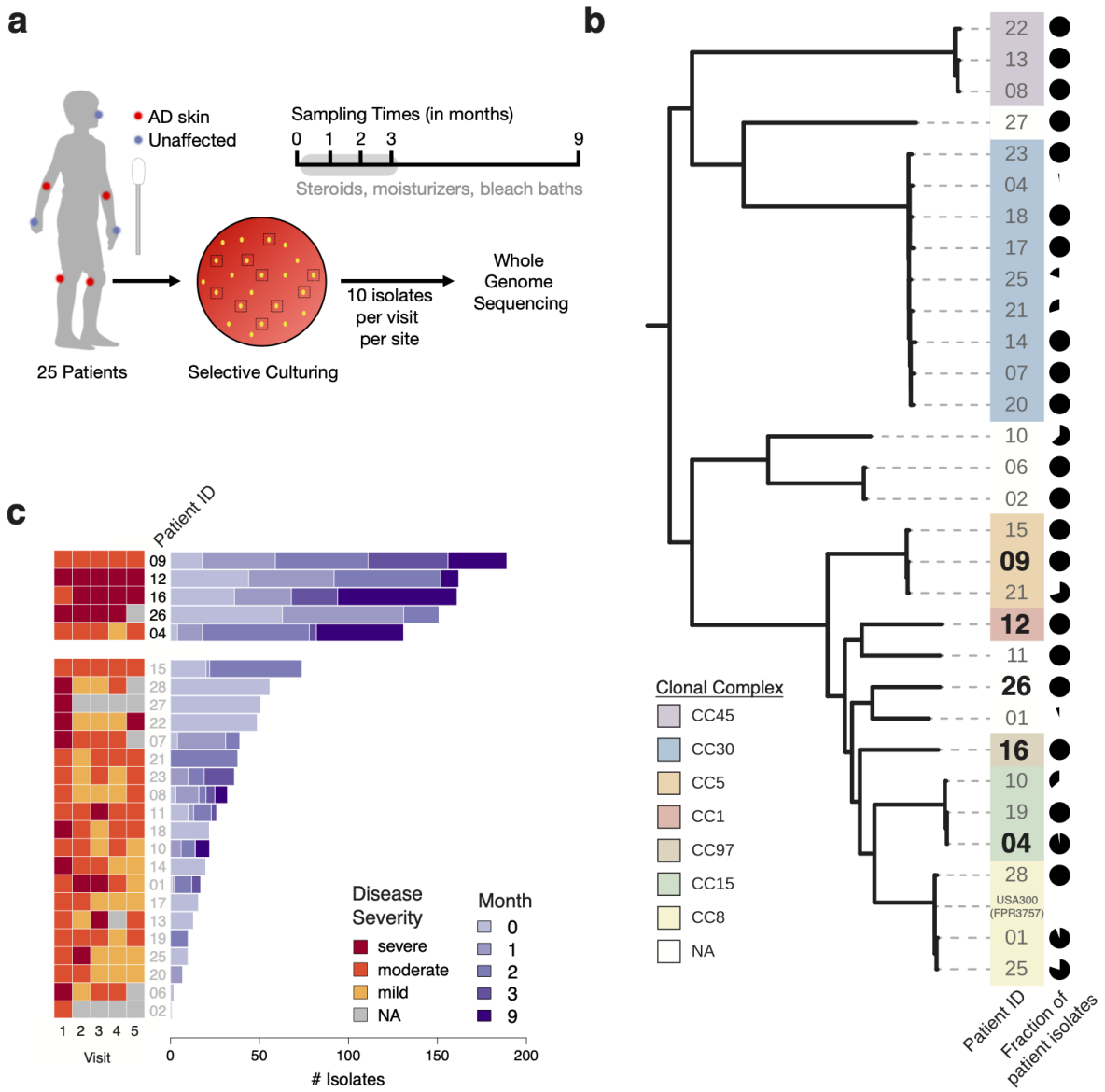


Figure 1: The *S. aureus* population on each child with atopic dermatitis (AD) is dominated by a single, stable, patient-specific lineage. a, Schematic of the longitudinal study of *S. aureus* evolution on children with AD. Swab samples were collected from each of 7 sites during each of 5 visits, including from nares and areas rarely affected by AD in children (forearms). From each swab, up to 10 colonies were cultured and processed for whole-genome sequencing. **b,** Phylogenetic tree showing relationship between all *S. aureus* isolates sequenced, generated using a reference-based approach and maximum-likelihood reconstruction with 91,783 SNVs (Methods). Isolates separated by fewer than 64 mutations are grouped into lineages and labeled by patient of origin. Pie charts indicate the fraction of patient isolates that come from each lineage, showing most patients have a single or dominant lineage. Lineages are colored according to their assigned clonal complex [26] (Methods), highlighting that strains colonizing AD patients come from a wide variety of global genotypes. **c,** For each patient, the disease severity over the course of 5 visits, as assessed by SCORAD, is plotted (left) alongside the number of isolates collected at each visit (right). We focus our analyses largely on the 5 patients with >100 isolates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Longitudinal tracking of *S. aureus* evolution in AD

We conducted a prospective, longitudinal study of 25 children (age 5 to 14 years old) with moderate to severe AD at 5 visits over the course of 9 months (**Figure 1a**). During each visit, AD severity was assessed using the SCOring Atopic Dermatitis (SCORAD) scale [27], and swabs were collected from seven affected and unaffected skin sites, including cubital and popliteal fossae, forearms and the nares (**Figure 1a**). From 205 swabs, we cultured and sequenced the genomes of 1,330 *S. aureus* colony isolates (up to 10 per swab). Isolates clustered into 30 patient-specific lineages separated by < 64 mutations (**Extended Data Figure 1**). Most patients were stably colonized by a single lineage, though we recovered a single minority clone from five patients (**Figure 1b; Supplementary Table 1**).

The *S. aureus* lineages recovered from patients span the diversity of the *S. aureus* species [28]. The largest fraction of lineages are part of clonal complex 30 (28% of lineages; 12% of isolates), in contrast with a previous finding of clonal complex 1 dominance among people with AD in the UK [25].

The number of isolates analyzed per patient visit is variable and correlates with disease severity (0-68 isolates/visit; r^2=0.36, **Figure 1b** and **Extended Data Figure 2**). While most patients provided isolates from intermittent or single timepoints, five patients suffering from severe AD throughout the study provided large numbers of isolates (131-189 isolates/patient; **Figure 1c**). We therefore focused most of our analyses on these patients. We generated patient-specific *S. aureus* pangenomes for all patients and used a rigorous alignment-based approach to identify single-nucleotide variants that occurred on each patient (**Figure 2a**, Methods).

S. aureus mutants sweep across the body

To investigate the spatiotemporal spread of *S. aureus*, we first estimated the molecular clock of these patients´ lineages. In four of these patients´ lineages, we observed a fast accumulation of mutations, with molecular rate estimates of 2.8 - 4.3 × 10^{-6} substitutions/site/year (corresponding to ~10 mutations/genome/year, **Figure 2b**), which largely agree with published rates for *S. aureus* (1.2 - 3.3 × 10^{-6} substitutions/site/year) [29-33]. One patient´s lineage had a significantly higher mutation rate of 18 × 10^{-6} substitutions/site/year (CI95 = 13 - 22 × 10^{-6} substitutions/site/year) (Patient 26; **Figure 2b**), suggesting a defect in DNA repair not apparent from our SNV-based analysis.

Accumulation of mutations in isolates can produce two different patterns of on-person evolution: diversification into coexisting genotypes or genotypic replacement. In all five heavily-colonized patients, phylogenetic reconstruction revealed a common pattern in which genotypes with newly acquired mutations repeatedly replace existing diversity (**Figure 2c-g**). These replacements were swift and spanned the entire body; the spread of new mutations was not spatially restricted and included the natural habitat of the nose (**Figure 2h, Extended Data Figure 2**). Similar genotypic sweeping likely occurred on most or all patients prior to the start of our study. For all 17 patients with at least 10 isolates at a visit, we found low diversity at the earliest such visit, indicating a recent reseeding of the population from a single genotype (**Figure 2i**). The frequency at which genotypes replace their ancestors is particularly striking

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

given the stability of on-patient *S. aureus* populations when considered at the lineage level (**Supplementary Table 1**).

