Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

_____

Signature of Student Date

BIOTIC AND ABIOTIC FACTORS ASSOCIATED WITH THE COLONIZATION OF STAPHYLOCOCCUS AUREUS AND ITS SUBTYPES IN HEALTHY HUMAN SUBJECTS

BY

Sandeep Jose Joseph PhD
Degree to be awarded: M.P.H.
Executive MPH

Lyndsey Darrow PhD	Date
Timothy D. Read	Date
Laura Gaydos PhD Associate Chair for Academic Affairs, Executive MPH program	Date

BIOTIC AND ABIOTIC FACTORS ASSOCIATED WITH THE COLONIZATION OF STAPHYLOCOCCUS AUREUS AND ITS SUBTYPES IN HEALTHY HUMAN SUBJECTS

BY

Sandeep Jose Joseph PhD
PhD., University of Georgia, 2007
BVSc & AH., Kerala Agriculture University, 2003

Thesis Committee Chair: Lyndsey Darrow, PhD

An abstract of
A Thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements of the degree of
Master of Public Health in the Executive MPH program
2016

Abstract

BIOTIC AND ABIOTIC FACTORS ASSOCIATED WITH THE COLONIZATION OF STAPHYLOCOCCUS AUREUS AND ITS SUBTYPES IN HEALTHY HUMAN SUBJECTS

BY
Sandeep Jose Joseph PhD

Background: *Staphylococcus aureus,* a human commensal and prevalent human pathogen, affects public health worldwide. It is a common asymptomatic colonizer predominantly in the nares, and also at the oral cavity and skin. Neither the role of carriage in the propagation of *S. aureus* infections nor the factors associated with the colonization of a particular subtype at a body site are well understood. The purpose of this study was to assess associations between demographic and life history characteristics and the profile of *S. aureus* subtypes identified at each body site using a metagenome-based subtyping scheme using data generated by the human microbiome project (HMP).

Materials and Methods: The metagenomic samples were collected from various body sites of healthy 18 - 40 years old adults. The exposure variables investigated in relation to the subtype profile of *S. aureus* in a body site were diet, breastfed, tobacco use, health insurance, history of surgery, age, BMI and ethnicity. Both binary (*S. aureus* +/-) and multinomial (4 outcomes: 3 subtypes of *S. aureus* (CC8, CC30, any other subtypes), vs. no detection of *S. aureus*) logistic regression were performed to identify predictors for *S. aureus* detection among HMP participants.

Results: In the binary outcome logistic regression model, main body site ($p < 0.001$), health insurance (OR for no health insurance=0.5 (0.2-1.0); p=0.0525) and BMI (OR for high BMI vs. normal BMI=1.7 (1.1-2.5); p=0.0276) were predictors of detection of *S. aureus*, whereas for the multinomial logistic regression model with 4 outcomes, only main site and BMI were significant ($p<0.05$) predictors of the presence of *S. aureus* at significance level of 0.1. Compared to subjects with normal BMI, the odds of detecting CC8 subtype tended to be higher in high BMI subjects (OR=1.4, 95% CI=0.6-3.0) while CC30 subtype detection was higher in those with low BMI (OR=1.6, 95% CI=0.6-3.8).

Conclusions: Results suggest that high BMI and health insurance are risk factors for *S. aureus* colonization. Larger studies with more heterogeneous subjects are needed to identify predictors of *S. aureus* subtype colonization in human body sites.

BIOTIC AND ABIOTIC FACTORS ASSOCIATED WITH THE COLONIZATION OF STAPHYLOCOCCUS AUREUS AND ITS SUBTYPES IN HEALTHY HUMAN SUBJECTS

BY

Sandeep Jose Joseph PhD
PhD., University of Georgia, 2007
BVSc & AH., Kerala Agriculture University, 2003

Thesis Committee Chair: Lyndsey Darrow, PhD

An abstract of
A Thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements of the degree of
Master of Public Health in the Executive MPH program
2016

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Lyndsey Darrow for her valuable advice and support through out the entire period of this study. I would also like to thank my mentor and thesis committee member Timothy D. Read for his guidance and inspiration that helped me to conduct this study.

I would like to thank my parents in India who always encouraged me to persue this MPH at Emory University.

Finally, I am gratefull to my wife, Sereen Abraham and my son, Nathan Joseph, who had sacrificed a lot of weekends and deserved family time for letting me finish my assignments. I would like to dedicate this thesis to my wife and son.

TABLE OF CONTENTS

Background and Literature Review

Staphylococcus aureus, a human commensal, also a prevalent human pathogen, affects public health worldwide. It accounts for most of bloodstream and soft tissue infections in developed countries [1][2][3][4] and is responsible for more deaths in the US than HIV [5]. This pathogen is also one of the most common hospital infections, often leading to chronic disease with poor outcome [6] [7] [8]. *S. aureus* also causes community-acquired and nosocomial bacterial infections in humans [5] [9]. It can cause infections that can range from mild skin infections to severe, highly invasive and necrotizing diseases [10]. Even though this organism is a part of the natural microbiome of humans, clones of epidemic drug-resistant *S. aureus* (healthcare-associated (HA) and community-acquired (CA) methicillin-resistant *S. aureus* (MRSA)) have emerged. *S. aureus* carries a higher mortality rate of 65 - 75% in the prebiotic era and currently 20-40% mortality at 30 days despite appropriate treatment [11] [12] [13].

S. aureus nasal carriage is a global phenomenon, which seems to be affected by various factors including, but not limited to, age, health, economic status and the country of residence [14]. Around 25-50% of the world population is persistently colonized with *S. aureus* in the nares, with 60-100% of individuals having *S. aureus* colonized at some point in their lifespan [15][16]. Global trends in *S. aureus* nasal carriage showed a larger variation were cohort studies found *S. aureus* carriage within continental USA vary from 26% to 32% [17], 10% in adults in Turkey [18] and 25% in Malaysia [19]. Two trends are evident in literature regarding *S. aureus* nasal carriage, 1) developed countries have high incidence rate for *S. aureus* nasal carriage (US [35%] [17], Netherlands [35%] [20], Norway [27%] [21], Switzerland [36.4%] [22] and Japan

[35.7%] [23] as compared to underdeveloped and developing countries (Nigeria [14%] [24], Pakistan [14.8%] [25], India [16%] [26], Tunisia [13%] [27], Malaysia [25%] and Indonesia [<10%] [28]); 2) increased rates of *S. aureus* nasal carriage was observed amongst intravenous drug users [29] and immunocompromised individuals [30]. From population based studies a number of modifiable risk factors have been found associated with carriage, including oral contraceptives use, smoking, crowding and health care exposure [31].

Apart from the most abundant presence of *S. aureus* in the anterior nares in the general adult population, *S. aureus* can also commonly be found at other body sites such as the axillae (8%), chest/abdomen (15%), perineum (22%), intestine (17-31%) [32] and vagina (5%) [33]. A recent study in Sweden have indicated the importance of throat colonisation, where they found *S. aureus* throat carriage was significantly higher than nasal carriage rate both for patients (40% vs. 31%) and staff (54% vs. 36%) [34]. In a Swiss study of risk factors for *S. aureus* throat carriage, including 3464 subjects, the prevalence of exclusive throat carriage was 30.2% among colonised individuals from the community and 18.4% among hospitalised patients and healthcare workers [22]. Analyzes of the metagenomic data generated by the human microbiome project (HMP) detected the presence of *S. aureus* in 4 other body sites of healthy individuals including retroauricular crease, tongue dorsum, supragingival plaque and gastrointestinal tract (stool) [35].

