Evolutionary Relationships Among Staphylococci And The Prevention Of Staphylococcus Aureus Nasal Colonization

2011

Ryan Paul Lamers
University of Central Florida

EVOLUTIONARY RELATIONSHIPS AMONG STAPHYLOCOCCI AND THE PREVENTION OF *STAPHYLOCOCCUS AUREUS* NASAL COLONIZATION

by

RYAN PAUL LAMERS
Honours B.Sc. Laurentian University, Canada, 2004
M.Sc. Lakehead University, Canada, 2006
M.S. University of Central Florida, United States of America, 2010

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Burnett School of Biomedical Sciences
in the College of Graduate Studies
at the University of Central Florida
Orlando, Florida

Fall Term
2011

Co-Major Professor: Alexander M. Cole, Ph.D.
Co-Major Professor: Christopher L. Parkinson, Ph.D.

ABSTRACT

Staphylococcus is a significant cause of human infection and mortality, worldwide. Currently, there are greater than 60 taxa within *Staphylococcus,* and nearly all are pathogenic. The collective potential for virulence among species of *Staphylococcus* heightens the overall clinical significance of this genus and argues for a thorough understanding of the evolutionary relationships among species. Within *Staphylococcus, aureus* is the most common cause of human infection, where nasal carriage of this bacterium is a known risk factor for autoinfection. The predisposition to infection by nasal carriers of *S. aureus*, and the ease with which strains are transferred between individuals, suggests that nasal carriage is a major vector for the transmission of virulent strains throughout the community. This hypothesis, however, has not been assessed in any great detail to identify the genetic relationships between clinical isolates of *S. aureus* and those strains being carried asymptomatically throughout the community. Also lacking within this field is a unified and robust estimate of phylogeny among species of *Staphylococcus*.

Here, we report on a highly unified species phylogeny for *Staphylococcus* that has been derived using multilocus nucleotide data under multiple Bayesian and maximum likelihood approaches. Our findings are in general agreement with previous reports of the staphylococcal phylogeny, although we identify multiple previously unreported relationships. Regardless of methodology, strong nodal support and high topological agreement was observed with only minor variations in results between methods. Based on our phylogenetic estimates, we propose that *Staphylococcus* species can be evolutionarily clustered into 15 groups, and six species groups. In addition, our more defined phylogenetic analyses of *S. aureus* revealed strong genetic

associations between both nasal carriage strains and clinical isolates. Genetic analyses of hypervariable regions from virulence genes revealed that not only do clinically relevant strains belong to identical genetic lineages as the nasal carriage isolates, but they also exhibited 100% sequence similarity within these regions. Our findings indicate that strains of *S. aureus* being carried asymptomatically throughout the community via nasal colonization are genetically related to those responsible for high levels of infection and mortality.

Due to nasal carriage of *S. aureus* being a risk factor for autoinfection, standardized preoperative decolonization has become a major consideration for the prevention of nosocomial infection. Toward this end, we have identified the macrocyclic θ-defensin analogue RC-101 as a promising anti-*S. aureus* agent for nasal decolonization. RC-101 exhibited bactericidal effects against *S. aureus* in both epithelium-free systems, and *ex vivo* models containing human airway epithelia. Importantly, RC-101 exhibited potent anti-*S. aureus* activities against all strains tested, including USA300. Moreover, RC-101 significantly reduced the adherence, survival, and proliferation of *S. aureus* on human airway epithelia without any noted cellular toxicity or the induction of a proinflammatory response. Collectively, our findings identify RC-101 as a potential preventative of *S. aureus* nasal colonization.

ACKNOWLEDGMENTS

Many people have guided, supported, or otherwise assisted me throughout my dissertation, and for that I would like to express my sincerest gratitude. In particular, I would like to extend a special thank you to my mentor, Dr. Alexander Cole. I am grateful for being given the chance to succeed in his laboratory and I appreciate all of the direction, guidance, and advice that he has given me; he has been nothing short of an excellent and much needed mentor.

I would also like to express sincere thanks to my mentor, Dr. Christopher Parkinson, for his endless help and guidance throughout my dissertation. He has always been eager to provide assistance and direction, and never turned away an unsolicited early-morning discussion.

My dissertation committee members, Drs. Karl Chai and Sean Moore, have been of tremendous help throughout my dissertation. They routinely provided positive feedback and guidance that ultimately made this dissertation stronger. Their suggestions and questions stimulated thought-provoking discussion and provided many considerations for future studies.

I would like to thank my labmates, Colleen Eade, Matthew Wood, Gowrishankar Muthukrishnan, Vanathy Paramanandam, Todd Penberthy, Ph.D., Julie Martellini-Moore, Ph.D., Alana Persaud, Austin Ellis, and Camilla Diaz for their help, support, and motivation. The continual lab banter has been a much-welcomed release from the more serious nature of our day-to-day activities. I am very grateful to Amy Cole, Ph.D., Nicole Rogers, and Christine Chong for their help and all that they do to keep the laboratory functional; this is no easy task.

I extend a very special thank you to one of my closest friends, Tisha Choudhury. We have spent some of the most difficult days of our lives together, but always managed to end up smiling no matter what. I could not have made it without your optimism and friendship.

My family has been the single most motivating factor throughout the duration of my dissertation, and without them this accomplishment would not have been possible. My parents, Harry and Teresa Lamers, have provided me with endless support, encouragement and motivation, for which I will always be appreciative and grateful. Lastly, I would like to extend my most heartfelt appreciation to Shana Hayter. Her continual support and motivation have helped me through the best and worst of times. Her strength and understanding have far exceeded anything I could have ever asked for. For all that she has done, I am eternally grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. GENERAL INTRODUCTION

1.1 *Staphylococcus aureus* nasal colonization

Staphylococcus aureus was first identified in 1880 (145) and since then has become a progressively more virulent disease-causing agent throughout the world. *S. aureus* is a primary human pathogen associated with high levels of morbidity and mortality worldwide (204), being responsible for tens of thousands of deaths each year (99). *S. aureus* is currently one of the world's leading causes of nosocomial infection (98), and is notable for its adaptability and increasing resistance to antibiotic treatments (203). Infections caused by *S. aureus* range from mild to severe, resulting in superficial skin and soft tissue infections, septicemia, endocarditis, toxic shock syndrome (43), bacteremia, and pneumonia, among myriad others (147). Outside clinical settings, *S. aureus* remains a prevalent bacterium within community settings with approximately 50% of healthy individuals being colonized (122, 197, 203).

Colonization of healthy individuals typically occurs within the anterior nares; however, colonization also occurs on other surfaces of the body (203), presumably due to the ease with which bacteria are transferred over the body from the nose. Colonization of the nose primarily occurs among the moist squamous epithelium on the septum, adjacent to the nasal ostium (30, 153). Nasal carriage of *S. aureus* has been implicated in aiding *S. aureus* infection as the bacterial strain found within an individual's nose is frequently the same strain responsible for causing infection elsewhere in the body (200). Based on this fact, nasal carriage of *S. aureus* is of particular clinical concern since asymptomatic carriage appears to provide an important vector for the transmission of virulent strains throughout the community. Interestingly, however, the

extent to which nasal carriage strains present within the healthy population contribute to clinical infection, or the evolutionary similarities between nasal carrier strains and clinical isolates has not previously been well established (113, 129).

The carriage statuses of *S. aureus* within the human nares has historically been categorized into three classes: non-carriers, intermittent carriers, and persistent carriers (197). Non-carriers comprise approximately 50% of the general population (197, 203) and are those individuals who appear to never experience colonization of *S. aureus* within their nares. Intermittent carriers comprise approximately 30% of the general population (197, 203) and are those individuals who are at times observed to be colonized by *S. aureus*, but also undergo periods during which colonization is not observed. Persistent carriers comprise the remaining approximately 20% of the general population (197, 203) and are those individuals who have been observed to be colonized by *S. aureus* within their nares at all times. More recently, however, the nasal carriage classification system has been called into question with the observation that only persistent nasal carriers exhibit an increase in the levels of infection by their endogenous *S. aureus* strain. Conversely, intermittent carriers experience only a low level of infection by their endogenous nasal strain, with no greater predisposition to infection than non-carriers (140, 197). Moreover, the elimination kinetics of *S. aureus* experimentally inoculated in the nasal vestibule are similar between intermittent and non-carriers. The elimination kinetics for intermittent and non-carriers both, however, are significantly higher than those observed in persistent carriers. Patterns of anti-staphylococcal antibodies within intermittent and non-carriers are also similar to one another; however, they are significantly lower than the same antibodies in persistent carriers (197). In particular, immunoglobulin (Ig) G

isotypes directed against *S. aureus* toxic shock syndrome toxin-1 (TSST-1) and surface protein G were found to be significantly higher in persistent carriers as opposed to intermittent and non-carriers. Similarly, levels of IgA isotypes directed against TSST-1, staphylococcal enterotoxin A and clumping factor A were also observed to be significantly higher in persistent carriers as compared to intermittent and non-carriers (which exhibit no significant difference compared to each other). Based on these observations, it has been suggested that individuals be reclassified as "persistent" or "other" when describing the host's nasal carriage status (197).

Given that nasal carriage is a major risk factor for staphylococcal disease, the prevention of *S. aureus* nasal carriage is desirable in preventing transmission and infection (150). Due to the large number of strains that colonize healthy hosts and the growing number of antibiotics to which they are resistant, community-associated infections are relatively easy to transmit while becoming more difficult to combat. As such, the eradication of *S. aureus* from noses would be a key step in preventing and combating infection throughout the community. Difficulties in treating infection arise from the impressive adaptability and evolution of this bacterium, which results in an increasing number of antibiotics to which it is resistant (117).

The remarkable speed at which *S. aureus* develops antibiotic resistance is a result of horizontal gene transfer facilitating allelic fixation within a population very rapidly (65). Short doubling times and asexual reproduction also facilitate fixation of alleles conferring antibiotic resistance within *S. aureus* clonal populations (117). The fact that *S. aureus* can adapt quickly to novel chemotherapeutics complicates treating infection and as such, the need for novel treatments is continually increasing.

Identifying treatment options for *S. aureus* infections and, more generally, for the

prevention of nasal colonization would undoubtedly be made easier by determining the factors that are responsible for nasal carriage in humans. To date, a multitude of research has focused on nasal carriage of *S. aureus*; however, the precise determinants have yet to be conclusively identified. One certainty, however, is that nasal carriage is a multifactorial condition, involving an, as yet, unknown number of determinants from both the bacterium and the host (30, 139, 203).

1.2 Factors affecting nasal colonization of *Staphylococcus aureus*

A large number of host and bacterial factors have previously been identified as contributing to the nasal carriage of *S. aureus*. It is evident that no single determinant will be entirely responsible for carriage, but a variety of both host and bacterial factors working in sync are likely codeterminants of nasal carriage. It has previously been postulated that four main events are required for successful nasal carriage (203). The bacterium must come into contact with the nose, adhere to nasal epithelial cells via specific receptors, evade the host defense system; and must be able to proliferate within the nose. Of these events, adherence to nasal epithelia and evasion of the host immune response are presumably the two most important elements for successful nasal colonization. A multitude of bacterial factors must, therefore, be expressed by the colonizing pathogen to facilitate epithelial adhesion and evasion of the host's immune system. The host must also be permissive enough to allow for colonization to occur followed by proliferation of the bacterium. Thus, the complex interplay between bacterial and host factors dictate the carriage status of an individual.