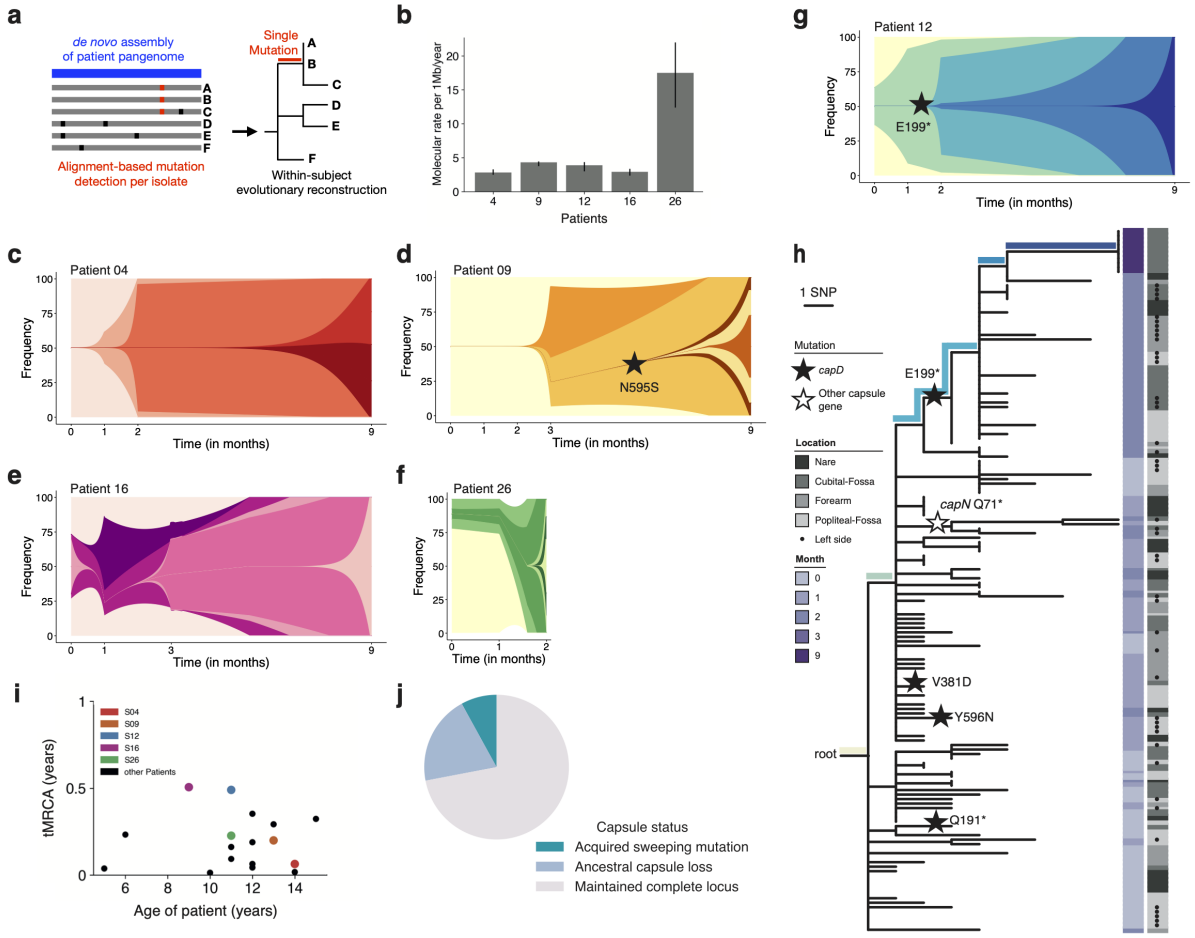


Figure 2. Rapid clonal spread of *de novo* mutations across the body implicates capsule loss as adaptive on AD skin. a, Schematic of evolutionary reconstruction within each lineage. We built a *de novo* reference genome for each patient (blue) and called *de novo* mutations by aligning raw reads from each sample individually to this reference genome, allowing the building of fine-resolution phylogenetic trees (Methods). **b,** The median molecular rate (Theil-Sen estimator) per patient in number of mutations per million base pairs. Error bars show the 95% confidence interval. **c, d, e, f, g** A schematic representing the on-person evolution for all patients with more than 100 isolates. Each color represents a unique genotype found in at least 3 isolates and with a frequency rise of ≥ 30%. The shape of lines between timepoints with data is speculative and modeled based on exponential growth (Methods). A star indicates a mutation in *capD*. **h,** Maximum parsimony tree of evolution of patient 12. The location from which each isolate was taken is indicated in gray (dots indicate left side of body) and timepoint is indicated by the purple heatmap. Stars indicate a mutation within the capsule locus. Branches representing mutations shown in (g) are colored accordingly. **i,** For each patient, the inferred time elapsed since the population most recently had a single-celled ancestor (Methods), using the earliest timepoint for which at least 10 isolates were available, is plotted relative to the patient's age. Mutational distances were converted to time using the calculated molecular clocks for the 5 highlighted patients; for all others, the median rate was used. This indicates that single-celled bottlenecks are common across all patients in our study. **j,** Pie chart indicating the capsule status for all 25 patients. Sweeping capsule mutations were observed in 2 patients' *S. aureus* populations, and another 5 patient populations were determined by *in vitro* phenotyping to have lost the capsule prior to the start of our study.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Adaptive mutations alter polysaccharide capsule

The speed of genotype replacement raises the possibility that the underlying mutations provide a competitive advantage on AD skin. To test if this process was adaptive, or arose from a neutral bottleneck (e.g. growth following a reduction in cell number during treatment), we investigated mutated genes for adaptive signatures. We first searched for evidence of parallel evolution, i.e. multiple mutations in the same gene within a single patient [34]. In patient 12, we observed four different mutations in *capD,* each in separate isolates. This mutational density is unlikely given the small number of mutations in this patient ($P = 4.4 \times 10^{-5}$, simulation), and all four mutations resulted in either amino acid replacement or a premature stop codon further supporting an adaptive role for mutations in this gene (**Figure 2h**). The *capD* gene encodes an enzyme that performs the first step in synthesizing the capsular polysaccharide of *S. aureus*. One of the *capD* mutations in patient 12, premature stop codon E199*, is associated with a spread and replacement event in this patient (1 of 9 mutations associated with replacement in this patient). Strikingly, a spread and replacement event in patient 9 is also associated with a mutation in *capD*: a nonsynonymous N595S mutation (**Extended Data Figure 3b**). These observations of parallel evolution within a patient, parallel evolution across patients, and association with replacement events strongly suggest that alterations in *capD* provide a survival advantage for *S. aureus* on AD skin.

We next sought to understand the generality of selection for *capD* alterations in our cohort. It is possible that replacement events involving *capD* emerged prior to the start of our study, given our inference of recent replacement events (**Figure 2i**). Alternatively, a patient's *S. aureus* population may have been initially founded by a strain with a nonfunctional *capD*. We tested all patient isolates for a complete *capD* open-reading frame and observed that all isolates from five additional patients had truncation mutations in this gene. Three of these patients share a frameshift mutation (single base insertion at an adenine hexamer of *capD*) also found in all isolates from the methicillin-resistant USA300 epidemic clone [35], indicating that this mutation was carried on their founding strains. To confirm capsule loss, we performed immunoblots and found that isolates from 24% of patients produce no detectable polysaccharide capsule (**Figure 2j, Supplementary Table 2**).

We did not detect comparably strong signals of adaptive *de novo* mutation across patients in any other gene (**Supplementary Table 3**). Notably, we did find a signature of adaptive maintenance of mobile elements (**Extended Data Figure 4**). While isolates from the first visit of several patients consisted of a mix of those with and without mobile elements (including prophages and plasmids; **Supplementary Table 4**), all mobile elements detected at a patient's first visit were found in all isolates at their last visit sampled. No novel mobile elements were gained in these five patients over the course of the study.

S. aureus capsule loss worsens disease on mouse skin

The polysaccharide capsule of *S. aureus* has been well studied, and it is generally considered a virulence factor that shields the pathogen from phagocytosis [10,36]. However, the loss of the *S. aureus* capsule has

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

been observed previously [35], including in USA300, and it has been shown that capsule-negative *S. aureus* exhibits improved adherence to tissues, including endothelial tissue [37].

The adaptive emergence of *capD* truncation mutations on AD skin, despite the known benefits of *S. aureus* capsule, suggests an advantage for the acapsular phenotype on AD skin or skin in general. To test this, we leveraged a well-established mouse model of AD based on epicutaneous *S. aureus* infection [38,39] (Methods). Consistent with our observations in AD patients, a Δ*capD*, acapsular, strain caused elevated disease severity scores (evaluated by erythema, edema, skin scale, and skin thickness, **Figure 3a**, $P = 2 \times 10^{-5}$; Wilcoxon test). There was also a higher bacterial burden recovered from mouse skin infected with the *capD* mutant (**Figure 3b**, n=12, P = .005 (Wilcoxon test) for replicate 1, Methods). Histological analysis revealed increased desquamation of the epidermis (skin scaling) and inflammatory infiltrates in skin from mice inoculated with the *capD* mutant compared to mice treated with the capsule producing strain (**Figure 3c**). These results support the hypothesis that *capD* mutants rise in frequency in AD patients because of their fitness advantage on skin and that the increased bacterial burden causes worse disease.

While the basis of this advantage remains unknown, previous studies suggest a possible mechanism: acapsular strains adhere better to certain cell types because the polysaccharide capsule interferes with the action of surface adhesin proteins [37,40]. This increased adherence is believed to be the basis for the success of acapsular phenotypes in endothelial[37] and mastitis models [41]. Adherence is particularly important on the skin surface, which may explain why contrasting results have been seen in a subcutaneous abscess model [42]. Alternatively, capsule negative strains might benefit from avoiding the metabolic costs of capsule production, which are likely to be enhanced on the nutrient-poor environment of skin.

