**Foodborne Origin and Local and Global Spread of *Staphylococcus saprophyticus* Causing Human Urinary Tract Infections**

Opeyemi U. Lawal, Maria J. Fraqueza, Ons Bouchami, Peder Worning, Mette D. Bartels, Maria L. Gonçalves, Paulo Paixão, Elsa Gonçalves, Cristina Toscano, Joanna Empel, Malgorzata Urbaś, M. Angeles Domínguez, Henrik Westh, Hermínia de Lencastre, Maria Miragaia

*Staphylococcus saprophyticus* is a primary cause of community-acquired urinary tract infections (UTIs) in young women. *S. saprophyticus* colonizes humans and animals but basic features of its molecular epidemiology are undetermined. We conducted a phylogenomic analysis of 321 *S. saprophyticus* isolates collected from human UTIs worldwide during 1997–2017 and 232 isolates from human UTIs and the pig-processing chain in a confined region during 2016–2017. We found epidemiologic and genomic evidence that the meat-production chain is a major source of *S. saprophyticus* causing human UTIs; human microbiota is another possible origin. Pathogenic *S. saprophyticus* belonged to 2 lineages with distinctive generic features that are globally and locally disseminated. Pangenome-wide approaches identified a strong association between pathogenicity and antimicrobial resistance, phages, platelet binding proteins, and an increased recombination rate. Our study provides insight into the origin, transmission, and population structure of pathogenic *S. saprophyticus* and identifies putative new virulence factors.

*S. saprophyticus* is the cause of uncomplicated urinary tract infection (UTI) in 10%–20% of young women (1). Despite a greater successful treatment rate, *S. saprophyticus* UTI has a higher recurrent infection frequency than *Escherichia coli* UTI (2). Rare complications of *S. saprophyticus* UTI include acute pyelonephritis, nephrolithiasis, and endocarditis (3).

*S. saprophyticus* frequently colonizes humans and can be found in the gastrointestinal tract, vagina, and perineum (4). *S. saprophyticus* also is part of the gut and rectal flora of livestock, including pigs and cattle, and a frequent contaminant of meat and fermented food products (1). *S. saprophyticus* also has been recovered from polluted aquatic environments (5).

The reservoirs of *S. saprophyticus* causing UTI in humans are believed to be endogenous, but evidence is lacking. Moreover, given the frequent bacterial contamination of meat through the meat-processing chain, meat and food are speculated to be sources of human gut colonization and human *S. saprophyticus* infection (6). Some studies have reported high genetic diversity among isolates from human infections, food products, and other sources (5,7,8). However, previous studies were performed with a limited number of isolates, which prevented the description of the global and local molecular epidemiology of *S. saprophyticus*.

We used phenotypic, genomic, and pangenome-wide association study (pan-GWAS) approaches to characterize *S. saprophyticus* both globally and locally. In addition, we identified adaptive features that drive *S. saprophyticus* evolution, defined the *S. saprophyticus* population structure, investigated dissemination routes, and identified new pathogenicity factors.

**Methods**

**Ethics Considerations**
The human isolates used in our study were recovered as part of routine clinical diagnostic testing; thus,
ethics approval and informed consent were not required. All data were handled anonymously. Sample collection was in accordance with the European Parliament and Council decision for the epidemiologic surveillance and control of communicable disease through the European Antimicrobial Resistance Surveillance Network (http://www.ecdc.europa.eu/en/activities/surveillance/EARS-Net/Pages/index.aspx). Slaughterhouse samples were part of the routine control practices for evaluation of good hygiene practices and programs to assure meat safety (European Parliament and Council regulation no. 853/2004).

Bacterial Isolates
The global S. saprophyticus collection we used included 299 isolates from humans collected in 7 countries during 1997–2017: 286 from UTIs, 12 from invasive disease, and 1 from colonization (Appendix 1 Table 1, https://wwwnc.cdc.gov/EID/article/27/3/20-0852-App1.xlsx). We also analyzed the genomes of S. saprophyticus for 38 isolates from 5 other countries: 35 isolates from human UTIs (8), 2 from human hand swabs (8), an isolate from Byzantine Troy (8), and ATCC 15305 (9), a previously investigated human UTI-causing isolate.

The local collection included isolates collected in Lisbon, Portugal, during 2016–2017: 128 human UTI isolates collected in 3 hospitals and 104 slaughterhouse isolates collected from equipment, pork samples, workers’ hands, and a pig’s rectum. In addition, we included 5 isolates from animals and 12 isolates from food used in other studies (8) (Appendix 2, https://wwwnc.cdc.gov/EID/article/27/3/20-0852-App2.pdf).

Whole-Genome Sequencing and Assembly

Phylogenetic Analysis
We performed whole-genome sequencing (WGS) on MiSeq (Illumina, https://www.illumina.com) and MinIon nanopore (Oxford Nanopore, https://nanoporetech.com) platforms, as described (10) (Appendix 2). We separately analyzed global population and local epidemiology of S. saprophyticus and their phylogeny by using single-nucleotide polymorphisms (SNPs). We identified SNPs by mapping the draft genomes to a reference genome, S. saprophyticus ATCC 15305 (GenBank accession no. AP008934.1) by using the web-based CSI phylogeny version 1.4 (11) with the default parameter, but we disabled the minimum distance between SNPs in the parameter. We used Gubbins version 2.3.4 (12) with default parameters to concatenate SNPs and remove recombination regions. We reconstructed the phylogenies by using RAxML version 8.2.4 (https://github.com/stamatak/standard-RAXML) and generalized time-reversible nucleotide substitution with gamma correction model with 100 bootstrap value. We visualized the maximum-likelihood trees by using Interactive Tree of Life (https://itol.embl.de) (Figures 1–3; Appendix 2 Figure 3). Recombination to mutation (r/m) ratios detected by using Gubbins were calculated as the average r/m of isolates in the entire collection and separately for each lineage by using as reference closed genomes of KS40 for lineage G and KS160 for lineage S, both obtained on the MinIon platform (Oxford Nanopore).

Pan-GWAS
We used Prokka version 1.13 (https://vicbioinformatics.com/software.prokka.shtml) to annotate genomes and defined the pangenome by using 85% blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) identity in Roary version 3.12 (http://sanger-pathogens.github.io/Roary). We performed GWAS by using Scoary version 1.6.16 (13) to identify genes associated with lineages and considered Bonferroni corrected p<0.05 (odds ratio [OR] >1) statistically significant; we identified genes associated with epidemiologic groups and considered Benjamini Hochberg corrected and pairwise comparison p<0.05 statistically significant. Sequence data from this study are available in GenBank (accession no. PRJNA604222).

Resistome and Virulome Analyses
We screened genomes for resistance and virulence genes by using ResFinder version 2.3 (http://cge.cbs.dtu.dk/services/ResFinder), an in-house virulence genes database, and the virulence factor database (14), integrated into ABRiccate version 0.5 (https://github.com/tseemann/abricate). We considered genes with a threshold ≥90% nucleotide identity and ≥60% coverage to be present.

Statistical Analyses
We used Prism 6.0 (GraphPad, https://www.graphpad.com) to compare the means of 2 groups. We used a 2-tailed unpaired Mann-Whitney test or χ² test for comparison and considered p<0.05 statistically significant.

Results

Lineages of S. saprophyticus Causing UTIs
Among study isolates, we first analyzed the diversity of S. saprophyticus causing UTIs by using genomic data of 321 human UTI isolates collected from 8 countries.
RESEARCH

on 4 continents during 1997–2017. From the SNPs initially detected, 42% arose from recombination events in the population, corresponding to a mean genome-wide r/m of 1.5:1, meaning that high accumulation of SNPs was due to recombination rather than mutation in UTI strains. The maximum-likelihood tree constructed from the 9,134 SNPs without recombination defined 2 lineages, which we called G and S (Figure 1). Most (74%, 236/321) UTI isolates were from lineage G and differed by 0–4,318 SNPs with an average nucleotide identity (ANI) of 98.5%–99.999%, whereas S isolates (26%, n = 85/321) were slightly less distantly