1.2.1 Bacterial factors affecting nasal colonization

A variety of bacterial factors have been implicated in aiding the nasal colonization of *S.*

aureus. The exact means by which this pathogen evades the host immune system and proliferates within the nose is not known; however, *S. aureus* does possess a variety of known factors that presumably aid in both of these events. Primary host defense evasion strategies employed by *S. aureus* are the activities of immunoglobulin G (IgG)-binding proteins. Two such proteins, staphylococcal protein A (SpA) (195) and the staphylococcal binder of IgG protein (Sbi) (209), have been identified in *S. aureus* as binding host IgG for the suppression of the host immune response, thereby evading phagocytosis and complement fixation (4, 209). The first discovered and best characterized IgG-binding protein is SpA. Protein A is membrane exposed and interacts with both IgG Fc receptors and Fab fragments (4). Through these interactions, SpA prevents IgG from interacting with neutrophils, which would otherwise lead to the phagocytosis of *S. aureus*. When bound to the Fab regions of IgG, SpA also induces a superantigen response, promoting the activation and depletion of the B-cell population within the host (4, 174). Interestingly, the binding of SpA to Fc and Fab portions of IgG are independent of one another and non-competitive. Thus, a single SpA can bind simultaneously to each of these regions without causing interference with the other, independently bound SpA molecules (159).

Sbi acts in a slightly different manner than SpA, although both function by binding to IgG to thwart host defense mechanisms (209). Sbi is a secreted protein that is active towards a variety of different effectors of host defense. As with SpA, Sbi binds to Fc regions of IgG; however, it does not bind to the Fab fragments. Sbi has two immunoglobulin binding motifs with homology to the immunoglobulin binding domains of SpA, but apart from that, exhibits no homology to any other known protein (15). In addition to binding Fc regions of IgG, Sbi also activates the alternative pathway of the complement system where it interacts with, and prevents

the activation of complement component, C3 (15). Collectively, the activities of both IgG-binding proteins expressed by *S. aureus* are responsible in large part for successful immune evasion capabilities.

In addition to evading the host immune response via IgG binding, proteins involved in the evasion of neutrophil-mediated phagocytosis are also known (176). Among these are Sbi, clumping factor A (ClfA), capsular polysaccharide, and iron regulated surface determinant protein IsdH. While these four act in different capacities to circumvent phagocytosis, all act to alter and evade the complement system by preventing opsonization and subsequent engulfment (176).

Concurrent to the initial immune evasion events, invading *S. aureus* must also adhere to nasal epithelial cells to successfully proliferate within the host. Myriad protein families have been implicated in *S. aureus* adhesion to eukaryotic cells, many of which are known as microbial surface components recognizing adhesive matrix molecules (MSCRAMMs). MSCRAMMs are surface exposed proteins that are anchored to the bacterium and interact with one or more known host ligands (149). In general, MSCRAMMs contain a number of different domains; however, they typically possess a signal sequence, one or more ligand-binding domains, a repetitive wall-spanning domain, a non-repetitive wall spanning domain, a membrane spanning domain, and a positively charged C-terminal LPXTG anchoring domain (149). While the functions of most domains within MSCRAMMs are known, the function of the repetitive wall-spanning domain is incompletely understood. This domain is hypothesized to serve as a stalk to extent the ligand-binding domain from the bacterium to the host, while another untested possibility is that the repeat domain length is directly correlated to bacterial wall thickness.

Among the MSCRAMMs of *S. aureus* are two primary virulence gene families, the clumping factor proteins (ClfA and ClfB), and the fibronectin binding proteins (FnbA, and FnbB) that derive notable significance due to their roles in human epithelial attachment, invasion, and subsequent virulence (22). Due to their adhesive properties, the clumping factor and fibronectin binding proteins are promising candidates for nasal colonization of *S. aureus*. Clumping factors are surface proteins that have been found to bind fibrinogen and promote adhesion to desquamated epithelial cells, and the nares' of both mice and humans. *clfB* has previously been observed essential for nasal adhesion in an *in vivo* model, while *clfA* has not been tested; however, in another study, *clfA* was identified as being essential for infection morbidity and mortality in mice models (36, 92, 168, 204). Fibronectin binding proteins (FnbA and FnbB) adhere to both fibronectin and fibrinogen (206), and have been shown to play a critical role in *S. aureus* adhesion to mammalian cells (63, 175). Adhesion to epithelia by fibronectin binding proteins is accomplished through the interaction between the ligand binding domain and the host fibronectin, which in turn interacts with its receptor, integrin (175).

In addition to the actions of bacterial effectors of nasal colonization, host factors also undoubtedly play a significant role in allowing or preventing carriage. While the exact means by which host factors limit nasal colonization of *S. aureus* is not known, a number of mechanisms are known to continuously function to prevent its occurrence.

1.2.2 Host factors affecting nasal colonization

A variety of host factors affect the ability of *S. aureus* to colonize human noses. Such factors are physical barriers, receptors that recognize the pathogen, and innate immune cells; all of which act to protect the host from invasion and colonization (153). Nasal secretions have a

prominent role in host defense and contain a variety of factors involved in both adaptive and innate immune systems to combat bacterial infection (27). Immunoglobulin A (IgA) and G (IgG) are major mediators of the adaptive immune system secreted by nasal epithelial cells. Mediators of the innate immune system in nasal secretions include: lysozyme, lactoferrin, uric acid, peroxidase, secretory leukoprotease inhibitor, defensins, and other antimicrobial peptides (27, 93); all of which exhibit differing antimicrobial and anti-staphylococcal activities. Importantly, most strains of *S. aureus* are resistant to lysozyme and lactoferrin (14) and of the defensins present within nasal secretions, only human beta-defensin 3 (hBD-3) has previously been reported as being anti-staphylococcal (128).

Detection of *S. aureus* in the nose is mediated by toll-like receptors (TLRs) (69); prominent pattern recognition molecules of the innate immune system that recognize common microbial motifs (52). The primary TLR responsible for detecting *S. aureus* in this milieu is TLR2 (69, 183). Interestingly, the expression of TLR2, as well as the antimicrobial activity of hBD-3, have been observed to be suppressed in the nasal carriage state by colonizing strains of *S. aureus* (153). Additionally, methicillin-resistant *S. aureus* (MRSA) strains are more resistant to hBD-3 than methicillin-sensitive *S. aureus* (MSSA) strains (128), adding to the likelihood that the strains more resistant to innate defense mechanisms are also the ones resistant to antibiotic treatments.

1.3 Antibiotic treatments and chemotherapies against *Staphylococcus aureus*

1.3.1 General therapies against *Staphylococcus aureus*

Myriad anti-*S. aureus* therapeutics have been utilized to treat *S. aureus* infection; however, resistance to all of these has been observed (74, 164). The first therapeutic strategies to be used against *S. aureus* were penicillin class β-lactam antibiotics. Penicillin and methicillin were the first anti-*S. aureus* agents employed; however, resistance to these rapidly emerged. Current estimates suggest that 90-95% of all *S. aureus* strains are resistant to penicillin and approximately ~60% are resistant to methicillin (164). β-lactam antibiotic resistance stems from the presence of the *mecA* gene, which is located on the 21-67 kilobase (kb) mobile genetic element referred to as the staphylococcal cassette chromosome *mec* (SCC*mec*) (90, 94).

The SCC*mec* element was acquired in *S. aureus* by unknown mechanisms; however, it is thought to be the product of a horizontal gene transfer event from a distantly related species (47, 74). Interestingly, SCC*mec* does not contain phage-related genes, virulence genes, or transposases; however, it does contain two recombinases (90). Once present in the bacterium SCC*mec* becomes inserted into the *S. aureus* genome near the origin of replication within a gene of unknown function (74). The *mecA* gene encodes an alternative penicillin binding protein (PBP) as compared to the endogenous PBP encoded by the *S. aureus* genome (126), and conveys resistance to methicillin. In methicillin susceptible *S. aureus* (MSSA), β-lactams bind to the endogenous PBP located in the bacterial cell wall and disrupt synthesis of the peptidoglycan layer, thereby killing the bacterium. In *S. aureus* containing *mecA*, the alternate PBP (PBP2)

binds β-lactams thereby sequestering these molecules and preventing their disruption of peptidoglycan synthesis (42).

Treatment of β-lactam resistant *S. aureus* (also referred to as methicillin or multidrug resistant *S. aureus*; MRSA) is typically achieved using the glycopeptide antibiotic, vancomycin. Vancomycin is typically used as a last resort for treating *S. aureus* and was first implemented in 1958 (81, 164). Resistance toward vancomycin was not observed until 1997 in Japan and 2002 in the United States (81). Since that time, resistance has not become widespread; however, increases in minimum inhibitory concentrations (MICs) have been observed. Interestingly, minor increases in MIC appear to have heightened effects on therapeutic outcome with minor increases in MIC reducing the efficacy of vancomycin even though the MIC remains within the susceptible range (81).

Vancomycin functions by binding to the D-alanyl-D-alanine terminus of the growing peptidoglycan layer, thereby preventing cross linking from occurring (115). In enterococci, resistance is caused by the presence of the *van* gene cluster. These genes encode enzymes that produce altered peptidoglycan precursors (e.g. D-alanyl-D-lactate or D-alanyl-D-ser) that have much lower affinity for vancomycin (164). While some strains of *S. aureus* have acquired genes from the *van* cluster (21), resistance toward vancomycin appears primarily a novel mechanisms resulting in modifications to the cell wall morphology (201). While different vancomycin intermediate or resistant strains exhibit differences in cell wall morphology, an overall thickening of the cell wall and an upregulation of D-alanyl-D-alanine precursors are the most common phenotypes (81). The mechanism of cell wall thickening is not currently known; however, it appears that the thickened wall hampers the diffusion of vancomycin to its active site

at the location of cell wall biosynthesis (81, 173). Vancomycin and other such therapeutics used for *S. aureus* infection are frequently issued intravenously. This is due to the metabolites from these molecules being ineffective against the bacterium. One other consideration is that many of these therapies, in particular vancomycin, are used as a last resort and thus, routine use is not recommended due to the potential for resistance to be acquired.

1.3.2 *Staphylococcus aureus* nasal decolonization using mupirocin

Nasal colonization of *S. aureus* increases the likelihood of autoinfection in postoperative patients and the immunocompromised (discussed in Section 1.1). As such a number of studies have detailed the benefits of nasal decolonization in clinical settings and the implementation of such a decolonization strategy in these settings is gaining increased attention. Various therapies exist for preventing *S. aureus* nasal colonization; however, none of which are more prominent than that of mupirocin ointment. Mupirocin ointment is typically applied to the nasal vestibule multiple times daily for multiple days prior to surgery. It is generally formulated as a 2% (w/w) nasal ointment and sold under the trade name, Bactroban® (GlaxoSmithKline, London, UK).

Mupirocin is a naturally occurring polyketide antimicrobial compound, synthesized by *Pseudomonas fluorescens*, originally identified as pseudomonic acid (55). It is an analogue of isoleucine (Ile) and is comprised of a pyran ring-containing monic acid joined to a 9-hydroxynonanoic acid (188). Antibiotic activity results from the irreversible binding of mupirocin to isoleucyl-transfer RNA synthetase (IleRS) which ultimately prevents protein synthesis (84, 148). When mupirocin is bound to IleRS, isoleucine-charged tRNA are not continually produced thus, depleted from the bacterium. Binding of mupirocin to IleRS occurs at two sites. The methyl terminus of the monic acid mimics the side chain of Ile and occupies the

Ile-binding site of IleRS while the pyran ring within mupirocin occupies the ATP-binding pocket of IleRS (133, 148, 188). Interestingly, the two amino acids on bacterial IleRS recognizing mupirocin are not present in eukaryotes (133).

The initial report of the anti-*S. aureus* activity of mupirocin in 1985 revealed a minimum inhibitory concentration (MIC) of 0.25 µg/mL (180); however, by the early 1990's resistance to this chemical therapy had emerged (34, 188). Currently, there are two levels of resistance for mupirocin in *S. aureus*. Low-level mupirocin resistance (LL-MR) is defined by a MIC of 8-256 µg/mL while high-level mupirocin resistance (HL-MR) is defined by a MIC ≥512 µg/mL (24). LL-MR is typically obtained with the acquisition of sporadic mutations among the IleRS gene, *ileS*; however, these mutations appear to only minimally affect the growth of the bacterium (87, 188). HL-MR among *S. aureus* has been attributed to the acquisition of plasmids containing a eukaryotic-like IleRS that possesses approximately 52% amino acid similarity and 30% nucleotide identity to the endogenous IleRS (75, 155, 188). The *mupA* gene encoding the alternative IleRS is highly conserved in *S. aureus*, and other staphylococci (32, 156); however, the remainder of the plasmid carrying this gene is variable in size, and genetic content (207).