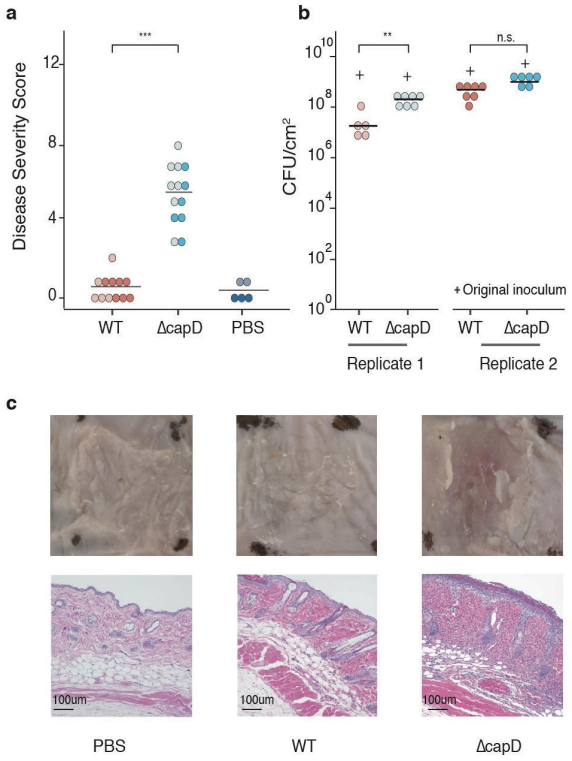


Figure 3. *S. aureus* strain with truncated CapD reaches higher bacterial burden and causes more severe disease in a murine epicutaneous infection model. a, An acapsular strain (Newman Δ*capD*) or capsule-producing strain (Newman WT) of *S. aureus* was topically applied to $1cm^2$ of mouse flank skin (n=13 per experimental group, two replicates; Methods). Disease severity scores were assessed on day 7 after association, measuring erythema (0-3), edema (0-3), scale (0-3), thickness (0-3). **b,** Bacterial load recovered from $1cm^2$ of mouse flank skin. Cell densities of inocula (1.3×10^9 CFU, 2.2×10^9 CFU for WT and 1.5×10^9 CFU, 4.2×10^9 CFU for Δ*capD*), are as indicated by black crosses. Significance values represented are: n.s. = $p > 0.05$, ** = $P < 0.01$, *** = $P < 0.001$ (Wilcoxon test) **c,** Representative images of mouse flank skin on day 7. Skin sections (below) were stained with hematoxylin and eosin.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Capsule loss is common in AD globally

To better understand if loss of *capD* is specifically advantageous on AD skin or is generally beneficial for *S. aureus in vivo*, we leveraged publicly available genomes of isolates from people with and without AD. We analyzed 276 *S. aureus* isolate whole-genomes from 110 AD patients, 67 healthy carriers, and 99 individuals who had other infections (blood-stream, soft tissue, bones, and joint; **Figure 4a**) [25,43-45]. While patients in our study were residents of Mexico, these samples were collected from individuals in Denmark, Ireland, and the UK.

Strikingly similar to our observation that 24% of AD patients carry a strain encoding a truncated CapD, 22.5% of AD-associated isolates in the public data lack a full-length *capD* gene This represents a significant increase relative to isolates from healthy controls (7.2%, $P < 0.001$; two-sided Fisher's exact test) and from those with other types of infections (10.5%, $P = 0.028$; two-sided Fisher's exact test) (**Figure 4b**). Phylogenetic analysis confirms many independent emergences of capD truncations, supporting the notion that *de novo* loss of *capD* can drive *S. aureus* adaptation (**Figure 4c**). Notably, 78% of capD-truncation isolates come from two independent, globally successful clones with recent expansions, including USA300 (CC8) and a CC1 lineage with a large deletion spanning *capD-capH.*

To understand if capsule loss via mutations in other genes is also enriched in AD, we repeated the same analysis for genes known to be associated with capsule regulation [46,47](**Figure 4d**). Interestingly, *agrC*, the histidine-kinase sensor in the *agr* quorum sensing system of *S. aureus,* also showed a significant enrichment of independent loss-of-function mutations in AD compared to other infections ($P < 0.001$; two-sided Fisher's exact test) but not isolates from healthy patients (**Extended Data Figure 5**). While recent work has suggested that retention of functional *agr* is associated with the onset of AD [24], *agrA* loss has been previously documented on the skin of a patient with AD [25]. Loss of the *agr* system is known to suppress capsule production, among several other pathways [46,48]. This finding of frequent *agr* loss in AD is consistent with an advantage for an acapsular phenotype on AD skin.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

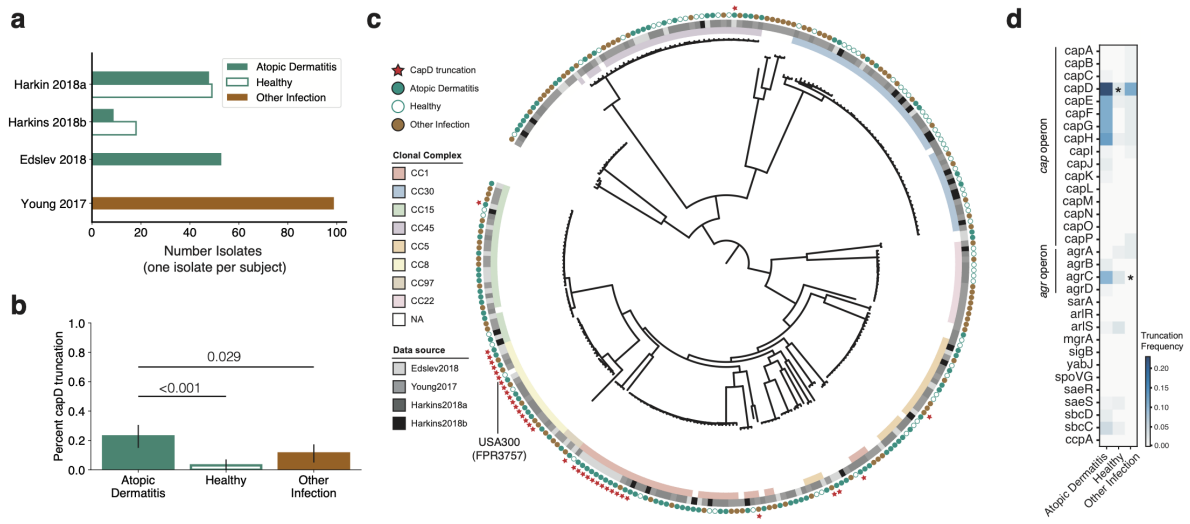


Figure 4. Meta-analysis of published data confirms enrichment of capsule loss on AD skin. a, The number of *S. aureus* genomes downloaded from each of 4 public studies are shown [25,43-45]. Isolates were collected either from people with AD (green), healthy individuals (white), or from a variety of other diseases, including bloodstream, bone/joint and soft-tissue infections (brown). **b,** The fraction of isolates with a truncated or missing *capD* is plotted as a function of isolation context (Methods). **c,** Phylogenetic tree showing relationship between all *S. aureus* analyzed, generated using a reference-based approach and maximum-likelihood reconstruction (Methods). Isolates are labeled with squares indicating their membership in global lineages and the study of origin, with circles indicating isolation context. Red stars indicate isolates without a full length *capD,* showing 9 different independent occurrences including two expanded clades. **d,** Analysis similar to that in (b) was performed for other genes involved in capsule regulation, and the percentage of isolates from each context lacking a full length copy of each gene is shown as a heatmap. * $P < 0.001$ vs. AD (two-sided Fisher's exact test).

Discussion

S. aureus is among the most successful opportunistic pathogens of humans, colonizing a third of the world's population, responsible for numerous outbreaks in healthcare facilities, and causing a variety of acute and chronic diseases[5]. Here, we report that loss-of-function mutations in a gene closely associated with virulence, *capD,* provide a competitive advantage to *S. aureus* in a particular microbiome niche: inflamed skin.

Loss of function mutations in *capD* are present in global lineages, and new bacterial strains with mutations in this gene also arise on individual patients with AD, spread across a patient's skin, and replace their ancestors within months. At a global scale, the advantage of being acapsular on skin may in part explain the success of the acapsular *S. aureus* USA300 lineage in causing skin infections and spreading to epidemic levels outside the hospital[35]; future studies are needed to test whether capsular mutants have an increased capacity for inter-person spread. The frequent emergence of capsule loss through premature stop codons in *capD* is particularly notable given *S. aureus*'s capacity for complex regulation of capsule production[46]. We cautiously propose that therapies that specifically target *S. aureus* strains with loss of function in *capD* may present a new strategy for treating AD or preventing skin infections. Future work characterizing the mechanistic basis behind the CapD-negative advantage on skin

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

will be critical to the design of such therapies, as unexpected consequences may emerge from selection for the capsular phenotype, which is more virulent in other disease contexts [10,49].

Together, our results highlight the potential of *de novo* mutations for altering bacterial competitiveness and disease severity in microbiomes, highlight the power of mutation tracking for identifying new therapeutic directions, and suggest that whole-genome resolution may be required for predicting the impact of microbial strains on complex diseases.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Methods

Study cohort and sample collection

Patients were recruited from the Dermatology Clinic at the National Institute for Pediatrics in Mexico City under a protocol approved by the Institutional Review Boards of the NIP (042/2016) and Massachusetts Institute of Technology. Inclusion criteria for enrollment were: ages 5 to 18 years, diagnosis of AD according to modified Hanifin and Rajka criteria [50], SCORAD ≥ 25 at first visit, and absence of topical or systemic antibiotics for the past month. Overview of patients' metadata is in **Supplementary Table 5**. Skin swabs were collected from seven different locations at up to 5 visits (Figure 1A), placed in Liquid Amies transport media, and sent to the laboratory for culture. Swabs were directly inoculated on mannitol salt and blood agar plates and cultured for 24h at 37℃. For each culture plate, up to 10 colonies suspected to be *S. aureus* by colony morphology were selected. These colonies were restreaked on ¼ of a blood agar plate and cultured for 24h at 37℃ to obtain sufficient material for DNA extraction. DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Corporation) for Gram-positive bacteria.