Figure 1. Maximum-likelihood tree of Staphylococcus saprophyticus isolates recovered from human infections and colonization globally, 1997–2017. The tree was constructed by using 9,134 SNPs without recombination. Among analyzed isolates, 321 were recovered from UTIs, 12 from blood, and 4 from colonization. Each node represents a strain; nodes with identical color belong to the same lineage. The assembled contigs were mapped to the reference genome S. saprophyticus ATCC 15305 (GenBank accession no. AP008934.1; black star). Polymorphic sites resulting from recombination events in the single-nucleotide polymorphism (SNP) alignments were filtered out by using Gubbins version 2.3.4 (12). Maximum likelihood tree was reconstructed by using RAxML version 8.2.4 (https://github.com/stamatak/standard-RAXML). We performed generalized time-reversible nucleotide substitution model with gamma correction with 100 bootstraps random resampling for support. We visualized the tree by using Interactive Tree of Life (iTOL; https://itol.embl.de). Black triangles represent isolates fully sequenced by using the long-read nanopore technologies and used as reference to estimate r/m in the respective lineage. Cream color represents clusters G1, G2, G3, G4, and S1, which had dissemination and transmission in same country and in different countries. The outer ring represents isolates’ country of origin; blocks with identical color represent isolates from the same country. Of note, cluster G4 contains a pair of isolates collected in 2016 that had only 10 SNPs difference; one is a blood isolate from Barcelona, Spain (KS266) and the other is a UTI isolate recovered in Lisbon, Portugal (KS135). Scale bar indicates number of substitutions per site. UTI, urinary tract infection; r/m, recombination to mutation ratio.
Figure 2. Phylogenomic analysis and distribution of *Staphylococcus saprophyticus* isolates collected from human infections and a slaughterhouse, Portugal, 2016–2017. A) Maximum-likelihood tree of 232 isolates from human infections or slaughterhouse contamination. The tree was constructed by using 14,110 single-nucleotide polymorphisms (SNPs) without recombination. Each node represents a strain; nodes with identical color belong to the same lineage. The assembled contigs were mapped to the reference genome *S. saprophyticus* ATCC 15305 (GenBank accession no. AP008934.1; black star). SNPs generated from each genome were concatenated to single alignment corresponding to position of the reference genome. Polymorphic sites resulting from recombination events in the SNP alignments were filtered out by using Gubbins version 2.3.4 (12). Tree was reconstructed by using RAxML version 8.2.4 (https://github.com/stamatak/standard-RAxML). The generalized time-reversible nucleotide substitution model with gamma correction was performed with 100 bootstrap random re-samplings for support. The tree was visualized by using Interactive Tree of Life (iTOL; https://itol.embl.de). The clusters highlighted in cream represent admixture of isolates recovered from different sources that are closely related by SNPs in clusters G5–G11 and S2–S4. The inner ring (ring 1) represents genetic relatedness of isolates recovered from different sites inside the slaughterhouses and those recovered from infection in the community. The center ring (ring 2) identifies the isolates recovered from different hospitals. The outer rings (rings 3–8) represent the distribution of 6 genes that convey antimicrobial resistance. Scale bar indicates nucleotide substitutions per site. B) Source-based distribution of *S. saprophyticus* isolates in the lineage G and lineage S. Lineage G consisted isolates from infections, colonization, and contamination. Almost all (97%) isolates in lineage S are from human colonization and infection.
related, differing by 0–3,540 SNPs with an ANI of 99.3%–99.991% (Appendix 2 Figure 1).

*S. saprophyticus* lineages we identified among UTI isolates worldwide had distinctive features indicative of 2 evolutionary histories. Although strains of both lineages had similar genome size as determined by their closed genomes (lineage G was 2.5 Mb and S 2.6 Mb), isolates from lineage S (n = 6) grew significantly faster in tryptic soy broth incubated at 37°C ($\mu_{\text{average}} = 0.34$ h$^{-1}$) than G strains (n = 8) ($\mu_{\text{average}} = 0.22$ h$^{-1}$; p = 0.0007) (Figure 4; Appendix 2). Moreover, we separately analyzed r/m of both lineages by using the respective closed genome of strains from each lineage as a reference and found that the mean estimated r/m was 9 times higher in S isolates ($r/m = 4.4:1$) than G isolates ($r/m = 0.5:1$). We did not detect a temporal signal when performing regression analysis (r) of tip-to-root distance versus isolation date, either in the entire collection (r = −0.2423) or for separate lineages (for G, r = −0.1314; for S, r = 0.1889) (Appendix). Hence, we could not determine when the 2 lineages diverged. The lack of temporal signal probably results from the limited number of isolates in each time point.

To further compare the 2 lineages, we constructed the pangenome of the 338 human *S. saprophyticus* genomes and identified 10,222 genes with 85% blastp clustering by using Roary. Among these, 8,351 genes,
Foodborne Origin of S. saprophyticus

present in <99% of the genomes, constituted the accessory genome. A gene accumulation plot of all the genes against the genomes sequenced showed that S. saprophyticus has an open pangenome (Appendix 2 Figure 2, panels A, B). The pan-GWAS analysis of the accessory genome indicated that G and S isolates have different genomic content (Bonferroni p<0.05, OR >1). A total of 128 genes were specific/enriched in the G lineage, including those encoding a type I restriction subunit (group_383), a defense mechanism against genetic transfer (15); metabolism of melibiose (ebgA, melB) (16) and inositol (iolE) (17), compounds that are excreted in urine; toxin-antitoxin systems (group_4685, group_5665), involved in stress response (18); and antimicrobial resistance (qacA) (19) (Table 1; Figure 5, panel A; Appendix 2 Table 1).

For S lineage, 237 genes were specific/enriched, some of which are involved in metabolism

Figure 4. Growth rate of Staphylococcus saprophyticus clonal lineages in tryptic soy broth (TSB) and in different concentrations of female sex hormones. All assays were performed in triplicate and each experiment was repeated 3 times. A) Growth rate of S. saprophyticus strains in different concentrations of progesterone. First panel represents growth rate in TSB at 37°C; isolates belonging to lineage S grew significantly faster (p = 0.0007) than isolates in lineage G in TSB without hormones. However, no statistically significant difference in the growth rate of either lineage was noted in physiologic (2.0–200 ng/mL) and higher concentrations of progesterone. B) Growth rate of S. saprophyticus strains in TSB (first panel) and different concentrations of estradiol. Lineage S isolates grew faster in physiologic concentrations (350 pg/mL–350 ng/mL) and higher of estradiol, suggesting that this lineage is better adapted to the hormone-rich environment of the urine and the vagina than lineage G. Error bars indicate 95% CIs; horizontal lines indicate medians. C) Growth rate mean values of S. saprophyticus strains in progesterone and estradiol.
of ascorbate (ulaA) and thiamine (tenl) (20); induction of phosphate starvation (psiE), which was previously linked to switch on virulence in other uropathogens (21); platelet binding (splE, sdrE) associated with virulence (22); steroid metabolism (group_7190); and resistance to trimethoprim (dfrG) and biocides (qacC) (Table 2; Figure 5, panel A; Appendix 2 Table 1). Additional genes associated with S lineage that could explain its increased growth rate include those linked to lactose and cellobiose metabolism (group_5572 gene) (23) and cell wall hydrolysis (lytN) (24), whereas the prevalence of transposases (group_3547 and group_1828) (25) and phage genes (recT and yueB) (27) could justify its increased recombination rate.

Local and Global Spread of S. saprophyticus

Clones Causing UTIs

We observed no time-based clustering of S. saprophyticus UTI isolates in the phylogeny, but we noted some geographic clustering. In particular, 89% (n = 78) of S isolates were found in Portugal and Spain (Figure 1). In addition, we identified clusters containing isolates from a single country. For instance, cluster G3 (1–8 SNPs) contained only isolates from Portugal and cluster G1 (0–63 SNPs) only isolates from Denmark (Figure 1).

We noted a high degree of isolate admixture in the maximum-likelihood tree of our global collection, suggesting that S. saprophyticus isolates of both lineages are disseminated geographically. G strains were distributed most widely, in 11 countries on 4 continents, whereas we found S isolates in only 6 countries. Despite the genetic diversity described, we still found isolates from different countries that differed by only a few SNPs. One pair in cluster G4 that had only 10 SNPs difference was a blood isolate from Barcelona, Spain (KS266), and a UTI isolate from Lisbon (KS135), collected in 2016 (Figure 1).

Although a relatedness cutoff is not defined yet for S. saprophyticus, the low number of SNPs observed between strains from the same and different countries is below the relatedness cutoff of 10–40 SNPs for most bacterial species (28). The apparent relatedness we noted implies that UTI isolates from different patients in the same country and in different countries are highly related and could belong to a cross-border chain of transmission. This finding challenges the assumption that S. saprophyticus causing UTIs were mainly endogenous (29).

Genetic Relatedness of S. saprophyticus from Slaughterhouses and UTIs

Pork is the most frequently consumed red meat in Europe (30) and is often contaminated with S. saprophyticus (1). We found that 35% of slaughterhouse samples (from meat, equipment, workers’ hands, and a live pig) were contaminated with S. saprophyticus.

To understand whether S. saprophyticus causing UTIs could be related to S. saprophyticus in pork, we compared 104 isolates collected from a slaughterhouse against 128 isolates collected from human UTIs in Lisbon during 2016–2017. Among the 104 isolates from the slaughterhouse, 39 (37.5%) were collected from meat, 32 (30.8%) from equipment, 32 (30.8%) from workers’ hands, and 1 (1%) from a live pig. SNP-based phylogenetic analysis with a tree-rooted at the midpoint showed that most (91%; 95/104) slaughterhouse isolates belonged to lineage G (Figure 2, panels A, B) and that a strain from slaughterhouse equipment was at the base of this lineage (bootstrap 100). In addition, the phylogenetic reconstruction including isolates from this study and other isolates

Table 1. List of genes exclusively associated with lineage G Staphylococcus saprophyticus strains in study of foodborne origin and local and global spread of S. saprophyticus causing human urinary tract infections*

| Genes         | Gene predicted function               | Biologic function group          | Frequency, % | Reference no. |
|---------------|--------------------------------------|----------------------------------|--------------|---------------|
| group_383     | Type I site-specific deoxyribonuclease | Endonuclease                      | 96           | (15)          |
| icaR          | Ica operon HTH-type negative transcription regulator |                                 |              |               |
| licT          | Transcription antiterminator LicT     | Transcription                     | 80           | NA            |
| ioIE          | Phage infection protein               |                                   | 86           | (17)          |
| group_4976    | Putative glucarate transporter        | Mobile genetic element            | 17           | NA            |
| group_5572    | Transposase for transposon Tn554      | Stress response                   | 20           | (18)          |
| group_7190    | Addiction module toxin Txe/YoeB family protein | Stress response                   | 19           | (18)          |
| group_3547    | Putative cysteine-rich protein YhjQ   | Putative function                 | 17           | NA            |

*List of genes also include 53 hypothetical proteins. Bonferroni adjusted p<0.032. NA, not applicable.
from production and companion animals (including 2 pigs, 2 bovine, and 1 canine) and food (8) showed that most (3/5) animal isolates clustered together at a basal clade of lineage G (bootstrap 100) (Appendix 2 Figure 2). Some clusters in the phylogenetic tree (e.g., G9) had slaughterhouse isolates at the base and UTI isolates at the tip. However, we also observed the opposite (e.g., G11), tree clusters with slaughterhouse isolates at the tip and UTI isolates at the base (Figure 2, panel A). Moreover, 41% of G strains included the antimicrobial resistance gene tetK (p<0.0001) (Figure 2, panel A), which is associated with resistance to tetracycline, an antimicrobial drug extensively used in animal production (31).