Recombination of the *mupA* gene occurs between plasmids due to the presence of recombination sequences flanking the gene (131). Transfer of a plasmid containing *mupA* between *S. aureus* and *S. epidermidis* has been observed *in vivo* in the hospital setting (88) contributing to the spread of mupirocin resistance among *S. aureus* and other staphylococci. The evolutionary relationships between staphylococci are incompletely understood; however, the exchange of genetic material between staphylococcal species is expected to enhance therapeutic resistance. Thus, a detailed understanding of the species phylogeny of *Staphylococcus* is a

necessity for understanding better the pathogen-pathogen and host-pathogen interactions within this genus. Supporting this claim is the observation that *S. aureus* strains found to previously colonize only ungulates and poultry have now been observed to colonize humans as well (204).

1.4 Evolutionary relationships among *Staphylococcus* species

Currently, greater than 60 recognized staphylococcal taxa exist with many of these responsible for human disease. With the exception of *S. aureus* most taxa of clinical importance are coagulase-negative. Indeed, coagulase negative staphylococci (CoNS) have become an increasing concern due to their emerging association with human infection (91) and their heightened resistance to conventional therapeutics with the acquisition of the SCC*mec* element (67). Multidrug resistant CoNS (MR-CoNS) are commonly associated with nosocomial infections due to their biofilm formation on medical devices (56). Interestingly, while *S. aureus* is known to be a highly adaptive pathogen, CoNS are historically more resistant to antibiotics (91).

The distribution of antibiotic resistance and transfer of the SCC*mec* element is not currently well understood and the origin of SCC*mec* remains unknown; however, there is some evidence to suggest that *Macrococcus caseolyticus* is the parent donor (6). Transfer between related staphylococci is presumed to occur via horizontal gene transfer where recipient staphylococci then act as *SCCmec* donors to other species (67). Thus, a thorough understanding of the relationships between staphylococci is a necessity for understanding and preventing the spread of antibiotic resistant strains.

Species discovery within *Staphylococcus* is a frequent occurrence with seven new species

identified within the last year. Typically, when novel species are discovered, a number of biochemical tests are performed (53) along with an assessment of the 16S ribosomal RNA gene (16S rDNA), to estimate which species of *Staphylococcus* the new taxon is most closely related. Historically, this has been sufficient to yield a general understanding of the staphylococcal phylogeny; however, high sequence similarity among 16S rDNA hampers a more detailed description of evolutionary relationships within *Staphylococcus.* As such, more variable genes have been assessed and used to infer species relationships in *Staphylococcus*. Previous estimates of the staphylococcal phylogeny have relied upon simple and fast methods of phylogenetic inference (neighbor joining (NJ)) as opposed to more complex approaches such as Bayesian inference (BI) or maximum likelihood (ML). Using BI and ML methodologies, it is possible to use complex modeling strategies for phylogenetic inference where dataset partitioning (for example, by stem and loop regions of 16S rDNA, and codon position for protein coding genes) facilitates the use of more biologically relevant models of sequence evolution. NJ on the other hand, estimates phylogeny based on a model of minimum evolution and does not allow for assessment using alternative models of sequence evolution. As such, previous analyses of the staphylococcal phylogeny may be under estimating the degree to which evolutionary changes have occurred within this genus.

Within *Staphylococcus*, there are two primary lineages. The most ancestral of these comprise the oxidase-positive species of the *S. sciuri* group (182), which are not frequently associated with human infection. This group is also the most closely related to *Macrococcus caseolyticus*, a species previously belonging to *Staphylococcus*, but subsequently reclassified (100). Most previous studies have grouped the remaining staphylococcal species (i.e. oxidase-

negative) into between three and ten different lineages. Among these are the prominent lineages represented by the clinically significant species of *S. epidermidis, S. aureus, S. saprophyticus, S. simulans, S. intermedius,* and *S. hyicus* (103, 182). Within these groups, however, relationships have been variable depending on the locus analyzed, with no robust overall estimate between species. Additionally, previous studies have typically only included species of human importance, failing to include a more complete set of staphylococcal taxa including those species not observed to directly colonize humans. Thus, a more comprehensive estimation of species phylogeny within *Staphylococcus* is required for a fuller understanding of species evolution within this genus.

2. EVOLUTIONARY ANALYSES OF *STAPHYLOCOCCUS AUREUS* IDENTIFY GENETIC RELATIONSHIPS BETWEEN NASAL CARRIAGE AND CLINICAL ISOLATES

2.1 Introduction

Staphylococcus aureus is a prevalent human pathogen of increasing concern to public health worldwide. This pathogen is one of the leading causes of hospital-acquired infection, and additionally leads to significant levels of infection via community transmission. Approximately 20-30% of the global population is persistently colonized with *S. aureus* in the anterior nares, with 60-100% of individuals projected to be transiently colonized at some point during their lives (197). Though nasal carriage of *S. aureus* is hypothesized to be a major vector for transmission throughout hospitals and the community, neither the determinants of nasal colonization nor the role of carriage in the propagation of *S. aureus* infection throughout these settings are well established (129).

Multiple studies have shown that nasal carriage of *S. aureus* is a risk factor for pathogenic infection (104, 200), but just recently was it observed that heightened risk is only evident in persistent nasal carriers whereas intermittent and non-carriers exhibit low levels of infection (197). Aside from straightforward incidences in which endogenous strains establish pathogenic infections in their hosts, the overall extent to which nasal carriage strains are responsible for transmissible infection is not currently known.

Population structure and genetic diversity of *S. aureus* has been extensively studied in the past using pulse-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST) (46, 49, 60, 107, 127, 166). While PFGE provides adequate strain resolution, it encounters difficulty

in reproducing and comparing data between laboratories. Thus, MLST is the primary means by which *S. aureus* strains have been analyzed for the past decade. Yet, because of the slow rate of molecular evolution within MLST genes, this methodology is most useful on a global epidemiology scale (60, 106, 170). When local investigations are carried out, where a greater level of strain resolution is desired (e.g. local applications such as patient to patient transmission), analyses of hypervariable virulence genes is required (109). Moreover, recent interest in sub-classifying sequence types (STs) has also identified virulence genes (e.g. clumping factor and fibronectin binding protein gene families) as appropriate targets for obtaining high levels of strain resolution (60, 107). In addition to their hypervariability, virulence genes are also attractive targets for the assessment of strain pathogenicity since these genes contribute to the invasiveness of the bacterium.

Previous studies that have focused on virulence genes have typically done so in large cohorts of clinical strains where methodologies such as amplified fragment length polymorphism typing, *spa* typing, or double locus sequence typing have been employed (60, 106, 107, 109, 110, 203). Few studies have analyzed virulence genes to examine *S. aureus* within the community, or to identify the genetic relationships between nasal carriage isolates and those isolated from the clinical setting. Previous studies have identified that most *S. aureus* strains, both nasal and clinical, belong to five major clonal complexes (CCs); CC5, CC8, CC22, CC30, and CC45 (203); however, it is not well established whether the genes responsible for the pathogenicity of *S. aureus* are genetically similar between clinical and nasal carriage isolates.

Here we have performed evolutionary analyses on the seven MLST gene fragments, as well as the hypervariable regions of virulence genes in a cohort of *S. aureus* nasal carriage

strains to analyze the genetic diversity present therein. Contrary to previous reports, we observe higher levels of nucleotide diversity among nasal carriage strains than those for clinical isolates. In addition to analyzing the genetic diversity in our cohort of nasal carriage strains, we also performed a genetic comparison between these strains and strains of clinical significance. We find that both nasal carriage strains from our cohort and clinical strains isolated from symptomatic patients around the world exhibit the same genetic makeup in housekeeping and virulence genes.

2.2 Materials and Methods

2.2.1 Ethics statement for collection of nasal carriage isolates

Nasal carriage isolates of *S. aureus* were collected from willing donors following University of Central Florida Institutional Review Board (IRB)-approved procedures. Written informed consent was obtained for all donors throughout the study. All study coordinators involved in the sample collection process were IRB-approved with Collaborative Institutional Training Initiative (CITI) certification.

2.2.2 Bacterial isolates

Two hundred and twenty-two healthy individuals at the University of Central Florida (Orlando, Florida, USA) were prescreened for the presence of *S. aureus* in their nares. Of these, nasal carriage isolates were obtained from 56 (25.2%) individuals and utilized for genetic analyses in this study. Isolates were collected by inserting a single cotton swab into each of a donor's nostrils and circulating for approximately five to ten seconds. As part of an ongoing

longitudinal study, we obtained multiple samples from repeat donors, at a minimum interval of one month, to monitor the population genetics of *S. aureus* over time. Thus, our data are a reflection of one representative strain from all individuals involved in the current study, unless multiple samplings identified different strains from the same individual. In those cases, all different strains from one individual were analyzed. Nasal samples were plated on Trypticase™ Soy Agar (TSA) containing 5% sheep's blood (Becton, Dickinson and Company, Franklin Lakes, New Jersey, USA), and incubated at 37°C for 16 hours. Bacterial colonies were identified as *S. aureus* using Staphyloslide™ Latex Test reagent (Becton, Dickinson and Company, Franklin Lakes, New Jersey, USA), and positive colonies were inoculated in 5 mL of Trypticase Soy Broth and grown for 16 hours at 37°C and 250 rpm. Following inoculation, 1.5 mL of bacterial culture was pelleted by centrifugation for two minutes at 16 000 x g and culture medium was discarded. Pellets were then stored at -80°C until DNA isolation.

Twenty-eight clinical isolates of *S. aureus* were also utilized in this study to determine the evolutionary relationships between clinical strains and strains present in the nasal carriage population. Gene sequences from 15 clinical isolates with complete genomes available were obtained from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nucleotide/). The previously sequenced clinical strains were N315, Mu50, COL, MRSA252, MSSA476, MW2, USA300_FPR3757, NCTC8325, JH1, JH9, Newman, Mu3, USA300_TCH1516, 04-02981, and TW20 (Appendix A; Table A.1). Thirteen additional clinical strains for which clumping factor A (*clfA*), clumping factor B (*clfB*), and fibronectin binding protein A (*fnbA*) repeat region sequences are available on the NCBI nucleotide database were also utilized in this study

(Appendix A; Table A.1) (109). Refer to Appendix A, Table A.2 for accession numbers to all DNA sequences utilized in this study.

2.2.3 DNA isolation/amplification

S. aureus genomic DNA was isolated using GenElute™ Bacterial Genomic DNA kit (Sigma-Aldrich Co., St. Louis, Missouri, USA), according to the manufacturer's instructions. Following DNA isolation, extracts were quantified and stored at -20°C until DNA amplification.

Amplification of multilocus sequence typing (MLST) gene fragments was carried out using primers and protocols described previously (46). Briefly, 402-516 bp fragments for the seven MLST housekeeping genes (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL*) were amplified and sequenced (see below). Sequence types (STs) were determined for each strain based on the alleles identified at each of the seven loci using the *S. aureus* MLST database (http://www.mlst.net) (Appendix A; Table A.1). For instances in which new alleles, or combinations of alleles (i.e. new STs), were identified the MLST database curator was contacted and new allele numbers and STs were obtained.