Even though majority of *S. aureus* nasal colonization happens asymptomatically in healthy individuals, such infections are thought to be a major source of bacterial transmission in the human population [36]. At the same time neither the factors of human colonization nor the

role of carriage in the propagation of *S. aureus* infections are well understood. Many studies have shown that persistent nasal carriage of *S. aureus* is a risk factor for pathogenic infection, but the overall association of these carriage strains to the presence of endogenous strains that establish pathogenic infections is currently unknown [15] [37] [13]. However, a recent study demonstrated lack of evidence of patient-to-patient intrahospital transmission of invasive *S. aureus* strains [38].

S. aureus is considered as a heterogeneous bacterial species, but the majority of human diseases are caused by a relatively small subset of clones. It is important to understand the genetic diversity (population structure) of *S. aureus* strains that colonize the different body sites of healthy individuals in order to study how the commensal *S. aureus* strains present in healthy human population might act as a predisposing factor for future invasive infections. Bacterial species are commonly comprised of multiple phylogenetic clades that have distinctive phenotypic properties. The process of identifying which clade a bacterial strain belongs in goes by several names but here we will refer to it as *subtyping*. Commonly used subtyping methods include multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), oligotyping and variable-number of tandem-repeat typing (VNTR) [39]. Each of these these methods was developed for bacteria first isolated in pure culture in the laboratory before DNA extraction. Current subtyping methods in *S. aureus*, including multilocus sequence typing (MLST), *spa* typing and SCC*mec* typing (used to identify methicillin-resistant *S. aureus* (MRSA), rely on sequencing short segments of a few genes and lack the resolution to differentiate related but distinct clones or isolates. For early disease diagnosis of pathogenic *S.*

aureus and to understand bacteria in the context of their natural community, it would be advantageous to subtype directly from clinical specimens such as blood and sputum. However, current direct identification options such as 16S rRNA gene sequence, FISH and REP-PCR, are not able to subtype *S. aureus* to below the species level taxonomic resolution, nor to deal with mixtures of subtypes of the same species being present. At the same time, several studies using next-generation DNA sequencing technology to generate whole genome sequences of *S. aureus* isolates have proven to be able to distinguish between isolates, which would have been grouped as identical strains using traditional typing methods.

Metagenomic approaches can be used to sequence libraries of DNA isolated from different body sites of human subjects for pathogen detection. Over the past few years, the expansion of "metagenomic" culture-free shotgun sequencing of biological samples have helped us to assay the presence and/or collective genome of the microbes living in and on human bodies (microbiome/microbiota). The first wave of bacterial metagenomics studies have predominantly sequenced 16S ribosomal genes to identify bacterial organisms. But again, less resolution of 16S ribosomal genes will be a drawback for this approach because low levels of genetic variation within *S. aureus* species (below that can be detected using 16S primers) are associated with different patterns of infection at multiple body sites. However, as the cost of genome sequencing has decreased over the years, the direct, unamplified shotgun sequencing of DNA samples extracted from different body sites (exemplified by the human microbiome project (HMP) [35]) would provide an unbiased snapshot of the biodiversity and population structure of the strains

within a species by tracking strain-specific single nucleotide polymorphisms (SNPs) obtained through reference mapping to already known high quality whole genome sequences.

Epidemiologists can characterize the complex microbiota among large human populations and help understand the direct and indirect impacts on disease outcome. Some of the questions that can be addressed by epidemiologists are: Are there measures of microbiome structure or function that corresponds to health or disease outcomes?; Are these measures risk markers, risk factors, or modifiers of either?, How amenable are they to intervention?; What is the variability of various measures of microbiota by person, place, and time and how do these change by host, agent/pathogen, and environment?

Introduction

S. aureus is one of the most common hospital infections, often causing chronic diseases with poor outcome [1][2][3][4] [6] [7] [8]. It is also a problem outside the hospital as a community-acquired bacterial infections in humans [5] [9], livestock [40] [41] and other animals[42]. *S. aureus* is a common asymptomatic colonizer of humans, with the nares (nose) believed to be the most important site. Estimates for human nasal carriage rates suggest ~20-50% of humans are persistently colonized with *S. aureus*, with 60-100% of individuals harboring *S. aureus* at some point in their lifespan [15][16][43]. The population of *S. aureus* asymptomatically colonizing the nose in healthy individuals is thought to be a major source for transmission [36]. Many studies have shown that persistent nasal carriage of *S. aureus* is a risk factor for pathogenic infection, but the overall association of these carriage strains to the presence of endogenous strains that establish pathogenic infections is currently unknown [15] [37] [13].

S. aureus strains can be classified into a limited number of clonal lineages of related MLST sequence types (clonal complexes[44]) , which differ in their geographical distribution and propensity to cause human diseases. The acquisition of the SCCmec cassette, producing the MRSA strains is more common in some clonal lineages than others[45], as is the acquisition of *vanA* genes to produce VRSA (vancomycin resistant *S. aureus*)[46]. It is important to understand the genetic diversity (population structure) of *S. aureus* strains that colonize the different body sites in order to understand how commensal strains present in healthy human population might act as a predisposing factor for future invasive infections.

Several studies have looked into the epidemiology, biotic and abiotic factors contributing to *S. aureus* nasal carriage. It was found that asymptomatic *S. aureus* nasal carriage was high in developed countries when compared with underdeveloped and developing countries [17] [47] [18] [19]. Other contributing factors towards *S. aureus* infestation in nostrils identified were ethnicity [24] [26] [28], availability of medical care [14], intravenous drug usage [29] and HIV infections [30]. Even though previous studies have inferred that host genetics could be of modest influence for the presence of *S. aureus* [48], a very recent study identified that *S. aureus* in the nostrils was an environmentally derived trait and it has nothing to do with the host genetics [43]. The same study also determined that the presence of *S. aureus* in the nostrils is influenced by the absolute abundance of nasal microbiota. Apart from the most abundant presence of *S. aureus* in the anterior nares, the metagenomic data generated by the human microbiome project (HMP) detected the presence of *S. aureus* in 4 other body sites of healthy individuals including retroauricular crease, tongue dorsum, supragingival plaque and gastrointestinal tract (stool) [35]. All these studies only looked into epidemiology and factors influencing the presence of *S. aureus* at the species level, but none looked into the strain level diversity at the various human body sites other than the nostrils.

To better understand whether subtypes/strains of *S. aureus* adapt to different niches in the healthy human body, Joseph et al. 2015 [49] developed a bioinformatics analysis strategy for mapping the shotgun metagenomic HMP data, and also implemented a statistical genotyping scheme that utilizes already known strain-specific SNPs to predict the most likely genetic background(s) of *S. aureus* at the strain/subtype-level resolution in each of the body samples. As

a follow up of Joseph et al, 2015 [49] findings of the different strains of *S. aureus* in various human body sites, we performed epidemiological modelling to understand whether there is any association between the demographic and life history characteristics collected using the responses that subjects gave to an extensive survey, and the different strains of *S. aureus* identified at each body site.