Phylogenetic reconstruction of all isolates from Lisbon based on SNPs provided additional examples of admixture of isolates recovered from different sampling sites in the slaughterhouse and from the slaughterhouse and humans. Isolates from meat were frequently intermixed with isolates from equipment and colonized workers as observed in cluster G6 wherein strains differed by only 1–65 SNPs (Figure 2, panel A; Figure 3). Likewise, cluster G9 included isolates from the slaughterhouse that were intermixed with human UTI isolates. Isolates collected from slaughterhouse equipment differed by only 19 SNPs from human UTI isolates. Likewise, human UTI isolates differed from meat isolates by 25 SNPs and from isolates of colonized workers by 26 SNPs. In addition, slaughterhouse and human UTI isolates had the same antimicrobial resistance profile, exhibiting resistance to fosfomycin, fusidic acid, and tetracycline (Appendix

![Figure 5. Genetic determinants that contribute to the distinction of clonal lineages and lifestyle of Staphylococcus saprophyticus. The graph displays determinants that contribute (A) and mediate (B) adaptation of S. saprophyticus to either infection or contamination. We used the genome-wide association study (GWAS) method to identify genetic factors by using 2 association comparisons: lineage G versus lineage S and human infection versus surface contamination. We used the pairwise comparison and included a core-SNP phylogenetic tree without recombination to remove the lineage effect in the analysis. Hits with Benjamini Hochberg corrected p<0.05 and odds ratio >1 were considered statistically significant. We grouped the identified genes into biologic functions based on gene annotation predicted by Prokka (https://vicbioinformatics.com/software.prokka.shtml). Some genetic factors that were associated with infections and contamination also were associated with the lineages despite subjecting the GWAS to lineage correction.](https://vicbioinformatics.com/software.prokka.shtml)

Genetic determinants that were also associated with lineage S(#) and G(*)

| Lineage G | Genes functional categories | Lineage S |
|-----------|-----------------------------|-----------|
| Tetracycline resistance |
| Antimicrobial resistance |
| Toxin/antitoxin module |
| Cell wall synthesis |
| Melibiose, galactose, Inositol, Inosose metabolism |
| Capsular polysaccharides |
| Oleate detoxification |
| Tn554, Recombinases |
| Type I restriction system |
| Antimicrobial resistance |
| Stress response |
| Nutrient utilisation |
| Virulence |
| Mobile elements |
| Genetic transfer barrier |
| Trimethoprim resistance |
| Phosphate starvation |
| Lactose/cellobiose, thiamine, ascorbate and steroid metabolism |
| Serine proteases |
| IS1181, IS431mec, Tn552 |
| Phage-related proteins |

| Infection | Genes functional categories | Contamination |
|-----------|-----------------------------|---------------|
| Trimethoprim resistance* |
| Macrolide resistance |
| Serine proteases (SraP, Sple*) |
| Accessory Secretory system II |
| IS431mec |
| Phages |
| Antimicrobial resistance |
| Virulence |
| Mobile elements |
| Tetracycline resistance* |
| Capsular polysaccharides |
| Recombinase (CcrA1, CcrB3)* |
mostly from UTIs and the slaughterhouse, S isolates
In contrast to isolates belonging to lineage G, which were
lack of temporal signal. However, we cannot ascertain directionality due to the
quently from the slaughterhouse and food to humans; (Figure 3). Transmission probably occurs more fre
are the primary sources of S. saprophyticus causing human urinary tract infections
photicus
The accumulation of substitutions and genetic dis
within the slaughterhouse, as demonstrated by
the highly related strains (<11 SNPs) on dirty and
clean equipment surfaces and similar strains (<16
SNPs) isolated 12 months apart. Carriage of the anti-
microbial resistance gene, gacA, by all G strains could
justify the observed unsuccessful cleaning procedures
(Figure 2, panel A).

The accumulation of substitutions and genetic dis-
tance evidenced by the phylogenetic analysis suggest
that isolates from slaughterhouses and food probably
are the primary sources of S. saprophyticus G strains
(Figure 3). Transmission probably occurs more fre-
quently from the slaughterhouse and food to humans;
however, we cannot ascertain directionality due to the
lack of temporal signal.

Evidence Supporting the Human Origin of Lineage S
In contrast to isolates belonging to lineage G, which were
mostly from UTIs and the slaughterhouse, S isolates
were almost exclusively of human origin (97%; n =
66/68), either from UTIs (n = 59) or human colonization
(n = 7) (Figure 2, panels A, B). When we reconstructed
the phylogeny of all isolates in this and other studies (8)
(Appendix 2 Figure 3), a human isolate was at the base
of the S lineage. The only 2 S isolates seen in animals
were from nonhuman primates. Furthermore, a resis-
tance determinant for trimethoprim (dfrG), which rou-
tinely is used to treat human UTIs, was associated with
this lineage (18%; p<0.05) (Figure 2, panel A; Table 2).

To determine whether S isolates could have origin-
ated in humans, we grew isolates from both lineages
in the absence and presence of human physiologic
concentrations (15–350 pg/mL) of estradiol (32), a
female hormone commonly found in urine and the
vagina, and in different pH values mimicking the
stomach (pH 2.5), vagina (pH ≤4.5) (33), skin (pH
5.5), and urine (pH 4.5–8.0) (34). The growth rate of
S isolates did not change significantly at the highest
physiologic concentration of estradiol specific to hu-
mans (0.34 h⁻¹ vs. 0.33 h⁻¹), but the growth rate for G
strains decreased by 59% at this concentration (0.22
h⁻¹ vs. 0.13 h⁻¹; p = 0.0007) (Figure 4, panels A–C; Ap-
pendix 2). All isolates grew at all pH levels assayed,
except for pH 2.5. At pH 4.5 and 5.5, isolates of both
lineages had similar growth rates, but S isolates had a
higher growth rate than G isolates at pH 8, although
this difference was not statistically significant (p =
0.133). These results suggest that lineage S isolates
are more adapted than lineage G strains to high estradiol concentrations found in women, but not found
in other female animal hosts, such as pigs, bovine, or
canines (35). GWAS also identified a gene involved in
steroid metabolism, 3-β hydroxysteroid dehydrogenase (HSD), associated with lineage S (Table 2; Figure 5, panels A, B). Steroids such as estradiol are primary signaling molecules for host–microbe interactions (26) and involved in the interconversion of active and inactive steroid hormones (26). Presence of HSD could be an adaptive evolution to colonization of the bladder, a hormone-rich environment. Evidence supports a human (primate) origin for lineage S, but studies sampling a wider range of ecologic sites and geographic regions are needed.

**UTIs and S. saprophyticus Dissemination among Humans in the Community**

To explore dissemination of *S. saprophyticus* causing UTIs in the community through human-to-human contact, we analyzed genomic data of the 128 UTI isolates from outpatients of 3 hospitals in the Lisbon area. Transmission of *S. saprophyticus* from lineage G and S between persons in the community was apparent, as demonstrated by the high relatedness of strains from UTI patients at different hospitals. In the G cluster, G10 isolates had 9–24 SNPs difference, and in the S cluster, S2, differed by 6–64 SNPs (Figure 2, panel A). However, due to data protection regulations, we could not ascertain whether the patients were epidemiologically linked. Results suggest that patients might have acquired these strains from the same reservoir or that direct and indirect cross-transmission could have occurred in the community.

**Disease Signatures among S. saprophyticus Populations**

Several virulence factors, including urease, have been described in *S. saprophyticus* (36), but the basis of pathogenicity in this species is mainly unknown. We used a pan-GWAS approach to compare the genetic content of 128 isolates from human infections to 104 isolates recovered from a slaughterhouse, all collected in Lisbon during 2016–2017. We identified 6 genes that appear to be associated with an increased pathogenic potential in *S. saprophyticus* (Figure 5, panel A; Appendix 2 Tables 2–5). These genes included those encoding an SplE-like protein and a gene cluster encoding a complete accessory secretory system associated with a serine-rich adhesion, similar to SraP. Previous studies have described highly homologous secretory system (>97% nucleotide identity) associated to serine-rich proteins that bind platelets, including SraP in *S. aureus* (37) and UafB in *S. saprophyticus* (38).

**Figure 6.** A proposed model for the dissemination and transmission of *Staphylococcus saprophyticus* in the community. The arrows represent the dissemination and transmission of *S. saprophyticus* isolates that belonged to lineage G (green) and lineage S (yellow). Lineage G *S. saprophyticus* strains are of animal origin and enter the slaughterhouse through production animals, such as pigs, persist on the equipment, and contaminate the meat in processing chain. Lineage G strains could enter the community through contaminated meat and workers colonized in the slaughterhouse. Lineage S strains most likely are of humans and primate origin and probably are disseminated by person-to-person contact within the community.
We also found other genes associated with infection that encoded phage proteins. An analysis on 128 UTI isolates using PHASTER (39) identified 5 phages in 48 (38%) strains. The phages were 38–125 Kb and had <50% identity with any known phage. The genomic vicinity of the identified phages varied in the chromosome, suggesting that lysogenic conversion was not the mechanism involved in pathogenicity of these strains; instead, hypothetical phage genes could be crucial for pathogenicity. In addition, genes encoding resistance to the antimicrobial drugs trimethoprim (dfrG), lincosamide (ltnA), streptogramin B (erm44v), or to macrolides (mphC-msrA) also were associated with strains from human infections (Figure 2; Figure 5, panel B; Appendix 2 Tables 2–5).