Library construction and Illumina sequencing

Dual-barcoded DNA libraries were constructed using the plexWell library prep system (SeqWell) for most samples and a modified version of the Nextera protocol for a small subset[51]. Libraries were sequenced on the Illumina NextSeq 500 using paired end 75bp reads to average of 3.1M reads per isolate (75 bp paired end). Demultiplexed reads were trimmed and filtered using cutadapt v1.18 [52] and sickle-trim v1.13 [53] (pe -q 20 -l 50).

Across-patient phylogenetic analyses

Reads were aligned using bowtie2 v2.2.6 against USA300-FPR3757 (RefSeq NC_007793). Candidate single nucleotide variants were called using samtools (v1.5) mpileup (-q30 -x -s -O -d3000), bcftools call (-c), and bcftools view (-v snps -q .75) [54]. For each candidate variant, information for all reads aligning to that position (e.g. base call, quality, coverage), across all samples, were aggregated into a data structure for local filtering and analysis. Isolates were removed from analysis if they had a mean coverage 8 or below across variant positions (185 isolates of an initial 1,531). Sporadic isolates that were phylogenetically confined within the diversity of another patient were likely mislabeled and removed (7 isolates). Isolates with a mean minor allele frequency above 0.04 across candidate sites (indicates contamination) were removed (5 isolates). We filtered candidate SNVs using a publicly available protocol (see Data Availability) similar to that previously published [34]. Basecalls were marked as ambiguous if the FQ score produced by samtools was below 30, the coverage per strand was below 2, or the major allele frequency was below 0.85. Any isolates for which >10% of candidate positions were marked as ambiguous at this stage were removed, presumably due to impurity of the starting colony (this removed 4 isolates, leaving 1,330 isolates). Remaining variant positions were filtered if 25% or more of all isolates were called as ambiguous or if the median coverage across strains was below 3, or if no unmasked polymorphisms remained. These filters retained 91,783 SNVs across 1,330 isolates. A maximum-likelihood tree was built with RAxML v8.2.12 [55] using the GTRCAT model, with rate heterogeneity disabled (-V). The phylogenetic tree was visualized with iTol [56].

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Within-patient phylogenetic reconstruction using de novo assemblies

We constructed a patient-specific pangenome in order to capture true genome-wide diversity; this approach can capture genetic information only found in a single isolate [3]. Filtered reads were taxonomically assigned using kraken2 (standard database build Sep 24 2018) [57], and only isolates with ≥ 80% reads assigned to *S. aureus* on the species level were included for *de-novo* assembly, leaving 1,271 isolates. For each lineage, we concatenated up to 250,000 reads from each member isolate and assembled a reference genome with SPAdes [58] (v.3.13, careful mode). All contigs of size 500b or larger were annotated using prokka [59] (v.1.14.6), which was supplied with a list of proteins (*--proteins*) from nine publicly available *S. aureus* genomes and six accompanying plasmids (NCBI accession: NC_002745, NC_003140, NC_002758, NC_002774, NC_009641, NC_003923, NC_002952, NC_002953, NC_005951, NC_002951, NC_006629, NC_007795, NC_007793, NC_007790, NC_007791, NC_007792). The assembled pangenomes for each patient are summarized in **Supplementary Table 6**.

We aligned the data of all initial 1,531 isolates to their respective patient-specific pangenomes using bowtie2 [60]. Isolates were removed if phylogenetically identified as minor lineages (25) or cross contaminants (30). Isolates and candidate single nucleotide variants were processed similar to the across-patient variant calling, but without an isolate filter based on minor allele-frequency. To remove variants that emerged from recombination or other complex events, we identified SNVs that were less then 500b apart and covaried highly across isolates within a patient (a SNV was considered to highly covary with another SNV if its covariance was in the 98% percentile across covariances calculated with the focal SNV); these positions were removed from downstream analysis. Across all patients we analyzed in total 1,327 isolates with 879 *de novo* on-person SNVs (**Supplementary Table 7**). Phylogenetic reconstruction was done using *dnapars* from PHYLIP v3.69 [61]. Trees were rooted using the isolate with the highest coverage from the most closely related lineage (based on across-patients analysis) as an outgroup.

Signatures of within-person adaptive evolution

Each within-patient dataset was searched for genes with either of two signatures of adaptive evolution: parallel evolution at the gene level [2] and hard selective sweeps [4,62,63]. First, we identified cases when two or more mutations arose in a single gene within a patient, with a minimum mutation density of 1 mutation per 1000 bp. We calculated a p-value for enrichment of mutations in each gene using a Poisson distribution. Only the *capD* gene in Patient 12 had a significant p-value after Bonferroni correction for the number of genes on the genome (**Supplementary Table 3**). Second, we searched for genomic positions at which the mutant allele frequency rose by at least 30% between visits. For each SNV, we assigned the ancestral allele as the allele found in the patient-specific outgroup, or, if that was not available, we used the allele present in the patient-specific pangenome. We compared lists of genes with candidate adaptive signatures across patients using CD-HIT (v4.7, 95% identity)[64]. All candidate signatures of adaptive evolution are reported in **Supplementary Table 3**.

Creation of evolution cartoons (Muller plots)

We visualized the change in frequency of mutations observed on each patient with > 100 isolates using Mueller plots. Patient 04 yielded only two isolates at visit 3; these were omitted from visualization. We only drew mutations with an observed variation in frequency of ≥ 0.3 between time points. We converted

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

mutation frequencies into genotype trajectories using *lolipop* v0.8.0 (https://github.com/cdeitrick/Lolipop), SNVs that differed by less than 8% derived allele frequency across timepoints were grouped into a single genotype. To make successive sweeps more visible, we used a custom-made python function to generate intermediate genotype frequencies between sampling timepoints. In brief, we created 100 time units per month, assumed mutations swept sequentially, and applied an exponential growth or decline following the function of frequency $x = \mathcal{E}\text{^}((\ln(f_1) - \ln(f_0)) / g)$, with f_0 is the frequency at preceding visit or 0, f_1 is the frequency at the next visit and g is the number of time units. The extended table was used again as the input for *lolipop*, which provided the input tables for plotting using the R package *ggmueller* v0.5.5 [65].

Molecular clock and TMRCA

We estimated the number of *de novo* mutations per isolate using all positions variable within a patient lineage. The ancestral state of each variant is defined by the allele called in the patient-specific outgroup. If this was not available, the major allele for all non-patient outgroup isolates was used. We normalized the number of mutations per isolate by dividing by the number of positions on the reference with at least 8X depth. We inferred the molecular rate using the Theil-Sen estimator [66] implemented in scipy v1.3.1 [67]. Time to the most recent common ancestor (tMRCA) was calculated for each patient data from each patient's earliest visit with at least 10 isolates. For each of the five highly colonized patients the tMRCA was calculated using their inferred mutation rate, while for all other patients the median molecular rate of the five highly colonized individuals was used (3.87×10^{-6} substitutions/site/year).

Mobile genetic element analysis

We identified genetic gain and/or loss events based on the depth of coverage of each isolate aligned to its patient-specific pangenome along the five individuals with more than 100 isolates. To avoid spurious results due to uneven coverage across samples or genomic regions, we calculated a z-score for each position by normalizing each isolate's depth of coverage at each position by sample and by position. We identified genomic regions greater than or equal to 5,000 bp long, for which each position's z-score was below a threshold of -0.5 with a median threshold below -1.0. Identified candidate regions were further filtered to have at least one isolate with a median coverage of 0 and at least one with an average depth of at least 10X. Candidate regions were confirmed using a custom, semi-automated python module that allowed interactive validation of each candidate (**Supplementary Table 4**).

Capsule typing

S. aureus isolates from 19 patients were cultivated at 37°C for 24 h on Columbia agar (Difco Laboratories) supplemented with 2% NaCl and capsule extracts were prepared as previously described [68]. Extracts were stored at –20C until further characterization. Two microliters of capsule isolates, as well as control strains, were applied to nitrocellulose membranes in a grid fashion. Capsule typing was performed by a previously described immunoblot method [69], using both polyclonal and mAbs specific to *S. aureus* CP types 5 and 8. Four replica experiments were performed. Results are summarized in **Supplementary Table 2**.

Bacterial Strains and Cultures for *in vivo* experiments

S. aureus strains Newman (WT) and a *cap5D* C-terminal deletion mutant in the Newman background (**Δ**Cap5D) were used as capsule-positive and capsule-negative strains respectively. Bacteria were grown to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

stationary phase overnight at 37°C in Tryptic Soy Broth (TSB) at 250 r.p.m. For inoculation, stationary phase cultures were diluted 1:100 in fresh TSB and grown for 3.5 hours to approximate mid-log phase. Cultures were then centrifuged at 800xg for 15 minutes, and the resulting pellet washed twice in phosphate buffered saline (PBS). The cells were then resuspended in PBS to an OD of $A_{600} = 0.5$, corresponding to ~4 x 10^8 CFU/mL. Serial dilutions were plated on ChromAgar *S. aureus* (ChromAgar, France) plates to confirm inoculum cell densities.