Material and Methods

Classifying *S. aureus* subtypes based on a binomial mixture model

Joseph et al, 2015 [49] used *binstrain* software [50], implemented in the R language [51] to perform *S. aureus* subtype classification. *binstrain* used a binomial mixture model to estimate the proportion of subtypes based on a DNA alignment against an ancestor *S. aureus* reference genome and a matrix of SNPs that distinguish different genetic subtypes (construction of the matrix described below). *binstrain* assumes a binomial probability distribution, p_i of observing a SNP, x_i in the entire genome and n_i denotes the total nucleotide coverage at position i. $Z_{i,j}$ is an indicator function specifying whether j^{th} strain has a SNP at i^{th} position. In the final version of the classifier, we used 102,057 SNP positions across the genome to classify *S. aureus* into 40 subtypes.

$$x_i \sim Binom(n_i, p_i), i = 1,.....102,057$$
$$p_i = \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + ... + \beta_{40} Z_{i,40}, i = 1,.....102,057$$

The estimation of β_i indicates the proportion of *S. aureus* reference strain-specific SNPs present in a clinical or purified sample. At the strain–specific SNP positions, there will be only a few β_i *s*

that affects $p_{i.}$ Other β_i *s* have no impact on p_i because their corresponding Z_{ij} are 0's, which makes it a sparse design matrix. We utilized this sparsity of the design matrix in order to perform a well-established step-by-step procedure to estimate all the β_i *s* using quadratic programming [50].

Sequence data analysis and statistical modeling for SNP based genotyping of *S. aureus* strains in the HMP

Joseph et al 2015 [49] obtained raw mwgs sequence data in fastq files for a total of 1265 samples (human DNA removed using NCBI's BMTagger tool) from the HMP ftp site (ftp://public-ftp.hmpdacc.org/Illumina). The HMP carried out 2 phases of metagenomic whole genome shotgun sequencing (mwgs), performed using the Illumina GAIIx platform with 101 bp paired-end reads. For Phase 1, 764 samples were chosen from 103 adults and for Phase II, 400 samples were chosen from 67 adults. Samples were chosen covering 16 body sites. The Phase 1 data sets have been described previously [35] [52]. In short, the reads in each of the sample FASTQ files were mapped against the ancestor *S. aureus* reference genome to generate the base call and coverage (average read depth) in each position in the mpileup output format using the Burrows-Wheeler Aligner (BWA) (Version: 0.6.1-r104) short-read aligner[53] by specifying the maximum number of gap extensions (e) to be 10. The resultant short-read alignment files for each samples were converted to mpileup format using the mpileup option in SamTools software along with the –B option that disables probabilistic realignment for the computation of base alignment quality (BAQ). The resultant mpileup file for each sample were used as an input for

the binStrain algorithm and the β values that indicates the proportion of a particular *S. aureus* strain present were estimated.

Selection criteria used in recruiting health subjects for the HMP

The HMP used rigorous and good clinical practice standards to complete comprehensive body site sampling in healthy 18 - 40 years old adults. Many subjects were students, staff and faculty at two major universities, Baylor College of Medicine at Houston and Washington University in St. Louis. To make sure that the specimens collected represented minimally perturbed microbiomes, HMP first screened potential participants using exclusion criteria based on health history and excluded those with hypertension, cancer or immunodeficiency or autoimmune disorders, recent use of immunomodulators and antibiotics or probiotics. Subsequent screening using physical examination excluded individuals based on body mass index (BMI), cutaneous lesions and oral health. Out of 554 subjects screened 300 were enrolled. There were 149 men and 151 women, mean age of 26 years, mean BMI of 24 kg/m^2 , 20.0% racial minority and 10.7% Hispanic. The specimens were obtained from the oral cavity, nares, skin, gastrointestinal tract and vagina (15 specimens from men and 18 from women). The HMP study evaluated longitudinal changes in an individual's microbiome by sampling 279 participants twice (mean 212 days after the first sampling; range 30-359 d) and 100 individuals 3 times (mean 72 d after the second sampling; range 30-224 d). This sampling strategy yielded 11,174 primary specimens, from which 12,479 DNA samples were submitted to 4 centers for metagenomic sequencing. After quality control, 6212 specimens were used for 16S rRNA sequencing via 454 pyrosequencing and 1263 samples were sequenced using 101bp paired-end Illumina shotgun metagenomic reads [54].

Selection of metadata from the HMP for epidemiological analysis

A large amount of demographic and clinical data were collected for each of the individuals sampled for the HMP. We obtained access to the most recent version of these data through a formal request to the dbGap database (accession phs000228.v3.p1). We used the metadata from the 170 healthy individuals (phase I and II) from whom the HMP mwgs sequence data were generated. Because of the generally high level of health of the subjects, for most clinical variables there were too few cases to have realistic odds of association. For the epidemiological analysis, we included only those body sites where *S. aureus* was detected in at least 20% of the samples. We also grouped the body sites into three superclasses based on their proximity within the human body as well as to increase power: airways (anterior nares); oral cavity (attached keratinized gingiva, buccal mucosa, palatine tonsils, saliva, supragingival plaque and tongue dorsum) and skin (right and left retroauricular crease). There were a total of 840 samples collected from 133 participants in the HMP, used in the epidemiological analysis (described below).

Epidemiological Modeling

The binary categorical variables (exposure variables) from the metadata, which we investigated in relation to presence of *S. aureus* and/or a particular *S. aureus* subtype in a body site were gender, breastfed or not, tobacco use, insurance information and history of previous surgery (Table 1). Other categorical variables used were diet (Meat/fish/poultry at least three days per week, Meat/fish/poultry at least one day but not more than two days per week and

Eggs/cheese/other dairy products, but no meat/fish/poultry), race/ethnicity (Hispanic, Asian, non-Hispanic Black and non-Hispanic white), BMI (< 22, 22 - 25 & > 25). Age was treated as a continuous variable that ranged from 18 to 40 years of age (Table 1).

We performed both binary (model 1) and multinomial (with 4 outcomes, model 2) logistic regression to identify predictors for *S. aureus* detection among HMP participants. The binary outcome indicated whether the presence of *S. aureus* was detected or not detected (reference) (model 1), while the 4 outcomes for the multinomial logit model were the presence/detection of *S. aureus* CC8, CC30, any other *S. aureus* CC types and no detection of *S. aureus* (reference) (model 2). Initially crude odds ratios were estimated for each of the 10 exposure variables for both the binary and multinomial outcomes. Adjusted odds ratios were also estimated by fitting the full multivariate logistic regression model with all the exposure variables, and the binary and multinomial outcomes separately.

Odds ratios were estimated by fitting generalized linear mixed models using SAS PROC GLIMMIX (Version 9.4, Cary, NC) with main site and other exposure variables (described above) as fixed effects and random effects for subject in order to assess any possible association of the exposure variables and the presence/detection of *S. aureus*.