For *clfA*, *clfB*, *fnbA*, and *fnbB*, the repeat-containing regions were chosen for molecular analysis within this study. Chromosomal DNA was amplified using primers and protocols previously described by Gomes *et al.* (60). All primers utilized in this study were synthesized by Integrated DNA Technologies, Inc. (Coralville, Iowa, USA). For PCR amplification, approximately 20-30 ng of template DNA was added to a 100 µL reaction containing 0.02 U/µL of Platinum® Taq DNA polymerase High Fidelity (Invitrogen Corporation, Carlsbad, California, USA), 1X PCR buffer (60 mM Tris-SO_4 (pH 8.0), 18 mM ammonium sulfate), 2 mM $MgSO_4$, 0.3 mM dNTPs, 0.3 µM of each primer, and 2% (v/v) dimethyl sulfoxide. PCR was conducted

using an iCycler™ thermal cycler (Bio-Rad Laboratories, Hercules, California, USA) with the following cycling parameters: 1 cycle of 5 min. at 95°C; 40 cycles of 30 sec. at 94°C, 60 sec. at annealing temperature (46, 60), 60 sec. at 72°C; 1 cycle of 10 min. at 72°C; hold at 4°C.

2.2.4 DNA sequencing

Following DNA amplification, PCR products were purified using isopropanol precipitation and subjected to Sanger sequencing (167) at The Florida State University DNA Sequencing Facility (Tallahassee, Florida, USA). Forward and reverse reads were generated for all amplicons and analyzed using BioEdit Sequence Alignment Editor (66) and MEGA 4.1 (184). While the use of these sequence analysis programs was sufficient for the *fnb* genes, additional DNA analysis tools were developed for the *clf* genes.

2.2.5 Sequence analysis of *clf* genes

To analyze the highly variable serine-aspartic acid (SD) repeat region of *clfA* and *clfB*, we developed sequence analysis software following that described by Koreen *et al.* (107). Briefly, the program analyzes a series of either *clfA* or *clfB* DNA sequences, beginning SD repeat profiling at the TCN-GAY (where N is any nucleotide and Y is either of the pyrimidines) in the first occurrence of GAT-TCN-GAY. The program then analyzes tandemly repeating blocks of 18 nucleotides (one repeat unit) unless nucleotides 13 to 15 are of the sequence TCN, in which case DNA strand-slippage had occurred and the previous 12 nucleotides are considered as one shortened repeat. Each unique repeat unit is then assigned a number, effectively converting the DNA sequence into a numeric profile. As described in (107), the analysis of *clfB* was terminated with the nucleotide immediately prior to the first occurrence of TCN-GAT-TCA-AGA. For *clfA*,

this is the first time such a program has been used in profiling the SD repeats, hence no prior termination sequence has previously been reported. Clumping factor A does not contain the same terminating sequence as *clfB*, and therefore, the program was modified to terminate the analysis with the nucleotide immediately prior to the first occurrence of TCN-AAC-AAT-AAT. Refer to Appendix A, sections A.1 and A.2 for the source code to both, *clfA* and *clfB* (respectively), SD repeat profiling programs.

Using this program, genes of interest were converted to a numeric profile based on the nucleotide sequence of each repeating unit, as well as the order and number of repeats. Therefore, two samples that share identical repeat profiles also share 100% nucleotide identity. The program assigns numbers to unique repeat sequences as they are encountered throughout the dataset, and as such, no inference can be made as to the percent nucleotide similarity between two different repeat numbers (e.g. repeat number one is not necessarily more similar to repeat number two than any other repeat number).

Following the generation of numeric *clf* repeat profiles, an additional program was generated to transform the numeric outputs to color-coded representations. As input, the plotting software uses the numeric *clf* repeat profiles along with a file containing hexadecimal color codes. All repeat units were assigned a uniquely colored box; therefore, all like-colored boxes are 100% identical in nucleotide sequence. Refer to Appendix A, section A.3 for the source code to the graphing software utilized in this study.

2.2.6 Phylogenetic reconstruction of MLST data

To determine the genetic relationship between nasal carriage isolates and those isolates of clinical origin, phylogenetic analyses of the concatenated MLST data were carried out for all

isolates analyzed in this study using the Metropolis-Hastings coupled Markov chain Monte Carlo method (BI)

in MrBayes v3.1.2 (83, 160). The concatenated MLST dataset was partitioned by locus with the nucleotide substitution model for each being determined using the Akaike Information Criterion (AIC) within jModelTest v0.1.1 (64, 151). For loci *arcC*, *glpF*, *pta*, and *yqiL* the K80 substitution model (95) was used. For locus *gmk*, the K80 plus Gamma substitution model was employed where Gamma indicates that, in addition to the substitution matrix determined by the model for specific nucleotide pairs, a gamma distribution was also applied to determine the overall substitution rate at each nucleotide site (11). For loci *aroE* and *tpi*, the SYM (210) plus Gamma substitution model was used in BI runs. Two independent BI runs were carried out using random starting trees with one cold chain and three heated chains. Each run consisted of 5 million generations with every 100 steps being sampled. As verified using Tracer v1.5 (157), stationarity was reached after 500 000 generations and a conservative burn-in of 1.25 million (25%) generations was performed.

2.2.7 Computational analyses of MLST data

Sequence types were assigned to groups using the eBURST v3 program (50) where all members of a group share six of seven identical loci with at least one other member of the group. Using eBURST to compare nasal carriage strains to clinical isolates, nasal carriage strains were treated as the reference set while clinical isolates were treated as the query set. To identify further relationships between isolates, a minimal spanning network of MLST data was generated using TCS v2.1 (23).

2.2.8 Statistical analyses of gene variability and evolution

Nucleotide diversities were determined for aligned DNA sequences using DnaSP v5 (116). Molecular evolutionary analyses, including codon-based Z-tests and dN/dS ratios, were conducted using MEGA 4.1 (184) under the Nei-Gojobori P-distance method (134) using 1 000 bootstrap replicates. Indices of discrimination were calculated for all loci using the Discriminatory Power Calculator (http://biophp.org/stats/discriminatory_power/demo.php), which is a modification of the Simpson's index of discrimination test (85, 86).

2.3 Results

2.3.1 Multilocus sequence typing reveals genetic associations between nasal carriage and clinical isolates

Multilocus sequence typing (MLST) of all 93 *S. aureus* strains analyzed in this study identified 34 different sequence types (STs). Among the 66 nasal carriage isolates, 26 different STs were observed, four of which were new. Additionally, three new alleles were also identified by this study, all at locus *tpi*.

Within the cohort of nasal carriage strains analyzed herein, ST30 was most prevalent, accounting for ~29% of all isolates. Sequence types 5 and 8 were also prevalent among the nasal carriage strains analyzed within this study, together accounting for ~20% of all isolates tested (Appendix A; Table A.1). While none of the clinical isolates analyzed in this study are of ST30, a combined ~32% belong to ST5 and ST8. The observation of nasal carriage and clinical isolates belonging to ST5 and ST8 is in agreement with previous reports in which both nasal carriage and clinical isolates belong to these same major clusters (49, 50, 203). Interestingly, at

the ST level, over half of the clinical isolates analyzed in this study (~54%) belonged to STs (such as ST105 and ST239) that do not contain nasal carriage strains (Appendix A; Table A.1). However, phylogenetic analyses of concatenated STs of all strains in this study revealed a close relationship among both nasal carriage and clinical isolates of *S. aureus* (Figure 2.1). As can be seen in Figure 2.1, the vast majority of clades containing clinical isolates (strain names in red text) also contain nasal carriage strains from the cohort analyzed in this study (strain names in black text).

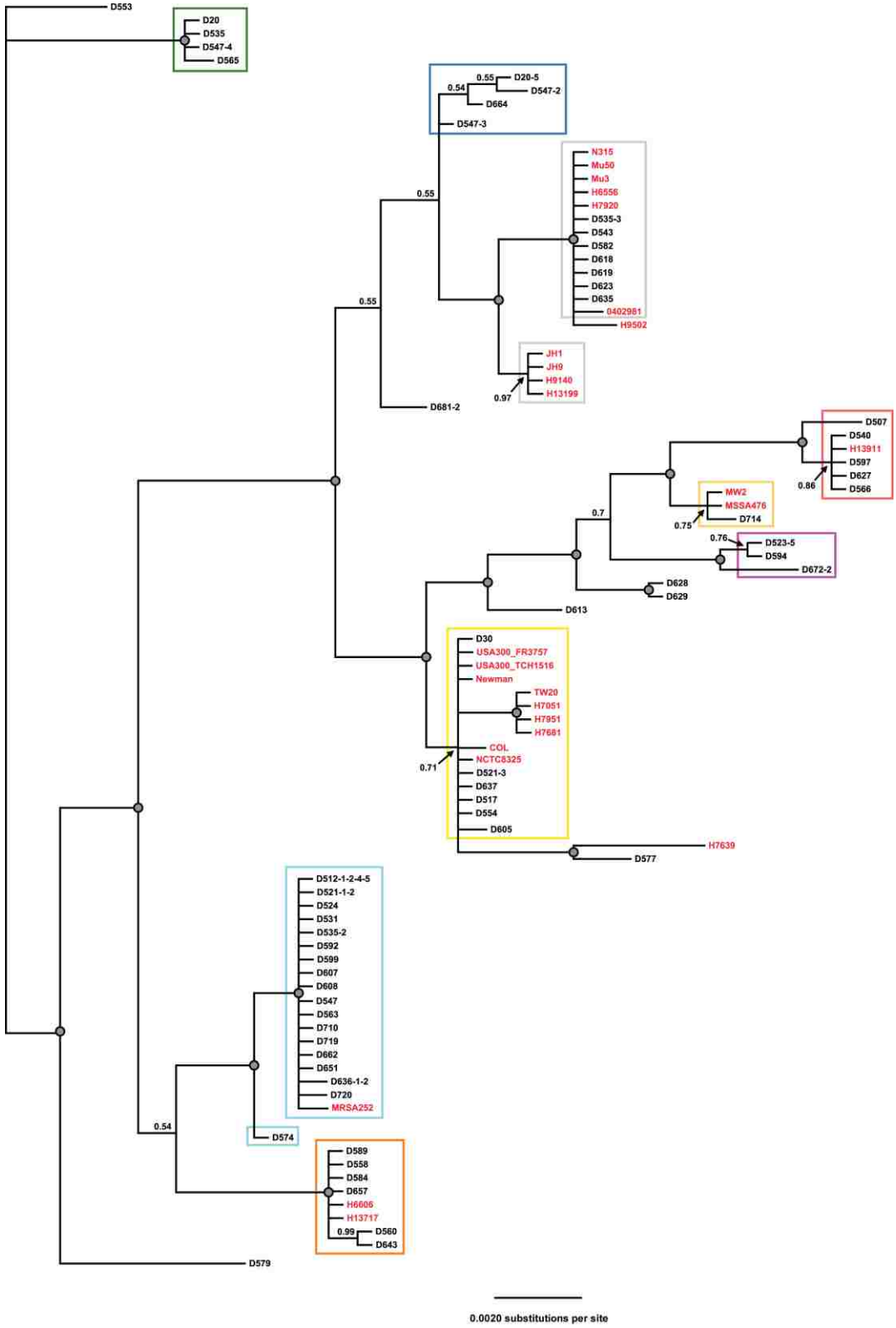
D553
D20
D535
D547-4
D565
0.55
D20-5
0.54
D547-2
D664
D547-3
N315
Mu50
Mu3
H6556
H7920
D535-3
D543
D582
D618
D619
D623
D635
0402981
H9502
0.55
0.55
JH1
JH9
H9140
H13199
0.97
D681-2
D507
D540
H13911
D597
D627
D566
0.86
MW2
MSSA476
D714
0.75
0.7
0.76
D523-5
D594
D672-2
D628
D629
D613
D30
USA300_FR3757
USA300_TCH1516
Newman
TW20
H7051
H7951
H7681
COL
NCTC8325
D521-3
D637
D517
D554
D605
0.71
H7639
D577
D512-1-2-4-5
D521-1-2
D524
D531
D535-2
D592
D599
D607
D608
D547
D563
D710
D719
D662
D651
D636-1-2
D720
MRSA252
0.54
D574
D589
D558
D584
D657
H6606
H13717
0.99
D560
D643
D579
0.0020 substitutions per site

Figure 2.1. MLST analysis reveals phylogenetic relationships between *S. aureus* nasal carriage and clinical strains. Bayesian phylogram indicating the evolutionary relationships of *S. aureus* strains analyzed in this study. Represented are 66 nasal carriage strains (strains colored in black) and 27 clinical strains (strains colored in red). Note that several clinical isolates cluster on the same genetic clade as major clinical strains. Numbers represent posterior probabilities and grey-filled circles represent nodes receiving 100% posterior probability support. Colored boxes are consistent with strain groupings in Figure 2.2 (below).