Another factor highly associated with human infections was the occurrence of recombination, which was 6 times higher in isolates from UTIs (r/m 1.7:1) than from isolates associated with colonization or contamination (r/m 0.3:1). This finding suggests recombination might be a strategy of \textit{S. saprophyticus} to evade the host immune system, as described elsewhere (40).

Discussion

We identified 2 \textit{S. saprophyticus} lineages, G and S, associated with human UTIs that appeared to have different evolutionary histories. Our data support a foodborne origin for lineage G and its transmission through food products to humans. We also found evidence of a human origin for lineage S and additional proof for the occurrence of direct or indirect human-to-human dissemination in the community, which could explain not only the local dissemination of both lineages but also the wide geographic dissemination, as described for \textit{S. aureus} (41) and \textit{E. coli} (42).

Our conclusions were limited by the characteristics of the sample collection analyzed. In particular, the lack of temporal signal did not enable inference of the direction of transmission between the sampling sites. We could establish obvious genomic relatedness between the meat production chain and human UTIs and UTIs from different persons in the community. However, lack of samples from other animal hosts, environments, or different ecologic niches in humans, did not enable us to establish pig meat and humans as the unique sources of the 2 lineages nor identify the preferred ecologic niche of \textit{S. saprophyticus} in humans.

Our observed \textit{S. saprophyticus} colonization rate among pigs was extremely low (1%) compared with previous studies (43), which could be explained by possible cross-inhibitory bacterial interactions (44); however, contamination within the slaughterhouse environment sometimes reached 35%. Amplification of \textit{S. saprophyticus} in the slaughterhouse environment was probably potentiated by resistance to biocides. We found that slaughterhouse workers’ washed hands were colonized with strains that were highly related to slaughterhouse environmental isolates and to strains causing UTIs, suggesting exposure of workers to the slaughterhouse environment as a risk factor for human colonization (Figure 6).

The transmission chain of S lineage isolates appears to be different and to have no link to meat production. S lineage UTIs might originate from human gut or vaginal colonization; both have been reported as possible human niches (2,4). Evidence for the human origin of S lineage included the almost exclusive (97%) identification in humans, where S lineage was better adapted to the physiologic high concentration of human female sex hormones; grew at vagina, skin, and urine pH values; and harbored 2 serine-proteases that presumably can bind to human platelets, as described for SraP (37). The distinct genetic content and phenotypic features of lineages G and S further reflected the diverse selective pressures of the human and animal, slaughterhouse, and food environments and suggest different evolutionary strategies toward pathogenicity. In particular, we found that tetracycline and antimicrobial resistance genes associated with colonization-contamination also were associated with isolates of the G lineage. Furthermore, genes encoding sugar metabolism, serine proteases, and a trimethoprim resistance associated with infection were also associated with isolates of the S lineage, implying distinctive specialization of the 2 lineages.

Our results also indicate a key role of setting-associated antimicrobial drug usage, especially for trimethoprim, macrolides, and tetracycline, in resistance development and pathogenicity. Subinhibitory concentrations of these drugs have been shown to promote virulence in bacteria through the induction of biofilm production (45), quorum-sensing (45), or phages (46) and might also increase \textit{S. saprophyticus} pathogenicity.

We found a high r/m rate in isolates of the S lineage and in strains from UTIs, comparable to naturally transformable bacterial species like \textit{Klebsiella pneumoniae} and \textit{Streptococcus pyogenes} (47). A similar phenomenon was observed for \textit{S. epidermidis} (48); variation in cell surface proteins was shown to contribute to evasion of human immunity (48), and a similar strategy might be advantageous for \textit{S. saprophyticus} in infection. The recombination might have resulted from defects in repair of double stranded DNA breaks from...
oxidative stress induced by leukocytes during infection, as previously described (48). However, the mechanism linking infection and recombination in S. saprophyticus remains elusive.

Last, we identified factors associated with infection that could represent new S. saprophyticus virulence factors, including 2 serine-proteases, sraP-like and splE-like, and phages SraP and SplE, which previously were connected to pathogenicity in S. aureus (22,37). In addition, phages have been described to transport pathogenicity islands in staphylococci (49). Our study constitutes a deep-structured analysis of S. saprophyticus population structure and genomic epidemiology, providing groundwork for future studies on the pathogenicity and population genetics of this bacterium.

O.U.L. was supported by PhD grants from the Fundação para a Ciência e Tecnologia (FCT) (grant no. PD/BD/113992/2015). This work was partially supported by FCT (project no. PTDC/CTM-CTM/29510/2017), Microbiologia Molecular, Estrutural e Celular (project nos. LISBOA-01-0145-FEDER-007660 and UID/Multi/04378/2019) and by the COMPETE2020 Programa Operacional Competitividade e Internacionalização; ONEIDA (project no. LISBOA-01-0145-FEDER-016417), and co-funded by Fundos Europeus Estruturais e de Investimento from Programa Operacional Regional Lisboa2020 and by national funds through FCT.

O.U.L. and O.B. cultured the isolates. O.U.L. performed the phenotypic experiments and bioinformatics analysis. O.U.L., H.W., P.W., M.D.B. performed the sequencing of the isolates. O.U.L. and M.M. carried out the data analysis and wrote the manuscript. M.J.F., M.L.G., P.P., E.G., C.T., J.E., M.U., H.M.L., H.W., M.D.B. provided the isolates. M.J.F., M.L.G., P.P., E.G., C.T., J.E., M.U., H.M.L., H.W., P.W., M.D.B. were involved in manuscript revision. All authors read and approved the final manuscript.

About the Author
Dr. Lawal is a postdoctoral researcher at the Instituto de Tecnologia Quimica e Biologica, Universidade Nova de Lisboa (ITQB-NOVA), Oeiras, Portugal. His primary research interests include antimicrobial resistance, bacterial evolution and transmission dynamics of Staphylococcus saprophyticus.