Results

Assignment of *S. aureus* subtypes in the HMP metagenomic dataset using whole genome subtyping

Joseph et al., 2015 [49] developed and tested the *S. aureus* binstrain classifier by calling the subtypes present in 1,263 whole metagenomic sequencing samples from the healthy human

cohort of the phase 1 and 2 of the HMP (170 subjects). They found at least one sequencing read mapping to the SA_ASR sequence in 348 of the samples (27.5%) isolated from 110 (36.3%) of the subjects. The presence of the species was variable across body sites, most commonly found in the left and right retroauricular creases and anterior nares (100%, 90% and 57%, respectively) and least common in the stool and subgingival plaque (6% and 5%, respectively) (Table 2). While the presence of the *S. aureus* reads in a sample will be dependent on factors such as the complexity of the microbiome and the amount of sequence data collected, this result was in line with estimates of *S. aureus* presence based on bacterial culture[15].

Of the 348 *S. aureus* positive samples, 321 had a *S. aureus* core coverage > 0.025X (Table 2). *S. aureus* was more prevalent at this level of coverage in the anterior nares, retroauricular creases and tongue dorsum. 165 (51%) of these samples were dominated by one subtype (largest β value > 0.8). In the other samples where there was a mixture of dominant subtypes, we used a conservation cutoff for a subtype being present as at least a minor component if the β value was > 0.2 (chosen to conservatively remove overcalls due to random errors in sequence reads). Based on these definitions, the most commonly detected subtypes were CC30, CC8, CC45, CC398 and CC5 (present in 112 (35%), 72 (22%), 32 (10%), 29 (9%) and 26 (8%) of samples, respectively) (Figure 1)

Biotic and abiotic factors associated with *S. aureus* and its subtypes

We performed epidemiologic modeling using generalized linear mixed models to assess whether any metadata variables on the subjects of the study collected by the HMP were associated either with the presence of *S. aureus,* or with a specific subtype. In order to increase

power we aggregated body sites into three categories: airways (anterior nares), oral cavity (attached keratinized gingiva, buccal mucosa, palatine tonsils, saliva, supragingival plaque and tongue dorsum) and skin (right and left retroauricular crease). In the binary outcome logistic regression model, at an alpha level of 0.1, main body site (p-value <0.001), having health insurance or not (p-value = 0.0525) and BMI (p-value = 0.0276) were predictors for the detection/presence of *S. aureus*, whereas for the multinomial logistic regression model with 4 outcomes, only main site (p-value = <0.001) and BMI (p-value = 0.0251) were predictors of the presence of *S. aureus* (Table 3). The estimated odds ratio for detecting the presence of any *S. aureus* subtypes in the airways compared to the oral cavity was 3.3 (95% CI: 2.2 - 5.0). This is consistent with our study and other previous studies showing that *S. aureus* is highly enriched in the anterior nares (nose) compared to any other body sites, and also body site could be a strong predictor for the presence of *S. aureus*. In the multinomial model where specific subtypes were examined, odds of detection of CC8, CC30 and other subtypes were all significantly elevated in the airways compared to the oral cavity (Table 3). Similarly, the odds of detecting any *S. aureus* subtypes in subjects with higher BMI (>25) was 70% higher when compared to subjects with normal BMI. In the multinomial model, odds of detection of CC8, CC30 and other subtypes were all elevated for higher (>25) vs. normal (22-25) BMI, but the higher odds of detection was more pronounced for the other subtype group (OR CC8=1.4, 95% CI=0.6-3.0; OR CC30=1.1, 95%CI= 0.5, 2.2; OR other subtype=2.4, 95% CI=1.3-4.5). Also the odds of detecting CC8 subtype tended to be higher in high BMI subjects while CC30 subtypes appeared to be associated with lower BMI. In the binary outcome model, subjects without health insurance had less detection of *S. aureus* compared to subjects with health insurance, with an estimated 50% lower odds of

detection of any *S. aureus* subtypes among the uninsured. Even though race and ethnicity overall was not a statistically significant predictor (p-value=0.28) for the detection of *S. aureus*, there was some indication that the odds of identifying any *S. aureus* subtype was higher among Hispanics compared to Non-Hispanic whites, with the odds ratio most elevated for detection of CC8 in the multinomial model (Table 3). However, we note that these odds ratios were based on only 45 samples from Hispanics included in our analysis (13 with detection of any *S. aureus*). Based on our analysis gender was not a significant predictor for the detection of any *S. aureus* subtypes (p-value=0.77). The odds of detecting *S. aureus* in females were 10% less than in males (Table 3). Age, breast fed or not, tobacco use (also previously identified in [43] and history of surgery were not significant predictors of association of the presence of *S. aureus* subtypes in any human body sites (p-value>0.1). Even though CC398 was enriched at the tongue dorsum we could not find any statistically significant association with the presence of CC398 in the tongue dorsum and eating a diet that contains meat. The unadjusted (crude) odds ratios are shown in Table 4.

Discussion and Conclusions

S. aureus is a versatile pathogen capable of growth and infection under various conditions. It is the most common cause of human skin and soft tissue infections, especially in post-surgical patients under an immunosuppressive drug regime, pediatric and geriatric patients, diabetics and immunocompromised patients. The pathogen can be contracted either in a hospital setting (nosocomial) or from the community (community acquired). In the hospital units, *S. aureus* infection is a growing threat because of the rapid acquisition and evolution of antibiotic resistance. The ‘community acquired’ infections are transmitted between people in a normal

population. Some people carry this pathogen in various body sites (carriers with *S. aureus* colonization), especially in the anterior nares, thus serving reservoirs of infections and transmission of the pathogen at the community level. There is considerable evidence indicating that such colonization in healthy people is an important risk factor for future invasive infection, while the reasons behind this phenomenon are unclear. Understanding such biotic and abiotic risk factors that leads to *S. aureus* colonization in healthy individuals are important to prevent the spread of the pathogen in a community as well as hospital setting.

Advances in DNA sequencing technologies have created a new field of research, called metagenomics, where we now have the tools to identify the various species of bacteria, viruses, archaea, and fungi that live in and on our various body sites, which is known as microbiota. The ability to conduct a census of human microbiota is unprecedented; until the development of genomic technologies, we were able to identify only those microbes that could be grown in the lab. The immediate outcome of such advanced technologies is the NIH Common Fund Human Microbiome Project (HMP) that characterized the microbial communities found at several different sites on the healthy human body: nasal passages, oral cavity, skin, gastrointestinal tract, and urogenital tract. The metagenomic data generated from the HMP is of tremendous use for the *S. aureus* community to understand the various *S. aureus* strains/CC types that colonizes the various body sites in a healthy cohort. The large amount of demographic and clinical data collected for each of the healthy individuals sampled for the HMP can be associated with the colonization of *S. aureus* in various body sites; thereby would help in understanding the various risk factors responsible for the subclinical colonization and would also allow policymakers to

draft efficient awareness campaigns of lifestyle modifications to control and subdue the spread of this pathogen in the community.