In addition to estimating phylogenetic relationships among strains we also utilized eBurst, which grouped the strains into nine clusters and eight singletons (Figure 2.2A). As with the BI, eBURST identified a high degree of relatedness between nasal carriage and clinical isolates. While only four STs contain both nasal carriage and clinical isolates, six of the major clusters contain both classes of isolates. eBURST groupings were based on strains sharing six of seven identical loci and takes into consideration the possibility for genetic recombination; therefore, the level of nucleotide divergence between two different STs contained within the same genetic cluster cannot be determined. To identify whether large-scale nucleotide differences were present between isolates grouped in the same cluster, but belonging to different STs, a minimal spanning network was generated (Figure 2.2B). This methodology does not take recombination events into consideration, but analyzes all mutations present between samples. The minimal spanning network revealed that many of the strains clustered by eBURST and BI contained only one or few polymorphisms between one another and may further indicate the clonality of the *S. aureus* genome. Collectively, all three computational approaches employed to identify strain relatedness were highly concordant in revealing the genetic associations between

nasal carriage strains belonging to the cohort generated for this study and clinical isolates from around the world.

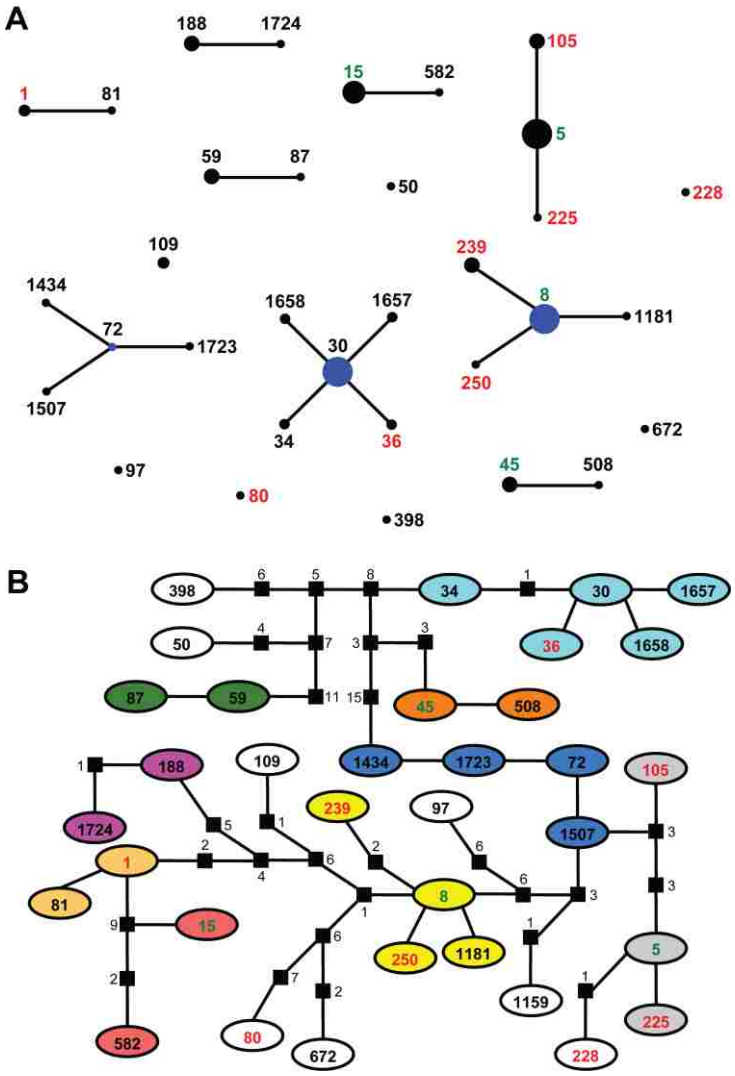

Figure 2.2. Nasal carriage and clinical isolates of *S. aureus* belong to the same genetic clusters. STs written in black are nasal carriage strains, those written in red are clinical strains and STs written in green indicate both nasal carriage and clinical isolates are contained within. (A) eBURST application of MLST data from all isolates analyzed in this study. Numbers represent ST. STs that are linked by a line belong to the same cluster. Circle sizes are proportional to the

number of strains within the ST. (B) Minimal spanning network of MLST data from the same isolates analyzed in (A). Circles represent STs with the numbers within. Branches represent a single nucleotide change between neighboring STs. Black squares indicate multiple nucleotide changes between adjoining STs with the number of differences indicated by the adjacent numbers. Like-colored circles represent STs belonging to the same cluster, as in (A). Non-colored circles are singletons by eBURST analysis.

Table 2.1. Diversity indices for virulence genes analyzed in this study.

Locus	Regions Analyzed	Includes clinical strains?	# of different strains	# of different haplotypes	Index of Discrimination (ID) (%)	dN/dS ratio
clfA	R	Yes	80	54	98.1	0.0750
		No	52	39	98.4	0.0758
clfB	R	Yes	89	58	98.4	0.0625
		No	61	44	98.6	0.0641
fnbA	D, W, & M	Yes	90	25	89.4	0.2017
		No	63	20	92.1	0.1934
fnbB	D, W, & M	Yes	67	22	90.3	0.1498
		No	56	22	91.5	0.1513

2.3.2 Virulence gene typing facilitates sub-sequence type strain resolution

Since MLST is based on the analysis of slowly evolving housekeeping genes within the *S. aureus* genome, we also analyzed hypervariable virulence-related genes to characterize further the genetic relationships between nasal carriage and clinical isolates. Virulence loci within *clf* and *fnb* gene families were chosen because they have previously been shown to facilitate strain resolution beyond that which is achievable using MLST alone (60, 107). The genetic diversity of *S. aureus* nasal carriage strains was assessed at *clfA*, *clfB*, *fnbA*, and *fnbB*. Between 52 (*clfA*) and 63 (*fnbA*) different isolates were typed over the hypervariable repeat regions of the *clf* and *fnb* genes (Figure 2.3), facilitating additional sub-ST strain resolution for 16 out of the 66

(~24%) nasal carriage isolates that underwent MLST (Appendix A; Table A.1). As summarized in Table 2.1, the *clf* genes were more variable than the *fnb* genes, although both gene families exhibited high levels of genetic diversity, overall. Both *clf* genes exhibited indices of discrimination (ID) of approximately 98.5%. Within the *fnb* loci, both *fnbA* and *fnbB* were highly variable; however, *fnbA* was slightly more variable than *fnbB*. With the exception of *clfB*, increased IDs were observed in this study as compared to those previously reported for epidemic strains of *S. aureus* where the IDs for *clfA*, *fnbA*, and *fnbB* were found to be 87.5%, 62.8%, and 67.9%, respectively (60). The elevated IDs observed here may, in part, be owed to the larger sample size analyzed within the current study. When epidemic strains were included in the analysis with nasal carriage strains, the IDs for all four genes remained relatively unaffected, albeit slightly lower than with nasal carriage strains alone (Table 2.1).

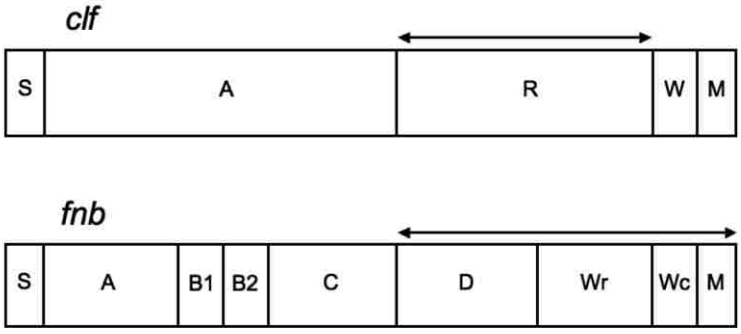

Figure 2.3. Genetic structure of *clf* and *fnb*. Double-ended arrows indicate region of analysis for this study. For clf genes: S, signal sequence; A, fibrinogen/fibronectin-binding domain; R, serine-aspartic acid repeat region; W, wall spanning domain; M, membrane spanning domain. For fnb genes: S, signal sequence; A, fibrinogen-binding domain; B, region containing 2 repeats of unknown function (only present in *fnbA*); C, region containing Du repeats with fibronectin-binding activity; D, region containing 4-5 repeats with fibronectin-binding activity; Wr, proline-

rich repeat region of the wall spanning domain; Wc, constant region of the wall spanning domain; M, membrane spanning domain.

2.3.3 Virulence genes in *S. aureus* provide evidence of purifying selection despite heightened nucleotide diversity

Among the virulence genes analyzed in this study, overall nucleotide diversities were relatively high, with the *clf* genes exhibiting approximately three times more nucleotide diversity than the *fnb* genes (data not shown). Nucleotide diversities represent the average number of nucleotide differences between two sequences at a given site. The nucleotide diversities for *clf* genes were approximately 0.15 while those for the *fnb* genes were approximately 0.05 (data not shown). Despite high nucleotide variability across all loci, strong purifying selection was observed (Table 2.1) by dN/dS ratios of less than 0.1 at both *clf* loci. Similarly, *fnb* genes exhibited evidence of strong purifying selection ($dN/dS \leq 0.2$) despite their heightened nucleotide diversity. These findings suggest that the repeat domains within the virulence genes exhibit a specific and essential function, such that natural selection maintains amino acid homology in spite of high levels of nucleotide substitution.

2.3.4 Virulence gene repeat domain lengths are identical between nasal carriage and clinical isolates

The R regions of the *clf* genes have previously been assumed to function as a stalk for the extension of the ligand binding domain from the bacterial cell wall (138). Thus, longer R region lengths may enhance bacterial adherence to nasal epithelia. Since the repeat region of the wall-spanning domain (Wr) within *fnb* genes may also serve a similar function, we extended this

hypothesis to include these domains as well. As such, we aimed to elucidate whether longer repeat regions associated with nasal carriage strains as compared to clinical strains of *S. aureus.*

Collectively, a large degree of length variability was observed among both nasal carriage strains of *S. aureus*, as well as clinical strains. For nasal carriage strains, *clf* R region lengths ranged from 816 bp to 1212 bp for *clfA* and 417 bp to 981 bp for *clfB*. Clinical strains analyzed in this study exhibited very similar (and often identical) R region lengths to those of the nasal carrier strains (Figure 2.4A). Repeat region lengths within clinical strains of *S. aureus* ranged from 666 bp to 1224 bp for *clfA* and 615 bp to 939 bp for *clfB*.