References
1. Becker K, Heilmann C, Peters G. Coagulase-negative staphylococci. Clin Microbiol Rev. 2014;27:870–926. https://doi.org/10.1128/CMR.00109-13
2. Latham RH, Running K, Stamm WE. Urinary tract infections in young adult women caused by Staphylococcus saprophyticus. JAMA. 1983;250:3063–6. https://doi.org/10.1001/jama.1983.03340220030128
3. Garduno E, Márquez I, Beteta A, Said L, Blanco J, Pineda T. Staphylococcus saprophyticus causing native valve endocarditis. Scand J Infect Dis. 2005;37:690–1. https://doi.org/10.1080/0365540510227200
4. Rupp ME, Soper DE, Archer GL. Colonization of the female genital tract with Staphylococcus saprophyticus. J Clin Microbiol. 1992;30:2975–9. https://doi.org/10.1128/JCM.30.11.2975-2979.1992
5. de Sousa VS, da-Silva APS, Sorenson L, Paschoal RP, Rabello RF, Campana EH, et al. Staphylococcus saprophyticus recovered from humans, food, and recreational waters in Rio de Janeiro, Brazil. Int J Microbiol. 2017;2017:4287547. https://doi.org/10.1155/2017/4287547
6. Hedman P, Ringertz O, Eriksson B, Kvarnforss L, Andersson M, Bengtsson L, et al. Staphylococcus saprophyticus found to be a common contaminant of food. J Infect. 1990;21:11–9. https://doi.org/10.1016/0163-4453(90)90554-L
7. Lee B, Jeong D-W, Lee J-H. Genetic diversity and antibiotic resistance of Staphylococcus saprophyticus isolates from fermented foods and clinical samples. J Korean Soc Appl Biol Chem. 2015;58:659–68. https://doi.org/10.1007/s13765-015-0091-1
8. Mortimer TD, Annis DS, O’Neill MB, Bohr LL, Smith TM, Poinar HN, et al. Adaptation in a fibronectin binding autolysin of Staphylococcus saprophyticus. mSphere. 2017;3:e00511-17. https://doi.org/10.1128/mSphere.00511-17
9. Kuroda M, Yamashita A, Hirakawa H, Kumanon M, Morikawa K, Higashide M, et al. Whole genome sequence of Staphylococcus saprophyticus reveals the pathogenesis of uncomplicated urinary tract infection. Proc Natl Acad Sci U S A. 2005;102:13272–7. PubMed https://doi.org/10.1073/pnas.0502950102
10. Hansen KH, Andreassen MR, Pedersen MS, Westh H, Jelsbak L, Schønning K. Resistance to piperacillin/tazobactam in Escherichia coli resulting from extensive blatem-associated gene amplification of blatem-1. J Antimicrob Chemother. 2019;74:3179–83. https://doi.org/10.1093/jac/dkz349
11. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Breaking through the barrier. Front Cell Infect Microbiol. 2017;2:e00511–17. https://doi.org/10.3389/fcimb.2017.000498
12. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15. https://doi.org/10.1093/nar/gku1196
13. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. et al. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 2016;17:1-9. https://doi.org/10.1186/s13059-016-1108-8
14. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis – 10 years on. Nucleic Acids Res. 2016;44(D1):D964–7. https://doi.org/10.1093/nar/gkv1239
15. Monk IR, Foster TJ. Genetic manipulation of Staphylococci-breaking through the barrier. Front Cell Infect Microbiol. 2012;2:49. https://doi.org/10.3389/fcimb.2012.00049
16. Tomita K, Nagura T, Okuhara Y, Nakajima-Adachi H, Shiogematsu N, Aritsuka T, et al. Dietary melibiose regulates cell response and enhances the induction of oral Defense.
tollerance. Biosci Biotechnol Biochem. 2007;71:2774–80. https://doi.org/10.1271/bbb.70372
17. Diniola S, Minini M, Unfer V, Verna R, Cucina A, Bizzarri M. Nutritional and acquired deficiencies in inositol bioavailability. Correlations with metabolic disorders. Int J Mol Sci. 2017;18:E2187. https://doi.org/10.3390/jms18102187
18. Gómez FA, Cárdenas C, Henríquez V, Marshall SH. Characterization of a functional toxin-antitoxin module in the genome of the fish pathogen Piscirickettsia salmonis. FEMS Microbiol Lett. 2011;317:83–92. https://doi.org/10.1111/j.1574-6968.2011.02218.x
19. Conceição T, Coelho C, de Lencastre H, Aires-de-Sousa M. High prevalence of bioicide resistance determinants in Staphylococcus aureus isolates from three African countries. Antimicrob Agents Chemother. 2015;60:678–81. https://doi.org/10.1128/AAC.02140-15
20. Costlow ZA, Degnan PH. Thiamine acquisition strategies impact metabolism and competition in the gut microbe Bacteroides thetaiotaomicron. mSystems. 2017;2:1–17. https://doi.org/10.1128/mSystems.00116-17
21. Chekabab SM, Harel J, Dozois CM. Interplay between genetic regulation of phosphate homeostasis and bacterial virulence. Virulence. 2014;5:786–93. https://doi.org/10.4161/viru.29303
22. Sharp JA, Echague CG, Hair PS, Ward MD, Nyalwidhe JO, Da Costa RR, Ferreira A, et al. Life cycle assessment of pigmeat production: Portuguese case study and proposal of improvement options. J Clean Prod. 2015;100:126–39. https://doi.org/10.1016/j.jclepro.2015.03.048
23. Ghali I, Sofyan A, Ohmori H, Shinkai T, Mitsumori M. Role of sex steroid hormones in bacterial-host interactions. Microbiology (Reading). 2011;157:1161–75. https://doi.org/10.1099/mic.0.047639-0
24. Frank LA, Mullins R, Rohrbach BW. Variability of estradiol concentration in normal dogs. Vet Dermatol. 2010;21:490–3. https://doi.org/10.1111/j.1365-3164.2010.00896.x
25. Frank LA, Mullins R, Rohrbach BW. Variability of estradiol concentration in normal dogs. Vet Dermatol. 2010;21:490–3. https://doi.org/10.1111/j.1365-3164.2010.00896.x
26. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol. 2015;13:269–84. https://doi.org/10.1038/nrmicro3432
27. Siibo OR, Chaffin DO, Rubens CE, Sullam PM. Characterization of the accessory Sec system of Staphylococcus aureus. J Bacteriol. 2008;190:6188–96. https://doi.org/10.1128/JB.00300-08
28. King NP, Beatson SA, Totsika M, Ulett GC, Alm RA, Manning PA, et al. UafB is a serine-rich repeat adhesin of Staphylococcus saprophyticus that mediates binding to fibronectin, fibrinogen and human uroepithelial cells. Microbiology (Reading). 2011;157:1161–75. https://doi.org/10.1099/mic.0.047639-0
29. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Bretscher AM, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44:W16–21. https://doi.org/10.1093/nar/gkw387
30. Yahara K, Didelot X, Jolley KA, Kobayashi I, Maiden MCJ, Dziejman M, et al. The landscape of realized homologous recombination in pathogenic bacteria. Mol Biol Evol. 2016;33:456–71. https://doi.org/10.1093/molbev/msw237
31. Tristán A, Bes M, Meugnier H, Lina G, Bozdogan B, Courvalin P, et al. Global distribution of Panton-Valentine leukocidin—positive methicillin-resistant Staphylococcus aureus, 2006. Emerg Infect Dis. 2007;13:594–600. https://doi.org/10.3201/eid1304.061316
32. Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L, Galanakis C, et al. Food reservoir for Escherichia coli causing urinary tract infections. Emerg Infect Dis. 2010;16:88–95. https://doi.org/10.3201/eid1601.091118
33. Hedman P, Ringertz O, Lindström M, Olsson K. The origin of Staphylococcus saprophyticus from cattle and pigs. Scand J Infect Dis. 1997;29:357–60. https://doi.org/10.1080/03085901.1997.10876858
34. Yahara K, Didelot X, Jolley KA, Kobayashi I, Maiden MCJ, Sheppard SK, et al. The landscape of realized homologous recombination in pathogenic bacteria. Mol Biol Evol. 2016;33:456–71. https://doi.org/10.1093/molbev/msw237
35. Tritan A, Bes M, Meugnier H, Lina G, Bozdogan B, Courvalin P, et al. Global distribution of Panton-Valentine leukocidin—positive methicillin-resistant Staphylococcus aureus, 2006. Emerg Infect Dis. 2007;13:594–600. https://doi.org/10.3201/eid1304.061316
36. Chopra I, Roberts M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev. 2001;65:232–60. https://doi.org/10.1128/MMBR.65.2.232-260.2001
37. Walker SW. Laboratory reference ranges. In: Endocrine self-assessment program. Washington, D.C.: Endocrine Society; 2015. p. 1–5 [cited 2019 Apr 29]. https://education.endocrine.org/system/files/ESAP%202015%20Laboratory%20Reference%20Ranges.pdf
38. Linhares IM, Minis E, Robial R, Witkin SS. The human vaginal microbiome. In: Faintuch J, Faintuch S, editors. Microbiome and metabolome in diagnosis, therapy, and other strategic applications. London: Elsevier; Inc.; 2019. p. 109–14. https://doi.org/10.1016/B978-0-12-815249-2.00011-7
39. Clarkson MR, Magee CN, Brenner BM. Chapter 2: Laboratory assessment of kidney disease. In: Clarkson MR, Magee CN, Brenner BM eds. Pocket companion to Brenner and Rector’s the Kidney 8th edition. London: Elsevier; Inc.; 2011. p. 21–41 [cited 2019 Apr 29]. https://www.sciencedirect.com/science/article/pii/B978141606480000026
40. Yahara K, Didelot X, Jolley KA, Kobayashi I, Maiden MCJ, Dziejman M, et al. The landscape of realized homologous recombination in pathogenic bacteria. Mol Biol Evol. 2016;33:456–71. https://doi.org/10.1093/molbev/msw237
41. Tristan A, Bes M, Meugnier H, Lina G, Bozdogan B, Courvalin P, et al. Global distribution of Panton-Valentine leukocidin—positive methicillin-resistant Staphylococcus aureus, 2006. Emerg Infect Dis. 2007;13:594–600. https://doi.org/10.3201/eid1304.061316
42. Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L, Galanakis C, et al. Food reservoir for Escherichia coli causing urinary tract infections. Emerg Infect Dis. 2010;16:88–95. https://doi.org/10.3201/eid1601.091118
43. Hemdan P, Ringertz O, Lindström M, Olsson K. The origin of Staphylococcus saprophyticus from cattle and pigs. Scand J Infect Dis. 1993;25:537–60. https://doi.org/10.1080/03085901.1993.10876858
44. Versstappen KM, Willems E, Fluit AC, Duin B, Martens M, Wagenaar JA. Staphylococcus aureus nasal colonization differs among pig lineages and is associated with the presence of other staphylococcal species. Front Vet Sci. 2017;4:97. https://doi.org/10.3389/fvets.2017.00097
45. Imperi F, Leoni L, Visca P. Antivirulence activity of azithromycin in Pseudomonas aeruginosa. Front Microbiol. 2014;5:178. https://doi.org/10.3389/fmicb.2014.00178
46. Goerke C, Köller J, Wolz C. Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in *Staphylococcus aureus*. Antimicrob Agents Chemother. 2006; 50:171-7. https://doi.org/10.1128/AAC.50.1.171-177.2006

47. Davies MR, McIntyre L, Mutreja A, Lacey JA, Lees JA, Towers RJ, et al. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. Nat Genet. 2019;51:1035–43. https://doi.org/10.1038/s41588-019-0417-8

48. Méric G, Mageiros L, Pensar J, Laabei M, Yahara K, Pascoe B, et al. Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. Nat Commun. 2018;9:5034. https://doi.org/10.1038/s41467-018-07368-7

49. Dearborn AD, Dokland T. Mobilization of pathogenicity islands by *Staphylococcus aureus* strain Newman bacteriophages. Bacteriophage. 2012;2:70–8. https://doi.org/10.4161/bact.20632

Address for correspondence: Maria Miragaia, Laboratory of Bacterial Evolution and Molecular Epidemiology, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa (ITQB-NOVA), Avenida da República, 2780-157, Oeiras, Portugal; email: miragaia@itqb.unl.pt