In this study, we utilized the high resolution strain/CC-type *S. aureus* subtyping information generated by Joseph et al., 2015 [49] from 1,263 whole metagenomic sequencing samples from various body sites of the healthy human cohort of the phase 1 and 2 of the HMP (170 subjects) and performed logistic regression to understand whether there is any association with the demographic and life history characteristics and their status of *S. aureus* colonization at the various body sites. Of the epidemiologic variables only body mass index (BMI) > 25 and possession of health insurance were associated with the presence of *S. aureus*. This study also confirmed previous results [43] that gender was not a significant predictor for the detection of any *S. aureus* subtypes. However, the odds of detecting *S. aureus* in females were 10% less than in males, which shows similar trends to previous culture-based studies that showed men are more likely to be colonized by *S. aureus* than females [48] [55] [56]. Based on the results of the multinomial logit model to understand the association of each of the *S. aureus* carriage subtypes and the exposure variables, there were no strong links to the carriage of the two major subtypes, CC30 and CC8. Understanding the reasons behind the distributions of *S. aureus* subtypes will take larger data sets.

The subjects chosen for the HMP study were fairly homogenous in terms of age, ethnicity [57], and absence of medical conditions, leaving little power to associate with conditions more prevalent in the general population. Majority of the subjects were educated professionals from two highly ranked universities in the US whose age ranged from only 18 - 40 years. Due to small sample size and less variation among the study subjects in the HMP, we did not assess the

interaction between the exposure variables selected in this analysis. Even though age, tobacco use and race/ethnicity were not significant predictors from our analysis, a previous study using 2,115 women and 1674 men and within a wide age range of 30 - 87 years found that sex (gender), age and smoking are risk factors for *S. aureus* carriage [56]. This study suggested interaction by age with the largest sex differences in carriage rates in younger adulthood and similar sex-specific carriage rates in children and the elderly. Olsen et al., 2012 [56] also reported *S. aureus* carriage rate was 28% lower in smokers than in nonsmokers ($P < 0.01$). Moreover, data from the large NHANES sample (9622 persons >= 1 year old) suggest interaction between sex and race/ethnicity; male gender was a significant risk factor for *S. aureus* carriage in the non-Hispanic White and Mexican American populations but not in the non-Hispanic black population. Being non-Hispanic white compared to non-Hispanic black seems to be a risk factor for *S. aureus* carriage among men (OR = 1.7, 95% CI 1.4–2.0), but not among women [9] [31]. One important think to note is that all these previous studies have been culture based and were unable to look at specific strains/subtypes, unlike a shotgun metagenomic dataset used in this study. All these indicates that this cohort is under sampled and definitely not a good representation to perform such epidemiological modeling. Most of the studies, including the HMP, identify a particular target population (examples, students, hospital workers, infants in neonatal ward or geriatric patients) and perform epidemiological analysis of the association of *S. aureus* colonization in that cohort with respect to certain standard variables collected from that population. Very few studies go beyond a particular population to select a heterogenous population were much more power and reliable results for epidemiological association can be achieved.

A major limitation of our epidemiologic analysis was the imprecision in our estimated odds ratios for detection of *S. aureus* driven mainly by the small number of study subjects and to some extent by the homogeneity of the HMP population for factors such as age, race, tobacco use, health insurance, diet, and history of surgery **(**Table 1**)**. Despite this, we did observe evidence for more detection of *S. aureus* among subjects with higher BMI compared to those with normal BMI, and suggestion towards lower detection of *S. aureus* among subjects without health insurance. We did not adjust for multiple comparisons and it is always possible that observed associations are due to chance. If high BMI and health insurance are in truth risks for *S. aureus* presence, these relationships may be connected to health factors outside those collected directly. For example, BMI may be associated with diet, exercise or other behaviors that are more directly to *S. aureus* carriage; likewise the lower detection among uninsured people may reflect less contact with the medical system or socioeconomic factors.

Even though microbiomic research is promising for epidemiological investigations, the fact that microbiome are dynamic in nature, and the variation within an individual can be high, could make such analysis more complex than expected. Our understanding of the factors responsible for such individual variations in microbiota is limited, which also leads to limited understanding of what factors might confound or modify observed associations between the microbiome and disease outcome. Well-conducted, population-based longitudinal studies are essential to filling these knowledge gaps. For epidemiologists, beyond the ability to process huge amounts of data from microbiome/metagenomics studies, the real challenge lies in the best way to achieve the data reduction needed to use these data in epidemiologic analyses. Epidemiologists can make an important contribution to microbiomic research by performing

well-designed, well-conducted, and appropriately powered studies and by including measures of microbiota composition, identifying important confounders and effect modifiers, and generating and testing hypotheses addressing the role of microbiome in health and disease [58].

References

1. King MD, Humphrey BJ, Wang YF, Kourbatova EV, Ray SM, Blumberg HM. Emergence of community-acquired methicillin-resistant Staphylococcus aureus USA 300 clone as the predominant cause of skin and soft-tissue infections. Ann Intern Med. Am Coll Physicians; 2006;144: 309–317.

2. Seybold U, Kourbatova EV, Johnson JG, Halvosa SJ, Wang YF, King MD. SM Ray, and HM Blumberg. 2006. Emergence of community-associated 17 methicillin-resistant Staphylococcus aureus USA300 genotype as a major cause of 18 health care-associated blood stream infections. Clin Infect Dis. 16AD;42: 647–656.

3. Mera RM, Suaya JA, Amrine-Madsen H, Hogea CS, Miller LA, Lu EP, et al. Increasing Role of Staphylococcus aureus and Community-Acquired Methicillin-Resistant Staphylococcus aureus Infections in the United States: A 10-Year Trend of Replacement and Expansion. Microb Drug Resist. 2011;17: 321–328.

4. O'Hara FP, Amrine-Madsen H, Mera RM, Brown ML, Close NM, Suaya JA, et al. Molecular Characterization of Staphylococcus aureus in the United States 2004–2008 Reveals the Rapid Expansion of USA300 Among Inpatients and Outpatients. Microb Drug Resist. 2012;18: 555–561.

5. Klevens RM, Morrison MA, Nadle J, Petit S, Gershman K, Ray S, et al. Invasive methicillin-resistant Staphylococcus aureus infections in the United States. JAMA. 2007;298: 1763–1771.

6. McBryde ES, Bradley LC, Whitby M, McElwain DLS. An investigation of contact transmission of methicillin-resistant Staphylococcus aureus. J Hosp Infect. 2004;58: 104–108.

7. PhD JAO, PhD SY, French GL, FRCPath. The Role Played by Contaminated Surfaces in the Transmission of Nosocomial Pathogens •. Infect Control Hosp Epidemiol. The University of Chicago Press on behalf of The Society for Healthcare Epidemiology of America; 2011;32: 687–699.

8. Dulon M, Haamann F, Peters C, Schablon A, Nienhaus A. MRSA prevalence in European

healthcare settings: a review. BMC Infect Dis. 2011;11: 138.

9. Kuehnert MJ, Kruszon-Moran D, Hill HA, McQuillan G, McAllister SK, Fosheim G, et al. Prevalence of Staphylococcus aureus Nasal Colonization in the United States, 2001–2002. J Infect Dis. 2006;193: 172–179.

10. Chambers HF, Deleo FR. Waves of resistance: Staphylococcus aureus in the antibiotic era. Nat Rev Microbiol. 2009;7: 629–641.