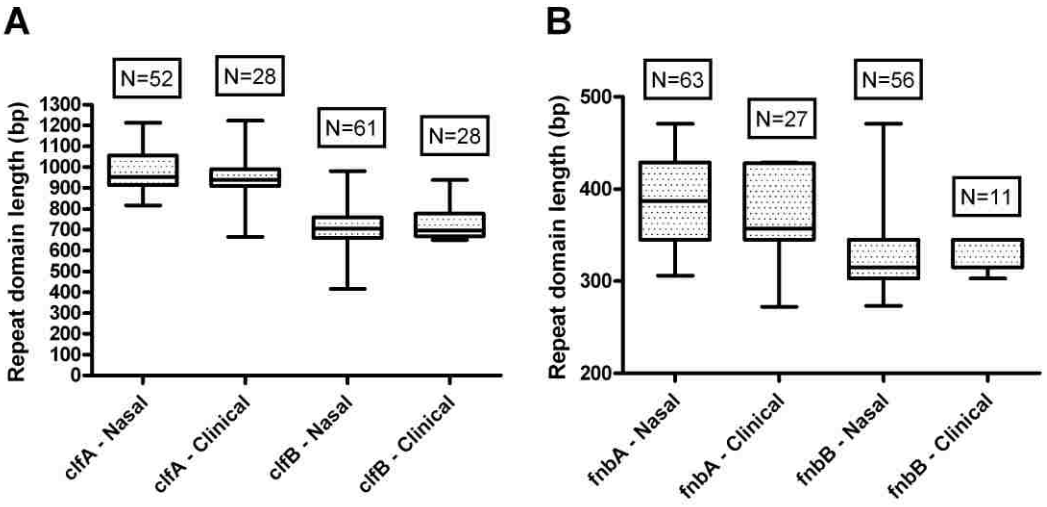

Figure 2.4. Repeat domain lengths of *clf* and *fnb* genes are indistinguishable between nasal carriage and clinical isolates. Shown are Box and Whisker plots comparing repeat domain lengths between nasal carriage strains of S. aureus and clinical strains. (A) Repeat domain lengths compared at *clf* loci; and (B) repeat domain lengths compared at *fnb* loci.

When considering the degrees of DNA strand-slippage at each *clf* locus, slippage events were three to four times more prevalent in *clfA* than *clfB*. Of 178 unique repeats observed in

nasal carriage strains at *clfA*, 32 (18%) were 12 nucleotides in length. Of 107 unique repeats observed at *clfB*, only five (4.7%) were 12 nucleotides in length. When all strains analyzed in this study (nasal carriage and clinical) were included in repeat profiling, a total of 185 unique repeats were observed at locus *clfA* with 34 (18.4%) of these being the result of slippage. When all strains were included in the analysis of *clfB*, at total of 109 unique repeats were observed with six (5.5%) being the result of slippage events (refer to Appendix A; Tables A.3 and A.4).

When analyzing the nucleotide sequences of the *fnb* genes, the majority of variation laid within the Wr domains in the form of indels, resulting in length differences. No difference was observed in total length of the Wr domains between nasal carriage and clinical isolates (Figure 2.4B). At *fnbA*, Wr domain lengths ranged from 345 bp to 471 bp for nasal carriage isolates and 273 bp and 429 bp for clinical isolates. At *fnbB*, Wr domain lengths for nasal carriage isolates ranged from 273 bp to 378 bp while clinical isolates ranged from 303 bp and 345 bp. Based on the analysis of repeat domain lengths alone at *clf* and *fnb* loci, it was not possible to distinguish between nasal carriage and clinical strains of *S. aureus*. As such, an analysis of nucleotide sequences was carried out to determine if, at the nucleotide level, nasal carriage and clinical strains belong to the same genetic lineages, or if distinct populations were evident.

2.3.5 Nasal carriage and clinical isolates of *S. aureus* belong to the same genetic lineages

To identify relatedness between nasal carriage and clinical isolates of *S. aureus*, *clf* and *fnb* nucleotide sequences were analyzed. Collectively, a large degree of nucleotide sequence diversity among the carrier strains present within our cohort was observed for the *clf* genes, as is indicated by the number of unique repeat units at each locus (refer to Appendix A; Tables A.3 and A4). For *clfA*, 52 nasal carriage isolates were genotyped and 178 unique repeat sequences

were identified. Interestingly, *clfB* exhibits a similar index of discrimination to that of *clfA* (Table 2.1); however, only 107 unique repeat sequences were identified from 61 isolates. While a large degree of variability within the *clf* gene fragments analyzed is the result of point mutations, insertions or deletions of repetitive units are the primary means of variability. As such, strain relatedness cannot be determined using algorithms that rely on sequence alignment (170). Therefore, in agreement with previous studies (37, 39, 106, 107, 170, 171), lineage assignments were carried out by visual inspection of R domain profiles. On the basis of R domain typing, the 52 nasal carriage isolates for *clfA* were grouped into six lineages (1-6), and 39 haplotypes (Appendix A; Table A.1). Lineage 1 was the largest within the sample set containing 26 of 52 (50%) nasal carriage isolates (Figure 2.5A). Interestingly, 24 of the 28 (85.7%) clinical isolates analyzed herein also belong to this same lineage, including the highly prevalent and virulent USA300 and MW2 (USA400) strains.

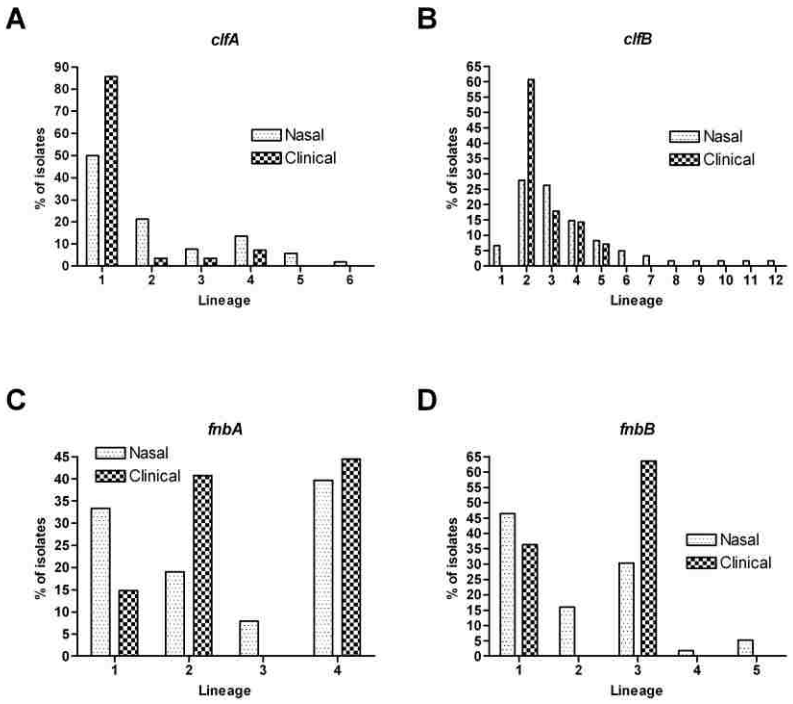

Figure 2.5. Classification of *S. aureus* strains reveals lineage associations between nasal carriage and clinical isolates. (A) *clfA*, N=52 nasal carriage and 28 clinical isolates; (B) *clfB*, N=61 nasal carriage and 28 clinical isolates; (C) *fnbA*, N=63 nasal carriage and 27 clinical isolates; and (D) *fnbB*, N=56 nasal carriage and 11 clinical isolates.

When performing the same R domain analysis at locus *clfB*, all 61 nasal carriage strains belonged to 12 lineages (1-12), and 44 different haplotypes (Figure 2.5B and Table S1). Lineages 2 and 3 contain the most nasal carriage isolates with 17 (27.9%) and 16 (26.2%), respectively. As with *clfA*, a large proportion of the clinical isolates analyzed within this study also belong to these two lineages (including again, USA300). Twenty-two of the 28 (78.6%) clinical isolates analyzed in this study belonged to lineages 2 and 3. Seventeen (60.7%) isolates share lineage 2 with nasal carriage strains while another five (17.9%) belong to lineage 3.

Nucleotide sequence analyses of *S. aureus* isolates were carried out using the D, W (Wr and Wc) and M domains of the *fnb* genes as well. DNA sequence analysis of these domains at locus *fnbA* made it possible to categorize all 63 strains analyzed into four lineages (1-4) while the 56 strains analyzed at *fnbB* were separated into five different lineages (1-5) (Appendix A; Table A.1). At both *fnb* loci, two separate lineages were identified that contained the majority of nasal carriage isolates. At *fnbA*, lineages 1 and 4 contained 73% of the nasal isolates with 33.3% and 39.7%, respectively (Figure 2.5C). At *fnbB*, lineages 1 and 3 were most prevalent within the data set (76.8% of nasal isolates) with 46.4% and 30.4%, respectively (Figure 2.5D). Within *fnbA*, a total of 59.2% of clinical strains were identified as belonging to lineages 1 and 4 (14.8% and 44.4%, respectively), while 100% of clinical strains, at *fnbB*, belonged to the two most prevalent nasal carriage lineages, 1 and 3 (36.4% and 63.6%, respectively). The lineage

assignment data for both *clf* and *fnb* genes identify further the genetic relatedness between the nasal carriage strains analyzed in this study and clinically relevant isolates of *S. aureus.* However, to further verify the relatedness we next sought to determine the prevalence of nasal carriage and clinical strains of *S. aureus* exhibiting identical virulence gene sequences.

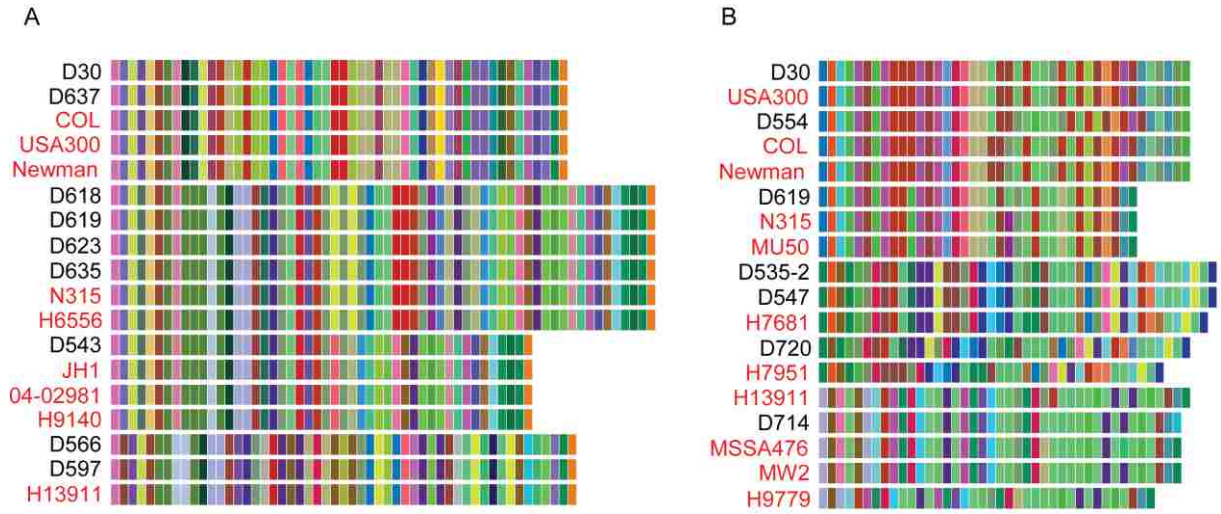

Figure 2.6. Nasal carriage and clinical isolates of *S. aureus* share (near-) identical *clf* repeat region sequences. Shown are *clf* repeat regions represented as color-coded bars. Like-colored boxes indicate 100% sequence similarity between isolates. (A) *clfA* and (B) *clfB* repeat region sequences from a representative sampling of all isolates analyzed within this study. Isolate names written in black are nasal carriage strains and those written in red are clinical strains.

2.3.6 Nasal carriage and clinical isolates of *S. aureus* are identical in *clf* and *fnb* gene sequences

At locus *clfA*, 11 (39.3%) of the clinical isolates (including USA300) analyzed exhibited identical nucleotide sequence to isolates from healthy donors. Refer to Figure 2.6A for a representation of the sequence similarities within a subset of the clinical and nasal carriage isolates analyzed herein. For the complete *clfA* data set, refer to Appendix A, Table A.5 and

Figure A.1. At *clfB*, nine (32.1%) clinical isolates, including USA300 and MW2, exhibit 100% sequence identity to nasal carrier strains analyzed in this study. Refer to Figure 2.6B for a comparison of R regions from a subset of nasal carrier and clinical isolates. Refer to Appendix A, Table A.6 and Figure A.2 for the complete *clfB* data set.