Animals

Eight-week old C57BL/6 female mice from Jackson Laboratories (Bar Harbor, ME, USA) were housed in specific pathogen free animal barrier facilities at Harvard Medical School in individually ventilated micro isolator cages under a 12 h light/dark cycle with ad libitum food and water access. Euthanasia was performed by CO_2 inhalation. All animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) at Harvard Medical School.

Epicutaneous Skin Infection and Bacterial Load Measurements

Two days prior to epicutaneous infection, mice were shaved and the remaining hair removed using depilatory cream (Nair) along the length of their back/flank. A sterile 1cm^2 square of gauze was soaked in 100uL of prepared *S. aureus* inoculum (~10^9 CFU/mL for each strain, see **Figure 3b** for exact starting inocula) and applied to the flank skin. The gauze was secured using bio-occlusive film (Tegaderm, 3M) and the tape checked and repaired daily to maintain its integrity. Gauze soaked in sterile PBS was used as a control. Seven days post-inoculation, mice were euthanized by CO_2 and the dressing removed. Skin under the gauze was immediately scored for disease severity according to the following criteria: oedema (0-3), erythema (0-3), skin scale (0-3), and skin thickness (0-3) and totaled for a max skin score of "12.". A higher score indicated more severe disease/inflammation. For bacterial load, the flank skin immediately under the gauze (~1cm^2) was excised and resuspended in 1mL cold PBS contained in a 2mL microcentrifuge tube. The tissue was cut into smaller pieces using sterilized scissors, two metal BB beads (Daisy Outdoor Products) were added to each tube, and the tissue was homogenised using a TissueLyser II (Qiagen, Germany) at 25 s^{-1} for 5 min. Homogenates were briefly spun down, serially diluted in PBS, and then plated for CFU counts. Bacterial identity was confirmed by plating on ChromAgar *S. aureus* (ChromAgar, France) plates to differentiate any native microbiota that may have been present. A total of two replicates were performed.

Histology

Flank skin samples were dissected, post-fixed overnight in 4% paraformaldehyde, embedded in paraffin, sectioned, and stained using hematoxylin and eosin (H&E) by the Harvard Medical School Rodent Histopathology Core. Stained sections were imaged using a Keyence BZ light microscope.

Public *S. aureus* isolate sequencing data analysis

We investigated publicly available data to verify if loss of CapD is associated with *S. aureus* colonizing AD skin. We obtained whole-genome sequence data from 4 different publications analysing *S. aureus* in healthy individuals, AD patients or individuals with other *S. aureus* infections [25,43-45]. When data from multiple isolates per patient was available, we used only one isolate in order to assess prevalence of variants across conditions; the isolate with highest coverage was chosen. To understand the relationship between isolates, we performed an alignment-based phylogenetic reconstruction using *S.aureus* COL

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(sequence version number: NC_002951.2, NC_006629.2) using the same filters as above. To examine gene content, we performed *de novo* assembly for each isolate using SPAdes (v3.13, --careful) [58] and annotated the assembly using Prokka (v1.14.6) [59] as described above. We removed 6 isolates with an assembly length of < 2.6M nucleotides. In total, we analyzed 276 isolates from 276 patients. In addition, we included USA300-FPR3757 (RefSeq NC_007793) as a sample, by simulating raw reads for input to our cross-lineage phylogenetic analysis (cutting its genome in segments of size 150 b in steps of 1 b to simulate reads). The maximum likelihood phylogeny was built using RAxML v8.2.12 (-m GTRCAT) [55] and visualized with iTol [56].

We inferred the ORF status for all capsule and capsule-associated genes using the annotated assemblies and BLAST+ v2.7.1 [70]. Details about each query gene are available in **Supplementary Table 8**. We compared the best BLAST match for an overlap with the annotated reading frame. We accepted an ORF as complete if the start and end of the best BLAST hit were each within 100 bp of a gene start or end of a gene in the annotated assembly. *S. aureus* genomes colonizing humans are known to carry 1 of 2 predominant *cap* loci (type 5 or 8) [10] and 1 of 4 *agr* types [71] containing homologous versions of the same genes; for these cassettes we performed analysis only for the respective loci with the best BLAST match to a given isolate. Results are reported in **Figure 4**, **Extended Data Figure 5** and summarized in **Supplementary Table 9**. When only papers reporting both AD and controls are included, the enrichment of *capD* premature stop codons in AD vs controls remains significant (P=0.02).

Data and software availability

All raw data is available at SRA under BioProject identifier: PRJNA715649 and PRJNA715375. All code needed to reproduce the results of this study, including snakemake pipelines, are available here: https://github.com/keyfm/aureus_ad (under construction).

Author contributions

T.D.L and M.T.G-R. designed the clinical cohort. M.T.G-R. enrolled patients and collected clinical samples. C.R-G. cultured bacteria from clinical samples, extracted DNA, and obtained polysaccharide extracts. T.C.L. and T.D.L. prepared genomic libraries. F.M.K. and T.D.L performed all genomic analysis and interpretation. J.C.L. performed capsule assays and provided *S. aureus* strains. V.K., K.J.B, L.D., and I.M.C. performed mouse experiments. F.M.K. and T.D.L wrote the manuscript with feedback from all authors.

Acknowledgements

We thank Mariana Matus and Eric Alm for assistance in the design of the clinical cohort, the BioMicroCenter at MIT for performing genomic sequencing, Samantha Choi for technical support on animal experiments, and members of the Lieberman lab for valuable advice and feedback on the manuscript. We acknowledge support from MISTI Global Seed Funds (to T.D.L. and M.T.G-R.), the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

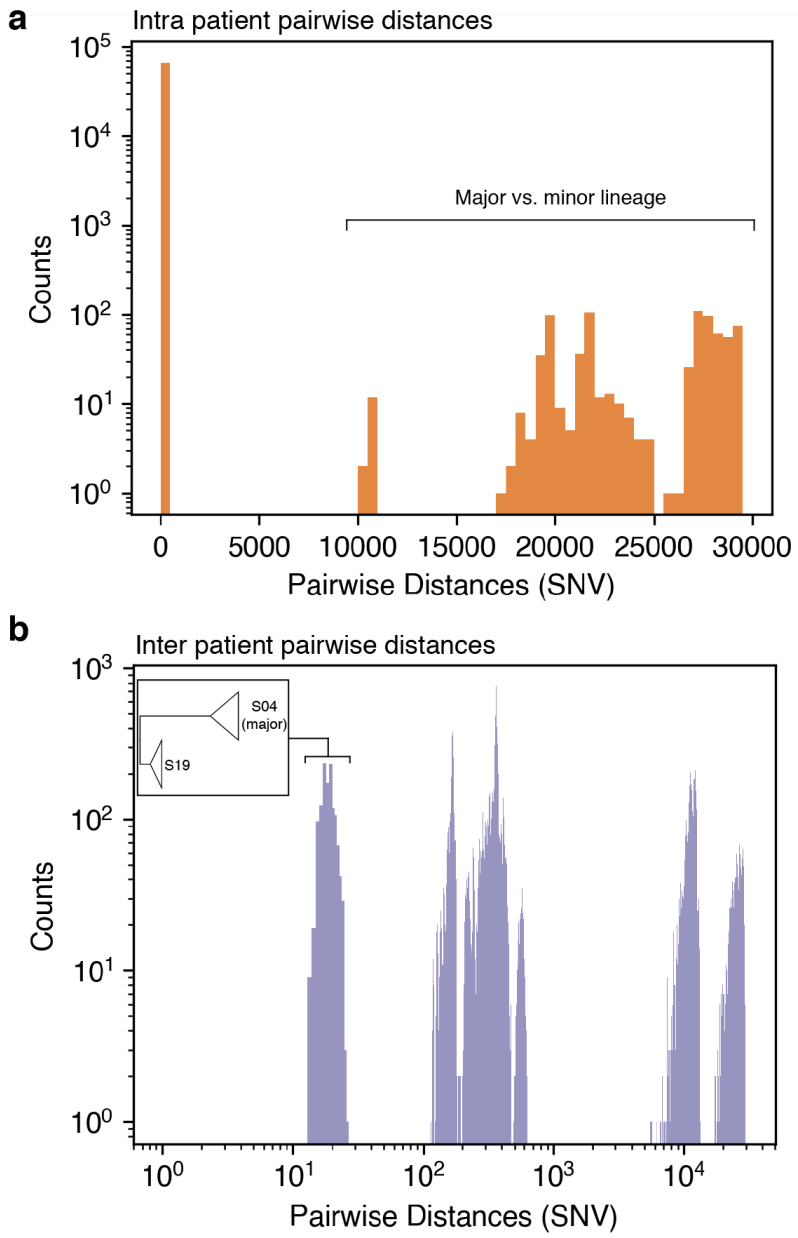
National Institutes of Health (DP2-GM140922 to T.D.L., R01AI30019 to I.M.C.), Burroughs Wellcome Fund (to I.M.C.), the Mexican Government Ministry of Taxes Program E022 for Health Research and Technological Development 2018 (to M.T.G-R.), and DFG research fellowship (KE 2408/1-1 to F.M.K.).