11. Melzer M, Welch C. Thirty-day mortality in UK patients with community-onset and hospital-acquired meticillin-susceptible Staphylococcus aureus bacteraemia. J Hosp Infect. 2013;84: 143–150.

12. MacNEAL WJ, Frisbee FC, McRAE MA. Staphylococcemia 1931-1940. Five hundred patients. Am J Clin Pathol. 1942;12.

13. Brown AF, Leech JM, Rogers TR, McLoughlin RM. Staphylococcus aureus Colonization: Modulation of Host Immune Response and Impact on Human Vaccine Design. Front Immunol. 2014;4: 507.

14. Sivaraman K, Venkataraman N, Cole AM. Staphylococcus aureus nasal carriage and its contributing factors. Future Microbiol. 2009;4: 999–1008.

15. van Belkum A, Verkaik NJ, de Vogel CP, Boelens HA, Verveer J, Nouwen JL, et al. Reclassification of Staphylococcus aureus nasal carriage types. J Infect Dis. 2009;199: 1820–1826.

16. Lamers RP, Stinnett JW, Muthukrishnan G, Parkinson CL, Cole AM. Evolutionary analyses of Staphylococcus aureus identify genetic relationships between nasal carriage and clinical isolates. PLoS One. 2011;6: e16426.

17. Cole AM, Tahk S, Oren A, Yoshioka D, Kim YH, Park A, et al. Determinants of Staphylococcus aureus nasal carriage. Clin Diagn Lab Immunol. 2001;8: 1064–1069.

18. Erdenizmenli M, Yapar N, Senger SS. Investigation of colonization with methicillin-resistant and methicillin-susceptible Staphylococcus aureus in an outpatient population in Turkey. Japanese journal of. nih.go.jp; 2004; Available: http://www0.nih.go.jp/JJID/57/172.pdf

19. Choi CS, Yin CS, Bakar AA, Sakewi Z. Nasal carriage of Staphylococcus aureus among healthy adults. , and infection= Wei …. europepmc.org; 2006; Available: http://europepmc.org/abstract/med/17164947

20. Wertheim H f. l., van Kleef M, Vos M c., Ott A, Verbrugh H a., Fokkens W. Nose Picking and Nasal Carriage of Staphylococcus aureus. Infect Control Hosp Epidemiol. [Cambridge

University Press, Society for Healthcare Epidemiology of America]; 2006;27: 863–867.

21. SKRAaMM I, Moen AEF, Bukholm G. Nasal carriage of Staphylococcus aureus: frequency and molecular diversity in a randomly sampled Norwegian community population. APMIS. Wiley Online Library; 2011;119: 522–528.

22. Mertz D, Frei R, Periat N, Zimmerli M, Battegay M, Flückiger U, et al. Exclusive Staphylococcus aureus throat carriage: at-risk populations. Arch Intern Med. 2009;169: 172–178.

23. Uemura E, Kakinohana S, Higa N, Toma C, Nakasone N. Comparative chracterization of Staphylococcus aureus isolates from throats and noses of healthy volunteers. Jpn J Infect Dis. 2004;57: 21–24.

24. Adesida SA, Abioye OA, Bamiro BS, Brai BIC, Smith SI, Amisu KO, et al. Associated risk factors and pulsed field gel electrophoresis of nasal isolates of Staphylococcus aureus from medical students in a tertiary hospital in Lagos, Nigeria. Braz J Infect Dis. 2007;11: 63–69.

25. Anwar MS, Jaffery G, Rehman Bhatti K-U-, Tayyib M, Bokhari SR. Staphylococcus aureus and MRSA nasal carriage in general population. J Coll Physicians Surg Pak. 2004;14: 661–664.

26. Vinodhkumaradithyaa A, Uma A, Shirivasan M, Ananthalakshmi I, Nallasivam P, Thirumalaikolundusubramanian P. Nasal carriage of methicillin-resistant Staphylococcus aureus among surgical unit staff. Jpn J Infect Dis. 2009;62: 228–229.

27. Balma-Mena A, Lara-Corrales I, Zeller J, Richardson S, McGavin MJ, Weinstein M, et al. Colonization with community-acquired methicillin-resistant Staphylococcus aureus in children with atopic dermatitis: a cross-sectional study. Int J Dermatol. Wiley Online Library; 2011;50: 682–688.

28. Severin JA, Lestari ES, Kuntaman K, Melles DC, Pastink M, Peeters JK, et al. Unusually high prevalence of panton-valentine leukocidin genes among methicillin-sensitive Staphylococcus aureus strains carried in the Indonesian population. J Clin Microbiol. 2008;46: 1989–1995.

29. Al-Rawahi GN, Schreader AG, Porter SD, Roscoe DL, Gustafson R, Bryce EA. Methicillin-resistant Staphylococcus aureus nasal carriage among injection drug users: six years later. J Clin Microbiol. 2008;46: 477–479.

30. Chacko J, Kuruvila M, Bhat GK. Factors affecting the nasal carriage of methicillin-resistant Staphylococcus aureus in human immunodeficiency virus-infected patients. Indian J Med Microbiol. 2009;27: 146–148.

31. Sollid JUE, Furberg AS, Hanssen AM, Johannessen M. Staphylococcus aureus:

determinants of human carriage. Infect Genet Evol. 2014;21: 531–541.

32. Williams RE. Healthy carriage of Staphylococcus aureus: its prevalence and importance. Bacteriol Rev. 1963;27: 56–71.

33. Guinan ME, Dan BB, Guidotti RJ, Reingold AL, Schmid GP, Bettoli EJ, et al. Vaginal colonization with Staphylococcus aureus in healthy women: a review of four studies. Ann Intern Med. 1982;96: 944–947.

34. Nilsson P, Ripa T. Staphylococcus aureus throat colonization is more frequent than colonization in the anterior nares. J Clin Microbiol. 2006;44: 3334–3339.

35. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486: 207–214.

36. von Eiff C, Becker K, Machka K, Stammer H, Peters G. Nasal carriage as a source of Staphylococcus aureus bacteremia. Study Group. N Engl J Med. 2001;344: 11–16.

37. Kluytmans J, van Belkum A, Verbrugh H. Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. Clin Microbiol Rev. 1997;10: 505–520.

38. Long SW, Beres SB, Olsen RJ, Musser JM. Absence of Patient-to-Patient Intrahospital Transmission of Staphylococcus aureus as Determined by Whole-Genome Sequencing. MBio. 2014;5. doi:10.1128/mBio.01692-14

39. Joseph SJ, Read TD. Bacterial population genomics and infectious disease diagnostics. Trends Biotechnol. 2010;28: 611–618.

40. Rinsky JL, Maya N, Steve W, Devon H, Dothula B, Price LB, et al. Livestock-Associated Methicillin and Multidrug Resistant Staphylococcus aureus Is Present among Industrial, Not Antibiotic-Free Livestock Operation Workers in North Carolina. PLoS One. 2013;8: e67641.

41. Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, et al. Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock. MBio. 2012;3. doi:10.1128/mBio.00305-11

42. Paterson GK, Harrison EM, Murray GGR, Welch JJ, Warland JH, Holden MTG, et al. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. Nat Commun. 2015;6: 6560.