When considering the D and W domains of the *fnb* genes, a large percent of clinical isolates exhibited similar, and in many cases identical, genetic sequences to the nasal carriage strains. In fact, 22 (81.5%) of the clinical isolates at locus *fnbA* shared 100% nucleotide sequence identity with nasal carriage isolates. Similarly, for *fnbB*, the same was observed for 8 (72.7%) of the clinical isolates analyzed in this study. Among the clinical strains exhibiting 100% sequence identity with nasal carriage strains at *fnbA* are USA300, N315 and COL (Figure 2.7). Shown in Figure 2.7 is an amino acid alignment of eight *S. aureus* strains analyzed in this study (four nasal carriage and four clinical strains) revealing the large degree of homology between the two classifications of strains. At *fnbB*, USA300 was again observed to exhibit 100% sequence identity with nasal carriage strains (data not shown), further supporting the relatedness of nasal carriage strains from the cohort analyzed in this study and clinical isolates of worldwide origin.

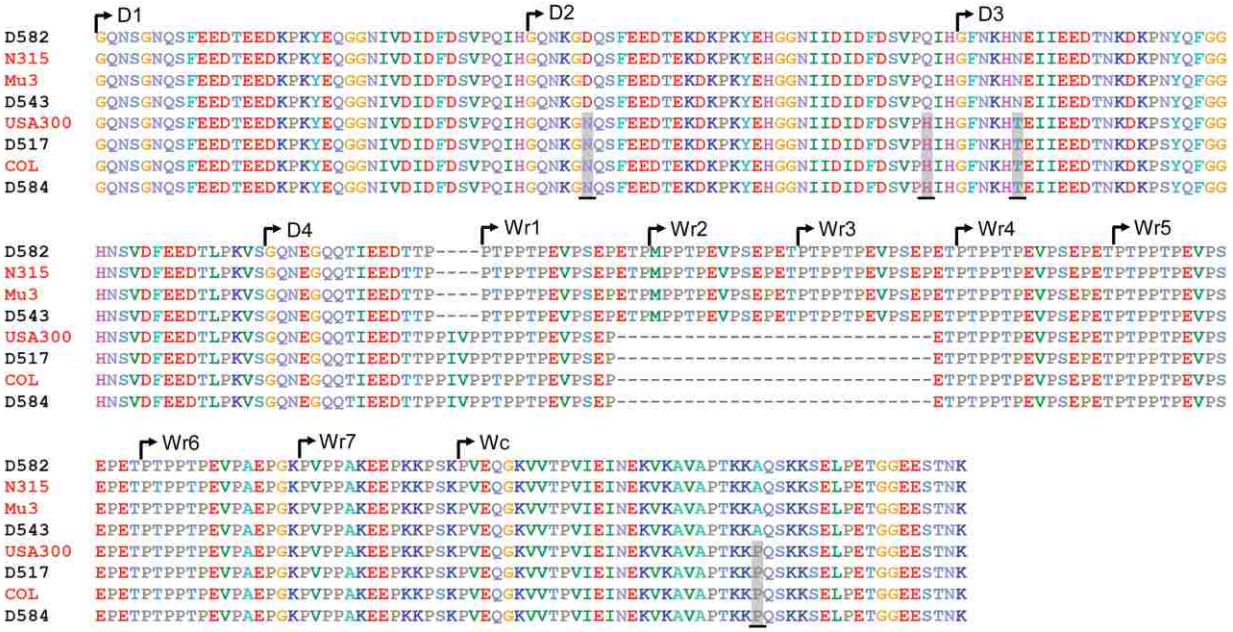

Figure 2.7. Comparison of FnbA amino acid sequences between four representative nasal carriage and four representative clinical *S. aureus* isolates. Note that identical sequences are evident between both nasal carriage and clinical isolates. D1-D4, repeat region containing fibronectin-binding activity; Wr1-Wr7, proline-rich repeat region of the wall spanning domain; Wc, constant region of the wall spanning domain. Isolate names written in black are nasal carriage strains and those written in red are clinical strains. Regions of amino acid variability are shaded grey and underlined.

2.4 Discussion

Due to the increasing public health concern regarding the severity and rates of *S. aureus* infection throughout the world, it was the goal of this study to perform an evolutionary genetic analysis of nasal carriage strains of *S. aureus* from a healthy population while also analyzing the relatedness of these isolates to strains responsible for pathogenic infection. Using MLST data,

computational analyses were conducted to determine the relatedness of clinical and nasal carriage isolates. Bayesian estimation of strain relatedness identified clades containing both nasal carriage and clinical isolates, indicating that these two groups share a recent common ancestor. One drawback to this approach is that, presumably due to the low variability and slow rate of evolution within housekeeping genes, low posterior probabilities are observed due to lack of genetic variability. As such, eBURST and a maximum parsimony minimum spanning network analyses were carried out. eBURST takes into consideration the fact that genetic recombination events within bacterial genomes occur and therefore, highly related strains may still exhibit large-scale nucleotide variation at one or few loci. Here, we generated eBURST clusters requiring six of seven identical loci between strains. Using this method, all of the strains analyzed in this study (nasal carriage and clinical) were clustered into few groups, identifying high genetic relatedness between the nasal carriage strains analyzed in this study and clinical isolates previously identified from around the world.

While many bacterial genomes exhibit high levels of genetic recombination, *S. aureus* has previously been shown to be highly clonal (49, 203), and therefore, a minimal spanning network based on maximum parsimony analysis was also conducted. Once again, the MLST data identified that nasal carriage and clinical strains are genetically related, with only few, if any, single nucleotide polymorphisms present between groups. Moreover, many of the sequence types (STs) that were clustered using BI and eBURST were also identified by this method as being closely related.

Collectively, MLST of nasal carriage and clinical strains in this study indicates that these strains share common genetic lineages based on housekeeping gene fragments in the common

minimal genome. This finding unto itself is not surprising given previous reports of nasal carriage and clinical isolates sharing common lineages (49, 50, 203) and that the nature of MLST limits the distinction between strains, which may in turn over represent the relatedness of nasal carriage and clinical isolates. To increase discrimination between strains virulence genes were also analyzed. Not only are virulence genes attractive because of their suitability for analyzing small geographic regions and providing sub-sequence type resolution (60, 106, 170), but also because of their involvement in the pathogenicity of *S. aureus*. Adhesion genes from the clumping factor gene family (specifically, *clfB*) are known to play a significant role in nasal colonization (204) while fibronectin binding protein genes are major contributors to the virulence of *S. aureus* in clinical settings (144). A number of reports utilized genetic typing of these virulence genes for their discriminatory capabilities between isolates, particularly those belonging to the same ST. By contrast, the current study is, to our knowledge, the first report on the sequence diversity of virulence-related genes from nasal carriage strains of *S. aureus* from a cohort of healthy individuals in the United States.

Here, we have observed a discriminatory power among the virulence genes superior to that reported previously for only clinical strains (60). This discrepancy may, in part, be owed to the ethnic diversity within our study area, but may also be an indication of the complex evolutionary processes of *S. aureus*; namely the potential for nasal carriage strains to continuously adapt to new environments and evade the host immune system (via continual host-to-host transmission). Using the enhanced discriminatory power of virulence genes, sub-ST strain resolution was achieved for a large percent (~24%) of strains analyzed in this study. The heightened discriminatory ability of virulence genes was attributable to the high degree of

genetic variability at these loci. Interestingly, while a high level of genetic variability was identified among the *clf* and *fnb* loci, a strong indication of purifying selection was still observed. Purifying selection is a reflection that, on average, little to no adaptive diversification (positive selection) of the gene's protein product is being maintained in the population (61), thus suggesting a possible role in the preservation of a specific function.

In order to assess the relatedness of nasal carriage isolates obtained for this study to clinical isolates, the first analysis identified length differences within the repeat regions of the *clf* and *fnb* genes. The repeat regions of the *clf* and *fnb* genes are highly variable in their number of repeating units and these regions are presumed to show the greatest variability between different strains. Discrimination based on the number of repeat units within these regions did not facilitate distinction between nasal carriage strains and clinical isolates. Perhaps the mere presence of determinant virulence genes is sufficient to confer pathogenicity and is not necessarily related to any genetic differences within virulence genes between strains. This hypothesis requires further exploration, as few studies have focused their efforts on the contributions of different genetic regions to pathogen virulence. Initial support to this hypothesis is observed with ClfA. Interestingly, the ClfA protein requires a minimum of 80 amino acid residues within its repeat domain (68); below this critical length, reduced ClfA activity is observed, while no obvious difference in activity is observed when the length increases. It is possible that the length of the repeat domains within *clfB* or *fnb* genes is sufficient for protein function and as long as they are greater than an, as yet unknown, minimum length, protein function and consequential virulence are indistinguishable between strains.

While differences in repeat length may not directly promote virulence, they may be important in the ability of a strain to colonize nasal epithelia. The difference between persistent, intermittent, or non-carriage may be dependent upon the repeat domain lengths of adhesion genes. Such a consideration has previously been addressed for the coagulase and protein A genes, with no correlation between carriage status and repeat length identified (196); however, never before has such a longitudinal study been conducted for the *clf* or *fnb* genes, which were previously identified as putative determinants of nasal carriage.

Variable number of tandem repeat profiling did not yield clear distinctions between nasal carriage and clinical isolates, and as such an analysis of nucleotide sequence variability was subsequently carried out for *clf* and *fnb* genes. With increasing frequency, analyses such as this for the *clf* genes are being conducted (60, 107, 161-163), and we feel that a database, much like that for *spa* typing and MLST, would be beneficial for the transfer and continuity of R domain data between laboratories. The software for R region profiling used in this study is publicly available, and could be combined in a database with all previous repeat sequences and profiles to facilitate faithful identification of strains worldwide. The large amount of data being obtained for *clf* genes and the observed importance of these genes in the classification of strains necessitates the transfer and continuity of genetic data between laboratories. Using the software for *clf* repeat region profiling developed in this study and multiple sequence analysis of *fnb* genes, the classification of nasal carriage and clinical isolates into genetic lineages was conducted.

A high proportion of clinical isolates analyzed in this study (>75%) belonged to the same genetic lineages as nasal carriage strains, revealing an evolutionary relationship stronger than has

heretofore been identified. That being said, it is recognized that a full appreciation for the genetic relatedness between nasal carriage and clinical isolates will require future endeavors with employment of large scale molecular typing on large cohorts of both nasal carriage and clinical isolates from the same geographic region. Here we have provided a foundation from which to build upon where a diverse population of clinical strains from around the world has shown genetic relatedness to strains from a nasal carriage population. Furthermore, as next generation sequencing becomes a more feasible option for bacterial typing, and strain collections become more prevalent, the relatedness of nasal carriage and clinical isolates will be more easily and reliably identified.

Of particular interest to this study are the molecular similarities between nasal carriage strains and the highly virulent community-associated methicillin-resistant (CA-MRSA) strains USA300 and MW2 (USA400). These are the two most prevalent strains responsible for CA-MRSA infection within the United States, responsible for 97-99% of all community-acquired skin and soft tissue infections (72, 96, 130). Interestingly, while USA300 and MW2 do not belong to the same ST, they do share common lineages to each other at *clfA* and *fnbB* (Table S1). They also share common lineages with many nasal carriage strains at all four virulence genes analyzed, belonging to the same lineages as the majority of nasal carriage strains in all cases. Collectively, nucleotide diversities within the repeat regions of *clf* and *fnb* loci facilitate high discriminatory power between strains; however, the molecular data do not distinguish the nasal carriage strains belonging to the cohort analyzed in this study from clinical isolates. In fact, the molecular population analyses indicate that nasal carrier strains share molecular lineages, and are often genetically identical to those strains of clinical significance. While the virulence of a

bacterial strain is undoubtedly multifactorial, involving an as yet unknown number of virulence genes, not to mention host factors, the genetic associations identified within this study between clinical and nasal carriage isolates suggest that strain relatedness between nasal carriage and clinical isolates may be higher than has previously been recognized.