April 2020

**Vectorborne Infections**

- Stemming the Rising Tide of Human-Biting Ticks and Tickborne Diseases, United States
- Ecology and Epidemiology of Tickborne Pathogens, Washington, USA, 2011–2016
- Imported Arbovirus Infections in Spain, 2009–2018
- Decreased Susceptibility to Azithromycin in Clinical Shigella Isolates Associated with HIV and Sexually Transmitted Bacterial Diseases, Minnesota, USA, 2012–2015
- High Incidence of Active Tuberculosis in Asylum Seekers from Eritrea and Somalia in the First 5 Years after Arrival in the Netherlands
- Severe Dengue Epidemic, Sri Lanka, 2017
- Severe Fever with Thrombocytopenia Syndrome, Japan, 2013–2017
- Comprehensive Profiling of Zika Virus Risk with Natural and Artificial Mitigating Strategies, United States
- Genomic Insight into the Spread of Meropenem-Resistant *Streptococcus pneumoniae* Spain-ST81, Taiwan
- Isolation of Drug-Resistant *Gallibacterium anatis* from Calves with Unresponsive Bronchopneumonia, Belgium
- Guaroa Virus and *Plasmodium vivax* Co-Infections, Peruvian Amazon
- Intensified Short Symptom Screening Program for Dengue Infection during Pregnancy, India
- Prevalence of Antibodies to Crimean-Congo Hemorrhagic Fever Virus in Ruminants, Nigeria, 2015
- Recurrent Herpes Simplex Virus 2 Lymphocytic Meningitis in Patient with IgG Subclass 2 Deficiency
- Health-Related Quality of Life after Dengue Fever, Morelos, Mexico, 2016–2017
- Person-to-Person Transmission of Andes Virus in Hantavirus Pulmonary Syndrome, Argentina, 2014
- Ebola Virus Neutralizing Antibodies in Dogs from Sierra Leone, 2017
- Outbreak of *Dirkmeia churashimaensis* Fungemia in a Neonatal Intensive Care Unit, India
- Rift Valley Fever Outbreak, Mayotte, France, 2018–2019
- Crimean-Congo Hemorrhagic Fever Virus in Humans and Livestock, Pakistan, 2015–2017
- Detection of Zoonotic Bartonella Pathogens in Rabbit Fleas, Colorado, USA
- Human-to-Human Transmission of Monkeypox Virus, United Kingdom, October 2018
- Whole-Genome Analysis of *Salmonella enterica* Serovar Enteritidis Isolates in Outbreak Linked to Online Food Delivery, Shenzhen, China, 2018
- Pruritic Cutaneous Nematodiasis Caused by Avian Eyeworm *Oxyspirura* Larvae, Vietnam
- Novel Rapid Test for Detecting Carbapenemase
- Arthritis Caused by MRSA CC398 in a Patient without Animal Contact, Japan
- Detection of Rocio Virus SPH 34675 during Dengue Epidemics, Brazil, 2011–2013
- Epidemiology of Lassa Fever and Factors Associated with Deaths, Bauchi State, Nigeria, 2015–2018
- Plague Epizootic Dynamics in Chipmunk Fleas, Sierra Nevada Mountains, California, USA, 2013–2015
- Knowledge of Infectious Disease Specialists Regarding Aspergillosis Complicating Influenza, United States

To revisit the April 2020 issue, go to: https://wwwnc.cdc.gov/eid/articles/issue/26/4/table-of-contents
Foodborne Origin and Local and Global Spread of Staphylococcus saprophyticus Causing Human Urinary Tract Infections

Appendix 2

Supplementary Methods

Bacterial Isolates

In addition to the global and local Staphylococcus saprophyticus collections stated in the manuscript, we included S. saprophyticus isolates from food production animals (1 pig and 3 bovine), 1 companion canine, 12 food isolates, and 1 isolate recovered from a river, which was described in previous studies (1). We also included 18 S. saprophyticus isolates recovered from food products and 2 isolates from nonhuman primates from our collection. These isolates, together with the global and local collection (n = 480) was analyzed to infer the origin of the S. saprophyticus lineages.

Isolation and Species Identification of Slaughterhouse Isolates

We enriched samples in peptone water and grew isolates on CHROMagar Staph aureus (CHROMagar Microbiology, https://www.chromagar.com) supplemented with 10% NaCl and 4 µg/mL novobiocin. We extracted genomic DNA by using methods previously described (2). We performed species identification by amplifying and sequencing the tuf gene (3).

Growth Rate in Different pH and Hormones

For representative isolates from the collection, we performed growth curves at different concentrations of female sex hormones, including progesterone at 20 ng/mL, 20 µg/mL, and 20 mg/mL, and estradiol at 350 pg/mL, 350 ng/mL, and 350 µg/mL; and at pH levels of 2.5, 4.5, 5.5, and 8.0. We performed growth assays by using an Infinite 200 PRO series microtiter reader (Tecan Group Ltd, https://www.tecan.com) in 96-well microtiter plates. For each strain, overnight culture was inoculated onto 5 mL Bacto tryptic soy broth (TSB; Becton Dickinson,
The OD$_{600\text{nm}}$ of the liquid culture was adjusted to an initial OD of 0.5 MacFarland with buffers for specific pHs and TSB containing varying concentrations of hormones and grown with aeration (180 rpm) at 37°C for 18 h. Assays were performed in triplicate and each experiment was repeated 3 times.

**Estimation of Evolutionary Rates**

To estimate the evolutionary rates in *S. saprophyticus* population, as a first approach, we explored the degree and pattern of temporal signal and determined whether sufficient temporal signals were available in the *S. saprophyticus* phylogeny. We performed a regression of the divergence of each tip from the root against the date of sampling, a root-to-tip plot, of the global collection and separately for the lineages using TempEst v1.5.3 (4). We used the phylogenetic tree without recombination and the date of isolation of the isolates as inputs.

**Average Nucleotide Identity Analysis**

We calculated average nucleotide identity (ANI) for representative strains of *S. saprophyticus* 40 G lineages and 20 S lineages by using a standalone Python program, pyani version 0.2.9 (https://github.com/widdowquinn/pyani) and the ANIb option, which compares genomes using BLAST program (https://blast.ncbi.nlm.nih.gov). The closed genome of KS40 was used as a reference for lineage G and closed genome of KS160 was used for lineage S.

**Intrasample Diversity**

We assessed the genetic diversity between isolates recovered from the same sample in the meat processing chain. We determined whether intrasample diversity existed by comparing the SNP differences between these isolates.

**Data Availability**

All raw sequence data are available in the SRA (https://www.ncbi.nlm.nih.gov/sra) under the study accession no. PRJNA604222. We also provide individual accession numbers for raw sequence data (Appendix 1 Table 1) and the SNP matrices and list of genes in the pangenomes (Appendix 1 Tables 2–6).
Results

Pangenome Analysis of *S. saprophyticus* Revealed an Open Pangenome

We annotated the 338 *S. saprophyticus* genomes by using Prokka (5) and constructed the pangenome by using Roary (6) with 85% blastp identity. A total of 10,222 genes were found, 48% (n = 4,925) of which were genes with unknown functions. The genes constituting the core of all isolates consisted of 1,871 genes. Also, we noted 118 soft core genes in 95%–99% of the isolates, 856 shell genes in 15%–94%, and we found 7,307 genes that constituted cloud genes in <15% of *S. saprophyticus* population. On average, 75% of *S. saprophyticus* genome is constituted by core genes and 25% of accessory genes. The plot of the total number of genes against the number of genomes indicate an open pangenome in which each genome sequence added several new genes. This finding implies that newly sequenced genomes will identify new genes and the pangenome size of this species will continue to increase (Appendix 2 Figure 2).

GWAS Revealed Genetic Factors Associated with *S. saprophyticus* Isolates from Different Genetic Lineages and Clinical Origins

We explored the pangenome gene presence to understand the difference in the genetic content of isolates from each of the genetic lineages defined by core SNPs. We used Scoary pipeline (7) and Bonferroni p<0.05 to identify genes that were exclusive or enriched in the *S. saprophyticus* genetic lineages. We categorized the hits into biologic function groups based on the annotations predicted by Prokka. For genes associated with different clinical origins (infection and colonization/contamination), we used Benjamini Hochberg and pairwise p<0.05 (Appendix 2 Tables 1–5).

References

1. Mortimer TD, Annis DS, O’Neill MB, Bohr LL, Smith TM, Poinar HN, et al. Adaptation in a fibronectin binding autolysin of *Staphylococcus saprophyticus*. MSphere. 2017;2:e00511–17. PubMed [https://doi.org/10.1128/mSphere.00511-17](https://doi.org/10.1128/mSphere.00511-17)

2. Couto I, Pereira S, Miragaia M, Sanches IS, de Lencastre H. Identification of clinical staphylococcal isolates from humans by internal transcribed spacer PCR. J Clin Microbiol. 2001;39:3099–103. PubMed [https://doi.org/10.1128/JCM.39.9.3099-3103.2001](https://doi.org/10.1128/JCM.39.9.3099-3103.2001)

3. Martineau F, Picard FJ, Ménard C, Roy PH, Ouellette M, Bergeron MG. Development of a rapid PCR assay specific for *Staphylococcus saprophyticus* and application to direct detection from urine
samples. J Clin Microbial. 2000;38:3280–4. PubMed https://doi.org/10.1128/JCM.38.9.3280-3284.2000

4. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2:vew007. PubMed https://doi.org/10.1093/ve/vew007

5. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9. PubMed https://doi.org/10.1093/bioinformatics/btu153

6. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3. PubMed https://doi.org/10.1093/bioinformatics/btv421

7. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Erratum to: Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 2016;17:1–9. PubMed https://doi.org/10.1186/s13059-016-1108-8

8. Tomita K, Nagura T, Okuhara Y, Nakajima-Adachi H, Shigematsu N, Aritsuka T, et al. Dietary melibiose regulates the cell response and enhances the induction of oral tolerance. Biosci Biotechnol Biochem. 2007;71:2774–80. PubMed https://doi.org/10.1271/bbb.70372

9. Costliow ZA, Degnan PH. Thiamine acquisition strategies impact metabolism and competition in the gut microbe Bacteroides thetaiotaomicron. [Internet]. mSystems. 2017;2:1–17 http://msystems.asm.org/lookup/doi/10.1128/mSystems.00116-17. PubMed https://doi.org/10.1128/mSystems.00116-17

10. Sharp JA, Echague CG, Hair PS, Ward MD, Nyalwidhe JO, Geoghegan JA, et al. Staphylococcus aureus surface protein SdrE binds complement regulator factor H as an immune evasion tactic. PLoS One. 2012;7:e38407. PubMed https://doi.org/10.1371/journal.pone.0038407

11. Hallet B, Sherratt DJ. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. FEMS Microbiol Rev. 1997;21:157–78. PubMed https://doi.org/10.1111/j.1574-6976.1997.tb00349.x