Competing interests

No competing interests to declare.

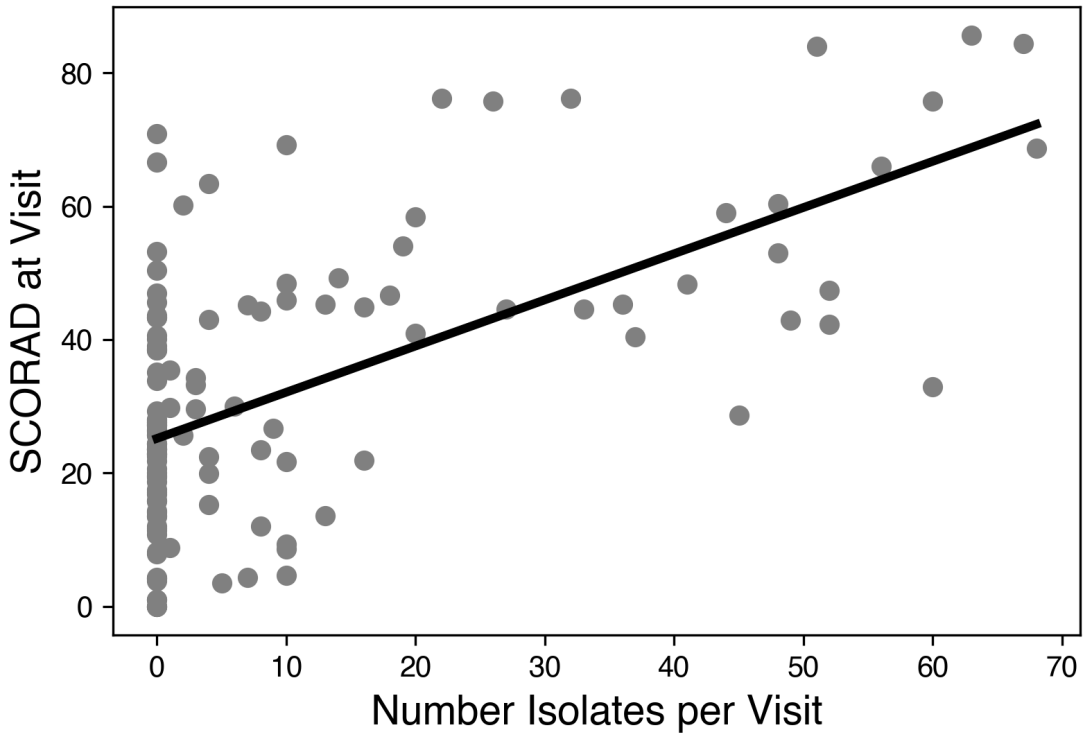
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Extended Data Figures



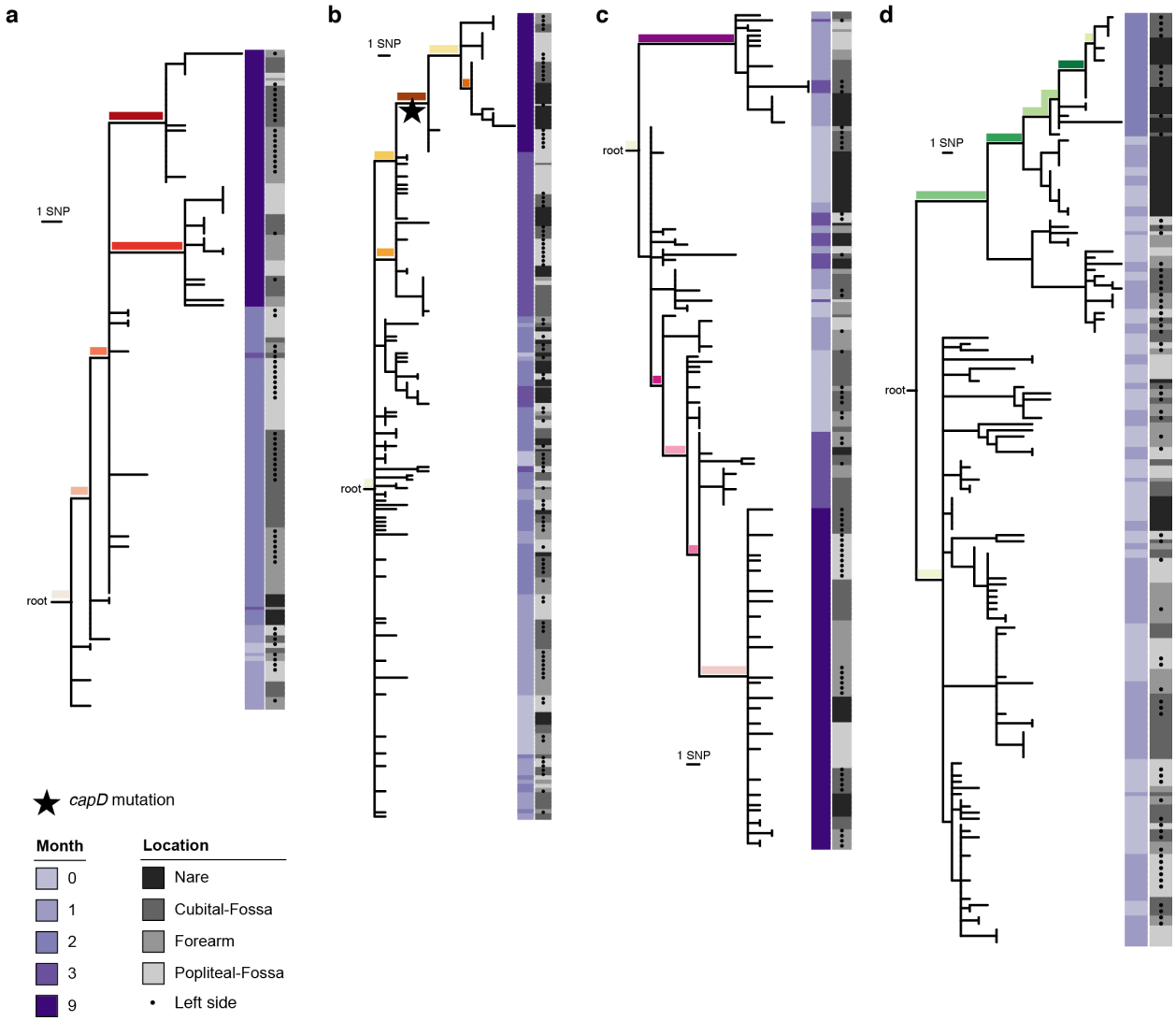
Extended Data Figure 1. Whole-genome pairwise differences between isolates of the same patient, or between patients. (**a**) Histogram of pairwise distances between isolates within a patient. Isolates from the same lineage have fewer than 70 SNV differences. The y -axis is shown on a log scale to make major-minor lineage pairs more visible. (**b**) Histogram of pairwise distances (SNV) between isolates of different patients. Patient 19 and the major lineage of Patient 04 are closely related but distinct (shown in insert). Isolates were included in this analysis only if at least 75% of all identified variable sites had coverage using an alignment-based approach (USA300-FPR3757). Both axes are on a log scale to enhance visibility.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



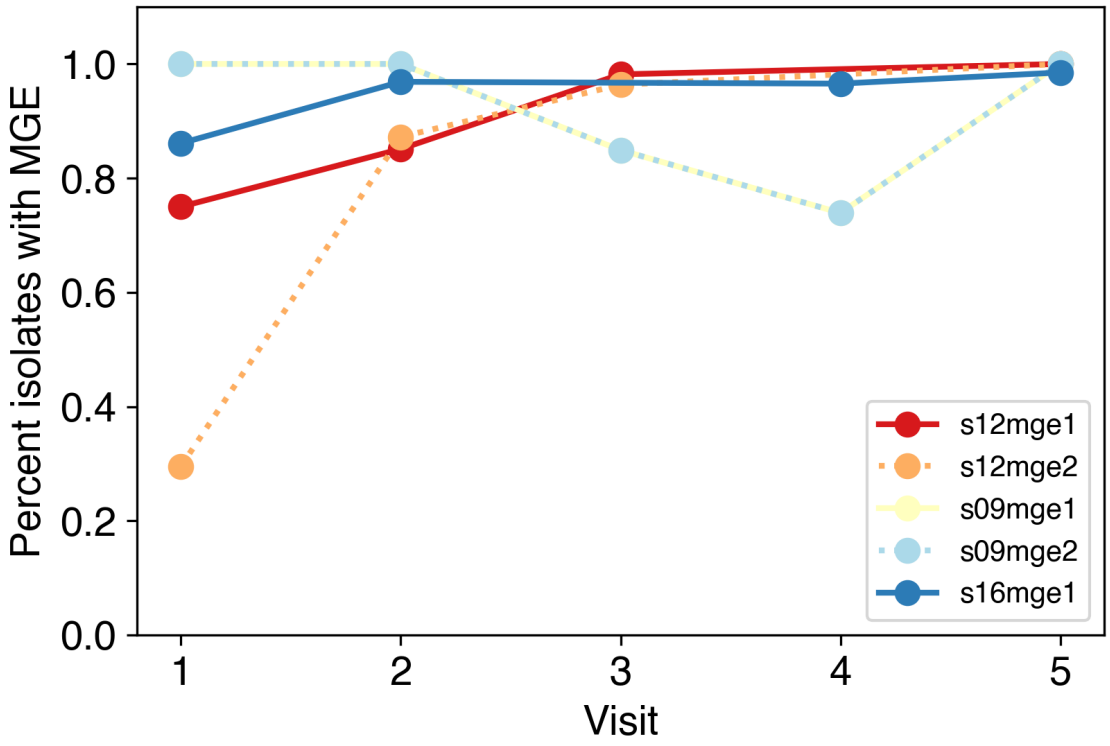
Extended Data Figure 2. Linear least-squares regression for the number of recovered isolates and disease severity (SCORAD) per visit (r^2=0.36). Regression analysis was done using stats.linregress() part of the scipy module in python [67].