43. Liu CM, Price LB, Hungate BA, Abraham AG, Larsen LA, Christensen K, et al. Staphylococcus aureus and the ecology of the nasal microbiome. Science Advances. American Association for the Advancement of Science; 2015;1: e1400216.

44. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004;186: 1518–1530.

45. Nübel U, Roumagnac P, Feldkamp M, Song J-H, Ko KS, Huang Y-C, et al. Frequent emergence and limited geographic dispersal of methicillin-resistant Staphylococcus aureus. Proc Natl Acad Sci U S A. 2008;105: 14130–14135.

46. Kobayashi SD, Musser JM, DeLeo FR. Genomic analysis of the emergence of vancomycin-resistant Staphylococcus aureus. MBio. 2012;3. doi:10.1128/mBio.00170-12

47. Mainous AG, Hueston WJ, Everett CJ. Nasal carriage of Staphylococcus aureus and methicillin-resistant S aureus in the United States, 2001–2002. The Annals of Family. Annals Family Med; 2006; Available: http://www.annfammed.org/content/4/2/132.short

48. Andersen PS, Pedersen JK, Fode P, Skov RL, Fowler VG Jr, Stegger M, et al. Influence of host genetics and environment on nasal carriage of staphylococcus aureus in danish middle-aged and elderly twins. J Infect Dis. 2012;206: 1178–1184.

49. Joseph SJ, Li B, Petit RA, Qin Z, Darrow L, Read TD. The single-species metagenome: subtyping Staphylococcus aureus core genome sequences from shotgun metagenomic data [Internet]. bioRxiv. 2015. p. 030692. doi:10.1101/030692

50. Joseph SJ, Li B, Ghonasgi T, Haase CP, Qin ZS, Dean D, et al. Direct Amplification, Sequencing and Profiling of Chlamydia trachomatis Strains in Single and Mixed Infection Clinical Samples. PLoS One. 2014;9: e99290.

51. R Core Team. The R project for statistical computing. R Foundation for Statistical Computing web-site www R-project org Accessed June. 2014;9.

52. Human Microbiome Project Consortium. A framework for human microbiome research. Nature. 2012;486: 215–221.

53. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760.

54. Aagaard K, Petrosino J, Keitel W, Watson M, Katancik J, Garcia N, et al. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. FASEB J. 2013;27: 1012–1022.

55. Wertheim HFL, Melles DC, Vos MC, van Leeuwen W, van Belkum A, Verbrugh HA, et al. The role of nasal carriage in Staphylococcus aureus infections. Lancet Infect Dis. 2005;5: 751–762.

56. Olsen K, Falch BM, Danielsen K, Johannessen M, Ericson Sollid JU, Thune I, et al.

Staphylococcus aureus nasal carriage is associated with serum 25-hydroxyvitamin D levels, gender and smoking status. The Tromsø Staph and Skin Study. Eur J Clin Microbiol Infect Dis. 2012;31: 465–473.

57. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. Nature. 2014;509: 357–360.

58. Foxman B, Martin ET. Use of the Microbiome in the Practice of Epidemiology: A Primer on -Omic Technologies. Am J Epidemiol. 2015;182: 1–8.

Tables

Table 1. Exposure variables assessed in this study, including the demographics and characteristics of the participants

Exposure Variable	Number present in each category (Total N=840)
Main Body Site	
Airways	132
Oral	658
Skin	50
Missing	0
Diet	
Meat/fish/poultry at least 3 days per week	767
Meat/fish/poultry at least 1 day but not more than 2 days per week	14
Eggs/cheese/other dairy products, but no meat/fish/poultry	37
missing	22
Gender	
Male	456
Female	384
missing	0
Age	
18 - 23 years	181
23 -28 years	398
28 - 33 years	150
33 - 38 years	69
>= 38 years	42
missing	0
Breast Fed or Not	
Yes	536
No	158
Don't know/remember	124
missing	22
Tobacco Use	
Yes	61
No	779
missing	0

Have Health Insurance or Not	
Yes	741
No	59
missing	40
BMI	
< 22	206
22 - 25	271
> 25	363
missing	0
Race	
Hispanic	45
Asian	65
Black	51
White	679
missing	0
Whether undergone any type of surgery	
Yes	32
No	808
Missing	0

Table 2. *S. aureus* positive HMP body sites based on reads mapping to the ancestral *S. aureus* genome sequence. Percentages based on total number of samples for that body site.

Body site	**Total number of samples**	**Number of *S. aureus* positive samples**	**Number samples with > 0.025X *S. aureus* coverage**
Anterior nares	137	78(57%)	68(50%)
Attached keratinized Gingiva	14	4(29%)	4(29%)
Buccal mucosa	185	56(31%)	56(30%)
Hard Palate	1	1(100%)	1(100%)
Left retroauricular crease	23	23(100%)	23(100%)
Palatine tonsils	19	7(37%)	6(32%)
Posterior fornix	108	11(10.20%)	9()
Right Antecubital fossa	1	1(100%)	1(100%)
Right retroauricular crease	31	28(90%)	28(90%)
Saliva	7	2(28.57%)	1(14%)
Stool	251	14(6%)	7(0.3%)
Subgingival plaque	19	1(5%)	1(5%)
Supragingival plaque	210	65(31%)	37(18%)
Tongue dorsum	221	82(37%)	82(37%)

Table 3. Estimated adjusted odds ratios with 95% confidence interval for models with binary outcome (presence/absence of any *S. aureus*) as well as multinomial outcomes (presence of strain-specific *S. aureus* vs. no detection*)* .

Exposure Variable	***Staphylococus aureus*** **Present/Not Present**	**p-value**	**Presence of *Staphylococus aureus* CC type**			
			CC8	**CC30**	**Other CC types**	
	OR (95% CI)		**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**	**p-value**
Main Body Site		<0.0001				<0.0001
Airways vs. Oral	3.3 (2.2 - 5.0)		2.7 (1.3 - 5.4)	2.5 (1.3 - 5.8)	4.6 (2.7 -7.7)	
Airways vs. Skin	0.1 (0.0 - 0.3)		0.2 (0.0 - 0.8)	0.1 (0.0 - 0.6)	0.0 (0.0 - 0.2)	
Oral vs. Skin	0.0 (0.0 - 0.0)		0.0 (0.0 - 0.3)	0.0 (0.0 - 0.2)	0.0 (0.0 - 0.1)	
Diet		0.4675				0.3228
Meat/fish/poultry at least 3 days per week	1.4 (0.6 - 3.0)		0.4 (0.1 - 1.5)	2.0 (0.4 - 10.4)	4.0 (0.7 - 21.0)	
Meat/fish/poultry at least 1 day but not more than 2 days per week	2.3 (0.6 - 8.9)		1.6 (0.1 - 15.9)	3.1 (0.3 - 31.9)	5.4 (0.5 - 62.5)	
Eggs/cheese/other dairy products, but no meat/fish/poultry (Reference)	1		1	1	1	
Gender		0.6434				0.776
Male (Reference)	1		1	1	1	
Female	0.9 (0.6-1.3)		1.0 (0.5 - 2.1)	0.7 (0.4 - 1.4)	0.9 (0.5 - 1.6)	
Age		0.7197				0.8055
3 years of age difference	1.0 (0.9 - 1.1)		1.1 (0.9 - 1.3)	1.1 (0.9 - 1.3)	1.0 (0.8 - 1.1)	