12. Datta S, Costantino N, Zhou X, Court DL. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. Proc Natl Acad Sci U S A. 2008;105:1626–31. PubMed https://doi.org/10.1073/pnas.0709089105

13. Imperi F, Leoni L, Visca P. Antivirulence activity of azithromycin in Pseudomonas aeruginosa. Front Microbiol. 2014;5:178. PubMed https://doi.org/10.3389/fmicb.2014.00178
14. Goerke C, Köller J, Wolz C. Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in *Staphylococcus aureus*. Antimicrob Agents Chemother. 2006;50:171–7. PubMed https://doi.org/10.1128/AAC.50.1.171-177.2006

15. Siboo IR, Chaffin DO, Rubens CE, Sullam PM. Characterization of the accessory Sec system of *Staphylococcus aureus*. J Bacteriol. 2008;190:6188–96. PubMed https://doi.org/10.1128/JB.00300-08

16. King NP, Beatson SA, Totsika M, Ulett GC, Alm RA, Manning PA, et al. UafB is a serine-rich repeat adhesin of *Staphylococcus saprophyticus* that mediates binding to fibronectin, fibrinogen and human uroepithelial cells. Microbiology (Reading). 2011;157:1161–75. PubMed https://doi.org/10.1099/mic.0.047639-0

17. Monk IR, Foster TJ. Genetic manipulation of Staphylococci-breaking through the barrier. Front Cell Infect Microbiol. 2012;2:49. PubMed https://doi.org/10.3389/fcimb.2012.00049

18. Chopra I, Roberts M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev. 2001;65:232–60. PubMed https://doi.org/10.1128/MMBR.65.2.232-260.2001
Appendix 2 Table 1. List of differentially enriched genes in *Staphylococcus saprophyticus* lineages in a study of isolates from human urinary tract infections and meat processing plants.*

| Gene      | Gene predicted function                  | Biologic function group         | Lineage 55%, % | Lineage 32%, % | Reference no. |
|-----------|------------------------------------------|---------------------------------|-----------------|----------------|---------------|
| meiB      | Melibiose carrier protein                | Sugar transport and metabolism  | 98              | 57             | (8)           |
| ebgA      | Evolved β-galactosidase subunit α        | Sugar transport and metabolism  | 98              | 26             | (8)           |
| csxA      | Exo-β-D-glucosaminidase                 | Sugar transport and metabolism  | 62              | 1              | (8)           |
| arsA      | Arsenical pump-driving ATPase            | Metal resistance                | 28              | 2              | NA            |
| arsD      | Arsenical resistance operon transacting | Metal resistance                | 28              | 2              | NA            |
| tenl      | Thiazole tautomerase                    | Thiamine biosynthesis           | 22              | 100            | (9)           |
| splE      | S1B family serine protease SplE         | Virulence                       | 15              | 100            | (10)          |
| sdrE      | Serine-rich repeat-containing protein    | Virulence                       | 35              | 74             | (10)          |
| mhpC      | Arylesterase                            | Hydrolase                       | 26              | 60             | NA            |
| group_1205| Transcriptional regulator               | Transcriptional regulator       | 81              | 1              | NA            |
| group_4356| Rho termination factor domain-containing | Transcriptional regulator       | 4               | 59             | NA            |
| qacA      | Antiseptic resistance protein           | Biocide resistance              | 100             | 5              | NA            |
| qacC      | Quaternary ammonium compound-resistance protein QacC | Biocide resistance | 35              | 87             | NA            |
| group_2160| Chaperone ATPase                        | Putative functions              | 16              | 84             | NA            |
| group_330 | Spore coat protein                      | Putative functions              | 12              | 46             | NA            |
| binS      | Recombinase/resolvase                   | Mobile genetic element          | 32              | 3              | NA            |
| group_4660| Putative replication-associated protein  | Mobile genetic element          | 28              | 1              | NA            |
| group_2182| Putative replication-associated protein  | Mobile genetic element          | 54              | 97             | NA            |
| group_3547| IS1161 transposase                      | Mobile genetic element          | 45              | 78             | (11)          |
| group_1828| Transposase-associated ATP/GTP binding protein | Mobile genetic element | 15              | 59             | (11)          |
| group_1679| Transposase for transposon Tn552        | Mobile genetic element          | 2               | 25             | (11)          |
| group_278 | Recombinase/resolvase                   | Mobile genetic element          | 1               | 20             | (11)          |
| yueB      | Phage infection protein                  | Phage-related protein           | 38              | 85             | (12)          |
| recT      | Putative phage-related DNA recombination protein | Phage-related protein | 15              | 59             | (12)          |
| group_2472| Phage N-acetylglucosaminidase           | Phage-related protein           | 15              | 59             | (12)          |
| group_2102| Phage protein                           | Phage-related protein           | 15              | 59             | NA            |
| group_2856| Phage protein                           | Phage-related protein           | 15              | 59             | NA            |
| group_2857| Phage protein                           | Phage-related protein           | 15              | 59             | NA            |
| group_3414| Phage protein                           | Phage-related protein           | 15              | 59             | NA            |
| group_1521| Phage tape measure protein              | Phage-related protein           | 15              | 59             | NA            |
| group_2854| Putative phage DNA-packaging protein    | Phage-related protein           | 15              | 59             | NA            |
| group_2855| Putative phage head-tail adaptor        | Phage-related protein           | 15              | 59             | NA            |
| group_4784| Putative phage minor structural protein | Phage-related protein           | 15              | 59             | NA            |
| group_1829| PVL phage protein                       | Phage-related protein           | 15              | 59             | NA            |
| group_2103| Phage N-acetylglucosaminidase           | Phage-related protein           | 15              | 59             | NA            |
| group_2855| Phage N-acetylglucosaminidase           | Phage-related protein           | 15              | 57             | (11)          |
| group_1640| Phage portal protein, SPP1 family      | Phage-related protein           | 15              | 56             | NA            |
| group_3412| Phage terminase, large subunit          | Phage-related protein           | 15              | 56             | NA            |
| group_2860| Phage protein                           | Phage-related protein           | 14              | 56             | NA            |
| group_4359| Holin protein                           | Phage-related protein           | 14              | 56             | NA            |
| group_3894| Phage minor structural protein GP20     | Phage-related protein           | 14              | 47             | NA            |
| group_3895| Phage minor head protein                | Phage-related protein           | 13              | 47             | NA            |
| group_4788| Phage transcriptional regulator         | Phage-related protein           | 7               | 46             | NA            |
| group_3800| Bacteriophage transcriptional regulator | Phage-related protein           | 5               | 24             | NA            |
| group_4678| Bacteriophage integrase                 | Phage-related protein           | 5               | 24             | NA            |
| group_6539| Bacteriophage terminase small subunit   | Phage-related protein           | 2               | 43             | NA            |

*Bonferroni p < 0.00002. Genes encoding hypothetical proteins enriched in lineage G = 55; genes encoding hypothetical proteins enriched in lineage S = 9. NA, not applicable.
### Appendix 2 Table 2. List of genes that were exclusively associated with *Staphylococcus saprophyticus* isolates recovered from urinary tract infections*%

| Gene            | Gene predicted function                                             | Biologic function group     | % Infection | Reference no. |
|-----------------|---------------------------------------------------------------------|------------------------------|-------------|---------------|
| group_1652      | Putative DNA primase-phage associated                               | Phage-related protein        | 15          | NA            |
| group_2800      | Putative phage leukocidin protein                                  | Phage-related protein        | 8           | NA            |
| group_4438      | Phage minor structural GP20                                         | Phage-related protein        | 8           | NA            |
| group_1406      | Hypothetical protein                                               | Uncharacterized protein      | 15          | NA            |
| group_1405      | Hypothetical protein                                               | Uncharacterized protein      | 15          | NA            |
| group_4443      | Hypothetical protein                                               | Uncharacterized protein      | 10          | NA            |
| group_3431      | Hypothetical protein                                               | Uncharacterized protein      | 10          | NA            |
| group_863       | Hypothetical protein                                               | Uncharacterized protein      | 9           | NA            |
| group_4447      | Hypothetical protein                                               | Uncharacterized protein      | 9           | NA            |
| group_4446      | Hypothetical protein                                               | Uncharacterized protein      | 9           | NA            |
| group_4445      | Hypothetical protein                                               | Uncharacterized protein      | 9           | NA            |
| group_4439      | Hypothetical protein                                               | Uncharacterized protein      | 9           | NA            |
| group_3617      | Hypothetical protein                                               | Uncharacterized protein      | 7           | NA            |
| group_3616      | Hypothetical protein                                               | Uncharacterized protein      | 7           | NA            |
| group_1466      | Hypothetical protein                                               | Uncharacterized protein      | 7           | NA            |

*Benjamini Hochberg p<0.02. NA, not applicable.