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



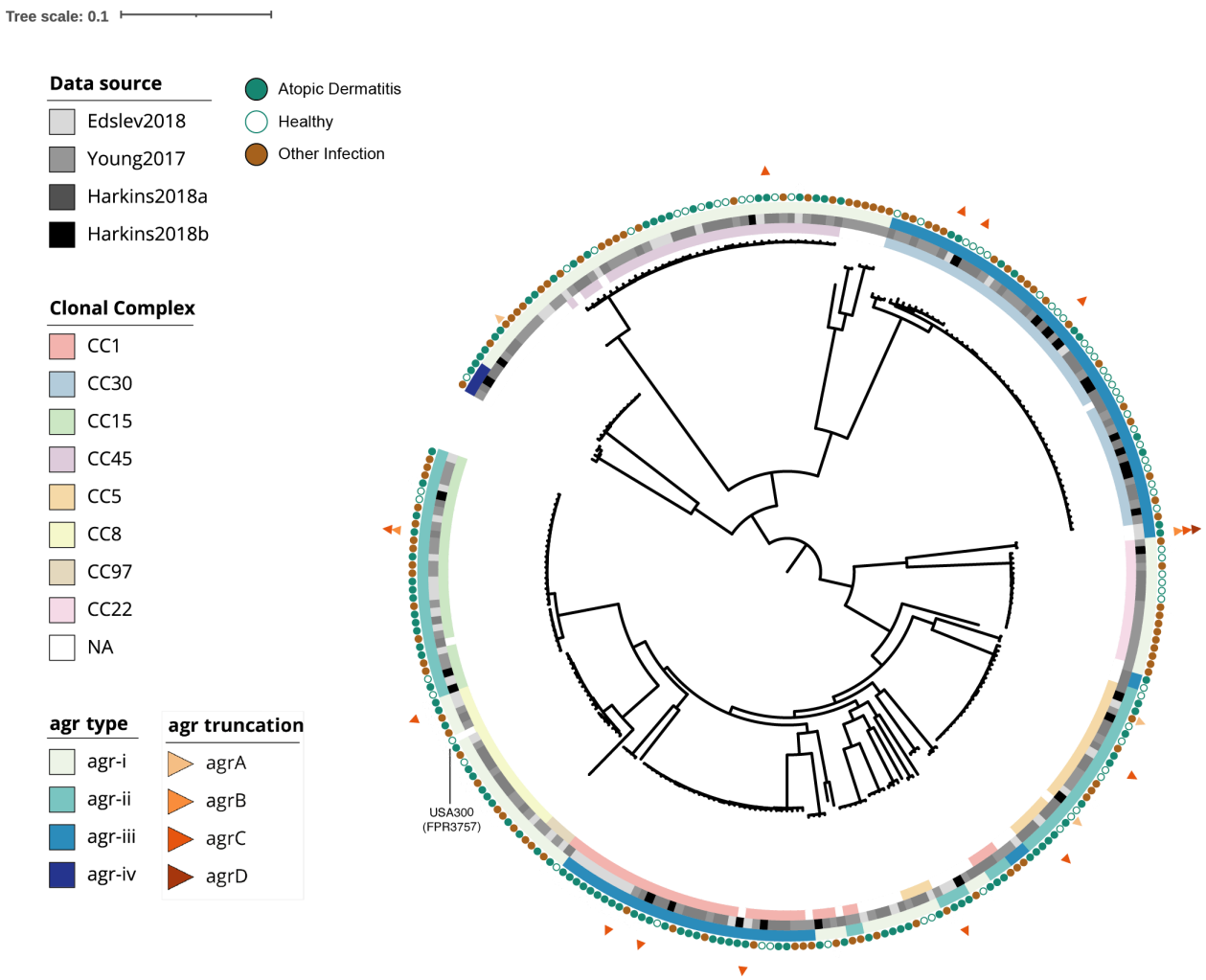
Extended Data Figure 3. Within-patient evolutionary reconstruction of *S. aureus* for other highly colonised patients. Within-person evolutionary trees for Patients 04 (**a**), 09 (**b**), 16 (**c**) and 26 (**d**), constructed similarly to **Fig. 2h.** The location from which each isolate was taken is indicated in gray (dots indicate left side of body) and timepoint is indicated with the purple heatmaps. Stars indicate the inferred location of mutations in the capsule locus. Branches with >30% change in frequency between timepoints are colored in accordance with **Fig**. **2**

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



Extended Data Figure 4. Frequency of identified mobile genetic element (MGE) among isolates per visit. Isolates were considered to harbor the MGE if the mean depth across the MGE was at least half of the mean genome-wide depth. To avoid artefacts, MGEs are reported only if at least two isolates had evidence of coverage and at least two isolates had evidence of absence. Gene annotations in each MGE are reported in **Supplementary Table 4**.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



Extended Data Figure 5. Maximum-likelihood phylogeny showing relationship for all 276 *S. aureus* isolates and strain FPR3757, using a reference-based approach (similar to **Figure 4c**). Isolates are labeled with squares indicating their membership in global lineages, the study of origin, and *agr* type, with circles indicating isolation context. Triangles indicate isolates with truncated *agr A*, *B*, *C*, or/and *D*, showing 16 different independent events.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1 Snitkin, E. S. *et al.* Genomic insights into the fate of colistin resistance and Acinetobacter baumannii during patient treatment. *Genome research* **23**, 1155-1162, doi:10.1101/gr.154328.112 (2013).

2 Lieberman, T. D. *et al.* Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet* **43**, 1275-1280, doi:10.1038/ng.997 (2011).

3 Zhao, S. *et al.* Adaptive evolution within gut microbiomes of healthy people. *Cell host & microbe* **25**, 656-667 (2019).

4 Garud, N. R., Good, B. H., Hallatschek, O. & Pollard, K. S. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLoS biology* **17**, e3000102 (2019).

5 Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. Staphylococcus aureus Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. *Clinical Microbiology Reviews* **28**, 603, doi:10.1128/CMR.00134-14 (2015).

6 Leyden, J. J., Marples, R. R. & Kligman, A. M. Staphylococcus aureus in the lesions of atopic dermatitis. *British Journal of Dermatology* **90**, 525-525, doi:https://doi.org/10.1111/j.1365-2133.1974.tb06447.x (1974).

7 Schultz Larsen, F., Holm, N. V. & Henningsen, K. Atopic dermatitis: A genetic-epidemiologic study in a population-based twin sample. *Journal of the American Academy of Dermatology* **15**, 487-494, doi:https://doi.org/10.1016/S0190-9622(86)70198-9 (1986).

8 Paternoster, L. *et al.* Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nature genetics* **47**, 1449 (2015).

9 Huang, J. T., Abrams, M., Tlougan, B., Rademaker, A. & Paller, A. S. Treatment of Staphylococcus aureus colonization in atopic dermatitis decreases disease severity. *Pediatrics* **123**, e808-e814 (2009).

10 O'Riordan, K. & Lee, J. C. Staphylococcus aureus capsular polysaccharides. *Clinical microbiology reviews* **17**, 218-234 (2004).

11 Ernst, C. M. *et al.* Adaptive evolution of virulence and persistence in carbapenem-resistant Klebsiella pneumoniae. *Nature medicine* **26**, 705-711 (2020).

12 Williams, M. R. & Gallo, R. L. The Role of the Skin Microbiome in Atopic Dermatitis. *Current Allergy and Asthma Reports* **15**, 65, doi:10.1007/s11882-015-0567-4 (2015).

13 Ni, J., Wu, G. D., Albenberg, L. & Tomov, V. T. Gut microbiota and IBD: causation or correlation? *Nature reviews Gastroenterology & hepatology* **14**, 573 (2017).

14 Nakatsuji, T. *et al.* Antimicrobials from human skin commensal bacteria protect against Staphylococcus aureus and are deficient in atopic dermatitis. *Science translational medicine* **9** (2017).

15 Byrd, A. L. *et al.* Staphylococcus aureus and Staphylococcus epidermidis strain diversity underlying pediatric atopic dermatitis. *Science translational medicine* **9** (2017).

16 Hall, A. B. *et al.* A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Medicine* **9**, 103, doi:10.1186/s13073-017-0490-5 (2017).

17 Asher, M. I. *et al.* Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *The Lancet* **368**, 733-743, doi:https://doi.org/10.1016/S0140-6736(06)69283-0 (2006).