Breast Fed or Not		0.6527				0.8913
Yes (reference)	1		1	1	1	
No	1.0 (0.7 - 1.6)		0.7 (0.3 - 1.8)	1.0 (0.5 - 2.1)	1.3 (0.7 - 2.5)	
Don't know/remember	0.8 (0.5 - 1.3)		0.6 (0.2 - 1.6)	0.9 (0.4 - 1.9)	0.9 (0.4 - 1.9)	
Tobacco Use		0.7522				0.7173
Yes	0.9 (0.4 - 1.8)		1.2 (0.3 - 5.4)	1.4 (0.5 - 3.8)	0.6 (0.2 - 2.0)	
No (Reference)	1		1	1	1	
Have Health Insurance or Not		0.0525				0.403
Yes (Reference)	1		1	1	1	
No	0.5 (0.2 - 1.0)		0.0 (0.0 - 13.0)	0.9 (0.3 - 2.6)	0.5 (0.2 - 1.5)	
BMI		0.0276				0.0251
< 22	1.1 (0.7 - 1.9)		0.4 (0.1 - 1.1)	1.6 (0.6 - 3.8)	1.7 (0.7 - 4.0)	
22 - 25	1		1	1	1	
> 25	1.7 (1.1 - 2.5)		1.4 (0.6 - 3.0)	1.1 (0.5 - 2.2)	2.4 (1.3 - 4.5)	
RACE		0.2815				0.2791
Hispanic	2.0 (0.9 - 4.1)		4.7 (1.2 - 18.9)	0.5 (0.1 - 2.8)	2.3 (0.7 - 6.9)	
Asian	1.3 (0.7 - 2.3)		2.6 (1.0 - 7.3)	1.1 (0.4 - 3.2)	1.1 (0.4 - 3.0)	
Black	1.2 (0.5 - 2.5)		1.0 (0.2 - 5.6)	1.9 (0.6 - 6.3)	0.9 (0.2 - 3.1)	
White (Reference)	1		1	1	1	
Whether undergone any type of surgery		0.9241				
Yes	1.0 (0.4 - 2.5)		2.6 (0.4 - 18.7)	0.0 (0.0 - 172.8)	1.6 (0.4 - 6.5)	0.5205
No (Reference)	1					

Table 4. Estimated unadjusted odds ratios with 95% confidence interval for models with binary outcome (presence/absence of any *S. aureus*) as well as multinomial outcomes (presence of strain-specific *S. aureus* vs. no detection*)* .

Exposure Variable	*Staphylococus aureus* Present/Not Present	p-value	Presence of *Staphylococus aureus* CC type			
			CC8	CC30	Other CC types	
	OR (95% CI)		OR (95% CI)	OR (95% CI)	OR (95% CI)	p-value
Main Body Site		<0.0001				<0.0001
Airways vs. Oral	3.5 (2.3 - 5.3)		2.8 (1.4 - 5.7)	2.5 (1.3 - 4.8)	4.8 (2.8 -8.0)	
Airways vs. Skin	0.1 (0.0 - 0.3)		0.2 (0.0 - 1.0)	0.1 (0.0 - 0.6)	0.0 (0.0 - 0.2)	
Oral vs. Skin	0.0 (0.0 - 0.0)		0.1 (0.0 - 0.3)	0.1 (0.0 - 0.2)	0.0 (0.0 - 0.1)	
Diet		0.5938				0.3998
Meat/fish/poultry at least 3 days per week	1.4 (0.6 - 3.2)		0.4 (0.1 - 1.8)	1.8 (0.4 - 8.7)	4.2 (0.8 - 22.2)	
Meat/fish/poultry at least 1 day but not more than 2 days per week	2.1 (0.5 - 9.5)		1.4 (0.1 - 18.0)	2.9 (0.3 - 30.7)	4.8 (0.4 - 58.0)	
Eggs/cheese/other dairy products, but no meat/fish/poultry (Reference)	1.0		1.0	1.0	1.0	
Gender		0.4345				0.8758
Male (Reference)	1		1.0	1.0	1.0	
Female	0.9 (0.6-1.2)		0.8 (0.4 - 1.7)	0.8 (0.5 - 1.5)	0.9 (0.6 - 1.5)	
Age		0.7197				0.8061
3 years of age difference	1.1 (1.1 - 1.2)		1.1 (0.9 - 1.3)	1.1 (0.9 - 1.2)	1.0 (0.9 - 1.2)	

Breast Fed or Not		0.6527				0.8913
Yes (reference)	1.0		1.0	1.0	1.0	
No	1.0 (0.7 - 1.6)		0.6 (0.2 - 1.7)	0.9 (0.5 - 2.0)	1.2 (0.7 - 2.3)	
Don't know/remember	1.0 (0.6 - 1.6)		0.8 (0.3 - 2.3)	1.0 (0.5 - 2.2)	0.9 (0.4 - 1.9)	
Tobacco Use		0.5167				0.5927
Yes	0.8 (0.4 - 1.7)		0.9 (0.2 - 4.4)	1.4 (0.5 - 4.0)	0.5 (0.1 - 1.7)	
No (Reference)	1.0		1.0	1.0	1.0	
Have Health Insurance or Not		0.0373				0.7749
Yes (Reference)	1.0		1.0	1.0	1.0	
No	2.0 (1.0 - 3.7)		0.0 (0.0 - 13.0)	1.3 (0.4 - 3.9)	1.6 (0.6 - 4.5)	
BMI		0.0744				0.3618
< 22	1.2 (0.8 - 1.9)		0.8 (0.3 - 2.3)	1.2 (0.6 - 2.5)	1.4 (0.7 - 2.8)	
22 - 25	1.0		1.0	1.0	1.0	
> 25	1.6 (1.1 - 2.3)		1.5 (0.7 - 3.6)	1.1 (0.6 - 2.1)	1.9 (1.1 - 3.3)	
RACE		0.0560				0.3284
Hispanic	2.3 (1.2 - 4.2)		3.3 (0.7 - 15.2)	0.7 (0.1 – 3.6)	3.1 (1.1 - 8.8)	
Asian	1.3 (0.7 - 2.1)		2.4 (0.8 - 7.6)	1.1 (0.4 - 3.0)	1.0 (0.4 - 2.4)	
Black	1.2 (0.6 - 2.4)		0.8 (0.2 - 5.5)	1.8 (0.6 - 6.0)	1.0 (0.3 - 3.4)	
White (Reference)	1.0		1.0	1.0	1.0	
Whether undergone any type of surgery		0.5144				0.9695
Yes	0.7 (0.3 - 1.9)		0.7 (0.1 - 6.3)	inf	1.3 (0.3 - 4.9)	
No (Reference)	1.0		1.0	1.0	1.0	

Figures

Figure 1. *S. aureus* subtypes in HMP samples. 321 samples from the HMP project with a *S. aureus* core coverage > 0.025X project were classified using binstrain with the v2 matrix. The figure shows counts of the number of samples with each subtype present with beta > 0.2.