### Appendix 2 Table 3. List of genes that were enriched in *Staphylococcus saprophyticus* isolates recovered from urinary tract infections*%

| Gene            | Gene predicted function                                             | Biologic function group     | % Infection | % Contamination | Reference no. |
|-----------------|---------------------------------------------------------------------|------------------------------|-------------|-----------------|---------------|
| group_4400      | Mph(C) macroline 2* phosphotransferase                             | Antimicrobial resistance     | 25          | 2               | (13,14)       |
| dfrG            | Dihydrofylate reductase                                            | Antimicrobial resistance     | 9           | 1               | (14)          |
| csoR            | Copper-sensing transcriptional repressor CsoR                      | Metal resistance             | 28          | 3               | NA            |
| cadX            | Putative cadmium efflux system                                     | Metal resistance             | 44          | 13              | NA            |
| rep             | Plasmid replication initiation protein                             | Mobile genetic element       | 18          | 3               | NA            |
| group_2868      | Protein rlx                                                        | Mobile genetic element       | 17          | 2               | NA            |
| group_422       | Transposase for IS431mec                                           | Mobile genetic element       | 14          | 2               | (11)          |
| group_425       | Transposase for IS431mec                                           | Mobile genetic element       | 12          | 2               | (11)          |
| group_1094      | Bacteriophage integrase                                            | Phage-related protein        | 38          | 8               | (12,14)       |
| group_4449      | DNA packaging protein Staph phage phiRS7                            | Phage-related protein        | 19          | 4               | (12,14)       |
| group_858       | Holin protein                                                      | Phage-related protein        | 16          | 2               | (12,14)       |
| group_1653      | Phage protein                                                      | Phage-related protein        | 15          | 2               | (14)          |
| group_1392      | Phage protein                                                      | Phage-related protein        | 13          | 2               | (14)          |
| group_800       | Bacteriophage tail tape                                            | Phage-related protein        | 13          | 2               | (14)          |
| splE            | S1B family serine protease SprE                                    | Virulence                    | 55          | 24              | (10)          |
| group_3377      | Accessory Sec system protein SprE                                  | Virulence                    | 27          | 3               | (15,16)       |
| secY_2          | Preprotein translocase subunit SecY2                               | Virulence                    | 27          | 3               | (15,16)       |
| secA2           | Sec family Type I general secretory pathway protein                | Virulence                    | 27          | 3               | (15,16)       |
| asp3            | Accessory Sec system protein Asp3                                  | Virulence                    | 27          | 3               | (15,16)       |
| asp2            | Accessory Sec system protein Asp2                                  | Virulence                    | 27          | 3               | (15,16)       |
| sraP            | Serine-rich repeat-containing protein                              | Virulence                    | 14          | 1               | (15,16)       |
| grxC            | Glutaredoxin 3                                                    | Stress response              | 30          | 7               | NA            |
| kefF            | Glutathione-regulated potassium-efflux system ancillary protein KeF | Stress response              | 16          | 3               | NA            |
| yhjQ            | Putative cysteine-rich protein                                     | Stress response              | 9           | 1               | NA            |
| splE            | S1B family serine protease SprE                                    | Virulence                    | 55          | 24              | (10)          |
| group_3377      | Accessory Sec system protein SprE                                  | Virulence                    | 27          | 3               | (15,16)       |
| Gene          | Gene predicted function                        | Biologic function group | % Infection | % Contamination | Reference no. |
|--------------|-----------------------------------------------|-------------------------|-------------|-----------------|---------------|
| secY_2       | Preprotein translocase subunit SecY2           | Virulence               | 27          | 3               | (15,16)       |
| secA2        | Sec family Type I general secretory pathway protein SecA2 | Virulence               | 27          | 3               | (15,16)       |
| asp3         | Accessory Sec system protein Asp3             | Virulence               | 27          | 3               | (15,16)       |
| asp2         | Accessory Sec system protein Asp2             | Virulence               | 27          | 3               | (15,16)       |
| sraP         | Serine-rich repeat-containing protein          | Virulence               | 14          | 1               | (15,16)       |
| grxC         | Glutaredoxin 3                                | Stress response         | 30          | 7               | NA            |
| kefF         | Glutathione-regulated potassium-efflux system ancillary protein KefF | Stress response         | 16          | 3               | NA            |
| yhjQ         | Putative cysteine-rich protein YhjQ           | Stress response         | 9           | 1               | NA            |

*Benjamini Hochberg p<0.01. Hypothetical proteins (n = 19 genes). NA, not applicable.

**Appendix 2 Table 4.** List of genes that were exclusively associated with *Staphylococcus saprophyticus* recovered from environmental sources*

| Gene          | Gene predicted function                        | Biologic function group | % Contamination | References |
|--------------|-----------------------------------------------|-------------------------|-----------------|------------|
| group_1467   | Hypothetical protein                          | Uncharacterized protein | 42              | NA         |
| group_466    | Hypothetical protein                          | Uncharacterized protein | 39              | NA         |
| group_1991   | Hypothetical protein                          | Uncharacterized protein | 32              | NA         |
| group_6148   | Hypothetical protein                          | Uncharacterized protein | 18              | NA         |
| group_6146   | Hypothetical protein                          | Uncharacterized protein | 18              | NA         |

*Benjamini Hochberg p<0.00006. NA, not applicable.

**Appendix 2 Table 5.** List of genes that were enriched for in *Staphylococcus saprophyticus* isolates recovered from environmental sources*

| Gene          | Gene predicted function                        | Biologic function group | % Contamination | % Infection | References |
|--------------|-----------------------------------------------|-------------------------|-----------------|-------------|------------|
| group_227    | Type I site-specific deoxyribonuclease restriction subunit | Restriction system      | 83              | 32          | (17)       |
| ccrB         | Cassette chromosome recombinase B             | Mobile genetic element  | 61              | 9           | NA         |
| ccrA         | Cassette chromosome recombinase A1           | Mobile genetic element  | 54              | 4           | NA         |
| group_1878   | Putative replication-associated protein       | Mobile genetic element  | 49              | 17          | NA         |
| group_3001   | Myosin-cross reactive antigen (Oleate hydratase) | Stress tolerance       | 60              | 6           | NA         |
| tetK         | Tetracycline resistance protein               | Antimicrobial resistance| 51              | 12          | (18)       |
| cap5O        | Capsular polysaccharide biosynthesis protein Cap5O | Capsule                | 15              | 1           | NA         |
| cap5M        | Capsular polysaccharide biosynthesis galactosyltransferase Cap5M | Capsule                | 13              | 2           | NA         |
| group_1472   | Hypothetical protein                          | Uncharacterized protein | 86              | 56          | NA         |
| group_1179   | Hypothetical protein                          | Uncharacterized protein | 69              | 24          | NA         |
| group_353    | Hypothetical protein                          | Uncharacterized protein | 61              | 9           | NA         |
| group_2224   | Hypothetical protein                          | Uncharacterized protein | 59              | 18          | NA         |
| group_1627   | Hypothetical protein                          | Uncharacterized protein | 55              | 25          | NA         |
| group_2291   | Hypothetical protein                          | Uncharacterized protein | 17              | 1           | NA         |
| group_3863   | Hypothetical protein                          | Uncharacterized protein | 10              | 1           | NA         |

*Benjamini Hochberg p<0.0001. NA, not applicable.
Appendix 2 Figure 1. Measure of genetic diversity between *Staphylococcus saprophyticus* lineage G (A) and lineage S (B) determined by using ANI. Lineage G strains had ANI values of 98.5%–99.999% and appear to be more diverse compared with lineage S strains, which had ANI values of 99.3%–99.991% and were slightly less diverse. Most isolates in lineage G had a lower ANI (99.6%) than the isolates in lineage S (99.7%). The ANI results were comparable to the genetic diversity observed with single nucleotide polymorphism analysis. ANI, average nucleotide identity.

Appendix 2 Figure 2. Pangenome of *Staphylococcus saprophyticus* inferred from 338 isolates recovered from human infections and colonization. Among analyzed isolates, 321 were recovered from UTIs, 12 from blood, 4 from colonization, and 1 from reference strain ATCC 15305 (https://www.atcc.org; GenBank accession no. AP008934.1). A) Distribution of genes in the pangenome generated using Roary (6). We found a total of 10,222 genes. The core genes shared by all isolates were constituted of 1,871 genes. We also found 188 soft core genes in 95%–99% of isolates, and 856 shell genes in 15%–94% of isolates. In addition, we noted 7,307 cloud genes <15% of *S. saprophyticus* population. B) Gene accumulation plot for *S. saprophyticus* pangenome as a function of genomes sequenced indicating that *S. saprophyticus* has an open pangenome.
Appendix 2 Figure 3. Single nucleotide polymorphism-based maximum likelihood tree of 480 *Staphylococcus saprophyticus* from different sources. Each node represents a strain. A node with identical color belongs to the same lineage. The assembled contigs were mapped to the reference genome *S. saprophyticus* ATCC 15305 (https://www.atcc.org; GenBank accession no. AP008934.1) and SNPs were called. SNPs generated from each genome were concatenated to single alignment corresponding to position of the reference genome. Polymorphic sites resulting from recombination events in the SNP alignments were filtered out by using Gubbins v2.3.4 (Sanger, https://sanger-pathogens.github.io/gubbins). Maximum likelihood tree was reconstructed using RAxML version 8.2.4 (https://github.com/stamatak/standard-RAxML). The generalized time reversible nucleotide substitution with gamma correction was performed with 100 bootstraps random resampling for support. The image was generated using Interactive Tree of Life (https://itol.embl.de). The colored ring represents the source.
of isolates. The green triangles represent strains recovered from pigs, green stars represent strains from bovines, and green circle a strain from a domestic canine. A slaughterhouse isolate recovered from equipment and those from pigs, bovines, and canine were at the base of lineage G, suggesting a probable foodborne origin of this lineage. Conversely, a human infection isolate was at the base of lineage S implying a human origin of this lineage. The green arrows in lineage S depict isolates recovered from small nonhuman primates.

Appendix 2 Figure 4. Growth rate of *Staphylococcus saprophyticus* clonal lineages in different pH levels. A) pH 2.5 representing pH of the human stomach; B) pH 4.5 and C) pH 5.5 representing pH of human skin; and D) pH 8.0 representing pH of urine from a healthy human. Isolates were completely inhibited at pH 2.5 but grew at a low rate when pH = 4.5 and 5.5. The 2 lineages behaved slightly differently in pH = 8.0 but this difference was not statistically significant. Assays were performed in triplicates and each experiment was repeated 3 times. Error bars indicate 95% CI; horizontal lines indicate median.