18 Hay, R. J. *et al.* The Global Burden of Skin Disease in 2010: An Analysis of the Prevalence and Impact of Skin Conditions. *Journal of Investigative Dermatology* **134**, 1527-1534, doi:https://doi.org/10.1038/jid.2013.446 (2014).

19 Baurecht, H. *et al.* Epidermal lipid composition, barrier integrity, and eczematous inflammation are associated with skin microbiome configuration. *Journal of Allergy and Clinical Immunology* **141**, 1668-1676.e1616, doi:https://doi.org/10.1016/j.jaci.2018.01.019 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

20 Kong, H. H. *et al.* Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Research* **22**, 850-859 (2012).

21 Tauber, M. *et al.* Staphylococcus aureus density on lesional and nonlesional skin is strongly associated with disease severity in atopic dermatitis. *Journal of Allergy and Clinical Immunology* **137**, 1272-1274 (2016).

22 Gorwitz, R. J. *et al.* Changes in the Prevalence of Nasal Colonization with Staphylococcus aureus in the United States, 2001–2004. *The Journal of Infectious Diseases* **197**, 1226-1234, doi:10.1086/533494 (2008).

23 Enright, M. C. *et al.* The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). *Proceedings of the National Academy of Sciences* **99**, 7687-7692 (2002).

24 Nakamura, Y. *et al.* Staphylococcus Agr virulence is critical for epidermal colonization and associates with atopic dermatitis development. *Science Translational Medicine* **12**, eaay4068, doi:10.1126/scitranslmed.aay4068 (2020).

25 Harkins, C. P. *et al.* The Microevolution and Epidemiology of Staphylococcus aureus Colonization during Atopic Eczema Disease Flare. *Journal of Investigative Dermatology* **138**, 336-343, doi:https://doi.org/10.1016/j.jid.2017.09.023 (2018).

26 Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* **3**, 124, doi:10.12688/wellcomeopenres.14826.1 (2018).

27 Chopra, R. *et al.* Severity strata for Eczema Area and Severity Index (EASI), modified EASI, Scoring Atopic Dermatitis (SCORAD), objective SCORAD, Atopic Dermatitis Severity Index and body surface area in adolescents and adults with atopic dermatitis. *British Journal of Dermatology* **177**, 1316-1321 (2017).

28 Turner, N. A. *et al.* Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research. *Nature Reviews Microbiology* **17**, 203-218 (2019).

29 Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469-474, doi:10.1126/science.1182395 (2010).

30 Holden, M. T. *et al.* A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Res* **23**, 653-664, doi:10.1101/gr.147710.112 (2013).

31 McAdam, P. R. *et al.* Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant Staphylococcus aureus. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9107-9112, doi:10.1073/pnas.1202869109 (2012).

32 Nübel, U. *et al.* MRSA Transmission on a Neonatal Intensive Care Unit: Epidemiological and Genome-Based Phylogenetic Analyses. *PloS one* **8**, e54898, doi:10.1371/journal.pone.0054898 (2013).

33 Uhlemann, A. C. *et al.* Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. *Proc Natl Acad Sci U S A* **111**, 6738-6743, doi:10.1073/pnas.1401006111 (2014).

34 Lieberman, T. D. *et al.* Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat Genet* **46**, 82-87, doi:10.1038/ng.2848 (2014).

35 Boyle-Vavra, S. *et al.* USA300 and USA500 clonal lineages of Staphylococcus aureus do not produce a capsular polysaccharide due to conserved mutations in the cap5 locus. *mBio* **6** (2015).

36 Kuipers, A. *et al.* The Staphylococcus aureus polysaccharide capsule and Efb-dependent fibrinogen shield act in concert to protect against phagocytosis. *Microbiology* **162**, 1185 (2016).

37 Pöhlmann-Dietze, P. *et al.* Adherence of Staphylococcus aureus to endothelial cells: influence of capsular polysaccharide, global regulator agr, and bacterial growth phase. *Infection and immunity* **68**, 4865-4871, doi:10.1128/iai.68.9.4865-4871.2000 (2000).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

38 Nakamura, Y. *et al.* Staphylococcus δ-toxin induces allergic skin disease by activating mast cells. *Nature* **503**, 397-401 (2013).

39 Nakagawa, S. *et al.* Staphylococcus aureus Virulent PSMα Peptides Induce Keratinocyte Alarmin Release to Orchestrate IL-17-Dependent Skin Inflammation. *Cell Host & Microbe* **22**, 667-677.e665, doi:https://doi.org/10.1016/j.chom.2017.10.008 (2017).

40 Risley, A. L., Loughman, A., Cywes-Bentley, C., Foster, T. J. & Lee, J. C. Capsular Polysaccharide Masks Clumping Factor A-Mediated Adherence of Staphylococcus aureus to Fibrinogen and Platelets. *The Journal of Infectious Diseases* **196**, 919-927, doi:10.1086/520932 (2007).

41 Tuchscherr, L. P. *et al.* Capsule-negative Staphylococcus aureus induces chronic experimental mastitis in mice. *Infect Immun* **73**, 7932-7937, doi:10.1128/iai.73.12.7932-7937.2005 (2005).

42 Portolés, M., Kiser, K. B., Bhasin, N., Chan, K. H. N. & Lee, J. C. Staphylococcus aureus Cap5O has UDP-ManNAc dehydrogenase activity and is essential for capsule expression. *Infection and immunity* **69**, 917-923 (2001).

43 Young, B. C. *et al.* Severe infections emerge from commensal bacteria by adaptive evolution. *elife* **6**, e30637 (2017).

44 Edslev, S. M., Clausen, M.-L., Agner, T., Stegger, M. & Andersen, P. S. Genomic analysis reveals different mechanisms of fusidic acid resistance in Staphylococcus aureus from Danish atopic dermatitis patients. *Journal of Antimicrobial Chemotherapy* **73**, 856-861 (2018).

45 Harkins, C. P. *et al.* The widespread use of topical antimicrobials enriches for resistance in Staphylococcus aureus isolated from patients with atopic dermatitis. *British Journal of Dermatology* **179**, 951-958, doi:https://doi.org/10.1111/bjd.16722 (2018).

46 Lei, M. G. & Lee, C. Y. MgrA Activates Staphylococcal Capsule via SigA-Dependent Promoter. *Journal of Bacteriology* **203**, e00495-00420, doi:10.1128/JB.00495-20 (2020).

47 Keinhörster, D. *et al.* Revisiting the regulation of the capsular polysaccharide biosynthesis gene cluster in Staphylococcus aureus. *Molecular Microbiology* **112**, 1083-1099, doi:https://doi.org/10.1111/mmi.14347 (2019).

48 Luong, T., Sau, S., Gomez, M., Lee, J. C. & Lee, C. Y. Regulation of Staphylococcus aureus capsular polysaccharide expression by agr and sarA. *Infection and immunity* **70**, 444-450 (2002).

49 Suligoy, C. M. *et al.* Acapsular Staphylococcus aureus with a non-functional agr regains capsule expression after passage through the bloodstream in a bacteremia mouse model. *Scientific Reports* **10**, 14108, doi:10.1038/s41598-020-70671-1 (2020).

50 Hanifin, J. M. Diagnostic features of atopic dermatitis. *Acta Derm. Venereol.* **92**, 44-47 (1980).

51 Baym, M. *et al.* Inexpensive multiplexed library preparation for megabase-sized genomes. *PloS one* **10**, e0128036, doi:10.1371/journal.pone.0128036 (2015).

52 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).

53 Joshi, N. A. & Fass, J. N. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. (2011).

54 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

55 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

56 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).

57 Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome biology* **20**, 1-13 (2019).

58 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.24.436824; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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

60 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357 (2012).

61 Felsenstein, J. *PHYLIP (phylogeny inference package), version 3.5 c*. (Joseph Felsenstein., 1993).

62 Pritchard, J. K., Pickrell, J. K. & Coop, G. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology* **20**, R208-R215, doi:10.1016/j.cub.2009.11.055 (2010).

63 Lang, G. I., Botstein, D. & Desai, M. M. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* **188**, 647-661 (2011).

64 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).

65 Noble, R. (2019).

66 Sen, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association* **63**, 1379-1389, doi:10.1080/01621459.1968.10480934 (1968).

67 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272, doi:10.1038/s41592-019-0686-2 (2020).

68 Tollersrud, T., Kenny, K., Reitz, A. J., Jr. & Lee, J. C. Genetic and serologic evaluation of capsule production by bovine mammary isolates of Staphylococcus aureus and other Staphylococcus spp. from Europe and the United States. *J Clin Microbiol* **38**, 2998-3003, doi:10.1128/jcm.38.8.2998-3003.2000 (2000).

69 Lee, J. C., Liu, M. J., Parsonnet, J. & Arbeit, R. D. Expression of type 8 capsular polysaccharide and production of toxic shock syndrome toxin 1 are associated among vaginal isolates of Staphylococcus aureus. *Journal of Clinical Microbiology* **28**, 2612 (1990).

70 Madden, T. in *The NCBI Handbook [Internet]. 2nd edition* (National Center for Biotechnology Information (US), 2013).

71 Jarraud, S. *et al.* Exfoliatin-producing strains define a fourth agr specificity group in Staphylococcus aureus. *J Bacteriol* **182**, 6517-6522, doi:10.1128/jb.182.22.6517-6522.2000 (2000).