3. CHARACTERIZATION OF THE RETROCYCLIN ANALOGUE RC-101 AS A PREVENTATIVE OF *STAPHYLOCOCCUS AUREUS* NASAL COLONIZATION

3.1 Introduction

Nasal colonization by *Staphylococcus aureus* occurs in approximately 20-30% of healthy individuals (197). The primary reservoir for *S. aureus* is the anterior nares, but the occurrence of nasal colonization also increases the prevalence of this bacterium on other surfaces of the body (104). As such, nasal carriage of *S. aureus* is a major risk factor for endogenous infection (autoinfection), ranging from minor skin and soft tissue infections to serious bacteremia (24, 200). Multiple studies have shown that removal of *S. aureus* from the nasal vestibule using antimicrobial agents (referred to as nasal decolonization), prior to hospitalization, significantly reduces incidences of nosocomial infection (10, 80, 105, 200, 202). For the past 25 years, the most common means of nasal decolonization prior to hospitalization has been the use of mupirocin ointment; however, resistance to this antibiotic is increasing (24, 35). Thus, there is urgent need to develop novel compounds to prevent or treat *S. aureus* nasal carriage, particularly in preoperative patients.

Toward this goal, we discovered that the retrocyclin class of θ-defensins is potently active against a broad spectrum of microbes including strains of *S. aureus* (28). Retrocyclins are 18-residue peptides that contain three intramolecular disulfide bonds, which stabilize a β-sheet conformation (191), and represent the first truly circular peptides of vertebrate origin (114, 185, 192). They are extremely stable and can resist boiling, acidic conditions, and other harsh environments. Notably, RC-101 has been recovered and found to remain bioactive after nine

days of treatment in an *ex vivo* model of organotypic human vaginal tissue containing vaginal mucus (25). Similarly, RC-101 was observed to be stable and bioactive after eight days of *in vivo* vaginal treatment in pigtailed macaques (29), thus highlighting the stability of this peptide in mucosal environments.

Whereas both human and non-human primates produce α- and β- defensin peptides (185), humans do not produce endogenous θ-defensin peptides because a premature stop codon precludes translation (28, 137). As such, these peptides have been recreated by solid-phase synthesis and found to exhibit broad-spectrum antimicrobial activity against a number of bacteria, fungi and viruses (28, 114, 185, 192, 208), without any noted cellular toxicity or inflammation *in vivo*, *ex vivo* and *in vitro* (25, 29). One retrocyclin analogue, RC-101, contains a single arginine to lysine mutation as compared to wild type retrocyclin, exhibits heightened activity in antiviral assays (146), is non-hemolytic to human red blood cells, and is not cytotoxic to a number of human cell lines at concentrations up to 500 μg/mL (25, 28, 29, 54, 199).

In the current study, we have characterized further the antimicrobial properties of RC-101 against both nasal carriage and clinical isolates of *S. aureus*. Importantly, RC-101 prevents the adherence and survival of *S. aureus* on cultured human nasal epithelia while inducing no noticeable cytotoxicity or proinflammatory responses. These findings identify RC-101 as a potentially promising therapeutic agent for nasal decolonization of *S. aureus* and supports further development of this peptide as an intranasal antibiotic.

3.2 Methods and Materials

3.2.1 Bacterial isolates

Nasal carriage strains of *S. aureus* were collected from the anterior nares of donors at the University of Central Florida (UCF; Orlando, Florida, USA) following the protocol described in (113). Written consent was obtained from all donors and samples were collected under a human subjects protocol approved by the UCF Institutional Review Board. The clinical strains, USA300, N315, and COL, were obtained from the Network on Antimicrobial Resistance in *Staphylococcus aureus* (Eurofins Medinet, Inc., Chantilly, Virginia, USA), of which our laboratory is a member.

3.2.2 Peptide synthesis, storage and utilization

The 18 amino acid peptide RC-101 was synthesized as previously described (26, 28). Following synthesis, lyophilized peptide was stored at -20^{o}C until use. Prior to use, RC-101 was reconstituted in sterile water/0.01% (v/v) acetic acid and diluted accordingly to desired working concentrations. Surplus peptide was aliquoted in single-use volumes and stored at -20^{o}C. As such, working volumes of peptide were limited to not more than one freeze-thaw cycle prior to use.

3.2.3 Turbidity assay

As an initial screen of the anti-*S. aureus* activity of RC-101, turbidity assays were performed as adapted from (135). Briefly, bacteria were grown to logarithmic growth phase and diluted in Mueller Hinton Broth (Sigma-Aldrich Co., St. Louis, Missouri, USA) containing 0.5%

sucrose. Aliquots of approximately 10^4 colony-forming units (80 µL) were added to the wells of a flat bottom 96-well plate (MidSci, St. Louis, Missouri, USA) with 20 µL of either vehicle or RC-101 (2.5-20 µM final). Plates were incubated in a SpectraMax 190 microplate reader (Molecular Devices, Sunnyvale, California, USA) at 37^{o}C for 16 hrs. Turbidity readings at 550 nm were acquired every five minutes following 15 seconds of agitation. Optical density data were plotted against time to generate growth curves for all samples. The time at which each sample entered logarithmic growth was considered the growth threshold (G_t) for these assays. The G_t for each sample was compared to the G_t of the vehicle treated sample and the retardation in growth was represented as delta G_t.

3.2.4 Tissue culture

Human nasal epithelia (RPMI 2650, American Type Culture Collection, Manassas, Virginia, USA) were grown to confluence on collagen-coated Transwell® inserts (12 mm diameter, 0.4 µm pore size, Corning Inc., Corning, New York, USA). Cell culture media contained Dulbecco's Modified Eagle's Medium (DMEM; Mediatech Inc., Manassas, Virginia, USA) with glucose (4.5 g/L), L-glutamine (584 mg/L), and sodium pyruvate (110 mg/L), supplemented with 10% (v/v) fetal bovine serum (FBS; Gemini Bioproducts, West Sacramento, California, USA), penicillin (100 U/mL), and streptomycin (100 µg/mL). Antibiotic-supplemented media were changed daily until epithelia reached confluence. Following confluence, cells were cultured in antibiotic-free media at the air-liquid interface at 37^{o}C and 5% CO_2 for four days prior to use in adhesion assays. During culturing at the air-liquid interface, antibiotic-free media were changed daily.

Organotypic airway epithelial tissues (EpiAirway^{TM}) were obtained from MatTek Corporation (Ashland, Massachusetts, USA) and maintained according to the manufacturer's instructions. These tissues resemble closely epithelial tissue of the respiratory tract. They are representative of healthy human donors and contain pseudo-stratified, highly differentiated tracheal/bronchial epithelia.

3.2.5 Colony forming unit (CFU) assay

To study the bactericidal effects of RC-101 on *S. aureus*, CFU assays were carried out with an adapted procedure from (27, 31). Briefly, nasal carriage isolates of *S. aureus* were grown to logarithmic growth phase in Trypticase Soy Broth at 37^{o}C and 250 rpm and diluted in minimal media containing DMEM with glucose (4.5 g/L), L-glutamine (584 mg/L), and sodium pyruvate (110 mg/L), supplemented with 0.05% (v/v) FBS to approximately 100 000 CFU/mL. *S. aureus* survivability reactions were prepared by incubating 4 µL of dilute bacteria with 1 µL of either vehicle or RC-101 at the appropriate concentration (1-10 µM final). Cultures of 5 µL were grown in sterile 72-well polystyrene Nunc^{TM} MiniTrays (Nalge Nunc International, Rochester, New York, USA) with 3 µL of liquid wax overlaid to prevent evaporation. Cultures were incubated at 37^{o}C and 5% CO_2 for 0, 0.25, 0.5, 1, 2, 3, 6, and 9 hrs. after which time the entire sample, or dilutions thereof, were plated on Trypticase Soy Agar (TSA) and incubated for 16 hrs. at 37^{o}C. The survival of *S. aureus* was determined by enumerating CFUs from RC-101 treated samples and comparing to that of the vehicle treated samples.

For studies assessing the effects of bacterial starting inocula on RC-101 activity, RC-101 (10µM final concentration) was incubated with increasing starting concentrations of *S. aureus.* Cultures were incubated at 37^{o}C and 5% CO_2 for 0, 0.5, 1, 2, 3, 6, and 9 hrs. after which time the

entire sample, or dilutions thereof, were plated on TSA and incubated for 16 hrs. at 37^{o}C. The survival of *S. aureus* was determined by enumerating CFUs from RC-101 treated samples and comparing to that of vehicle treated samples.

3.2.6 Epithelial cell adhesion assays

Epithelial cell adhesion assays have previously been used to assess the binding of bacteria to human epithelia under a number of different conditions (33, 153, 154, 189, 190). Here, we have employed this assay to elucidate the nasal epithelial cell adhesion properties of *S. aureus* in the presence of RC-101. Adhesion assays were carried out as previously described (153, 154) using two different models of human airway epithelia. EpiAirway^{TM} tissues (MatTek tissues) or confluent layers of human nasal epithelia (RPMI 2650), exposed to the air-liquid interphase for four days (described above) were inoculated with 10-50 bacteria (100 µL) in minimal media containing DMEM with glucose (4.5 g/L), L-glutamine (584 mg/L), and sodium pyruvate (110 mg/L), supplemented with 0.05% (v/v) FBS. Inoculated epithelial layers were incubated at 37^{o}C and 5% CO_2 for 15 min. prior to treatment with either vehicle or RC-101 (1-20 µg/tissue). At 0, 3, 6, and 9 hrs. post-treatment, wash fractions containing non-adherent bacteria were collected by rinsing the apical epithelial cell surface three times in 300 µL of minimal media (900 µL total). Following collection of the wash fraction, adherent bacteria (adhere fraction) were isolated by scraping the cell layer, and the sample was collected in 900 µL of minimal media. To liberate bacteria from epithelia, the adhered fraction was then sonicated (Model 100 Sonic Dismembrator; Fisher Scientific, Pittsburgh, Pennsylvania, USA) using 10 X 0.5 sec. pulses on power setting three. Samples were then plated on Trypticase Soy Agar and

incubated for 16 hrs. at 37^{o}C. The survival of *S. aureus* was determined by graphing CFU from RC-101 treated samples and comparing to that of the vehicle treated samples.

3.2.7 Epithelial cell viability assays

Human nasal epithelial cell viability was quantified using a MTT (3-[4,5-Dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) reduction assay, according to the manufacturer's instructions (Trevigen Inc., Gaithersburg, Maryland, USA). Cytotoxicity was also measured by protease release using the CytoTox-GLO™ Cytotoxicity Assay (Promega Corp., Madison, Wisconsin, USA) according to the manufacturer's instructions. Cell viability for organotypic airway epithelia was measured using a MTT assay from the tissue supplier (MatTek Corporation, Ashland, Massachusetts, USA), following the manufacturer's instructions.

3.2.8 Detection of proinflammatory cytokines

To identify possible proinflammatory effects of RC-101 on human nasal epithelia or organotypic human airway epithelial tissues, multiplex enzyme-linked immunosorbent assays were performed. For these assays, conditioned underlay (basal media) was collected after 24 hr. and/or 72 hr. incubations of human nasal epithelia, or organotypic airway epithelia, treated with RC-101 or vehicle. Following collection, basal media was subjected to multiplex suspension bead arrays assessing 27 human proinflammatory cytokines and analyzed using a Bio-Plex 200 System (Bio-Rad Laboratories, Inc., Hercules, California, USA) following the manufacturer's instructions.

3.2.9 Statistical analyses

Statistical analyses were conducted throughout this study using GraphPad Prism 4 software (GraphPad Software, La Jolla, California, USA). Bacterial counts from CFU and adhesion assays were $\log_{10}$ reduced and statistical analyses were performed on the transformed data. For turbidity, CFU, and adhesion assays, one-tailed Student's t-tests were performed assuming a two-sample unequal variance (heteroscedastic). For epithelial cell viability assays and cytokine analyses, two-tailed Student's t-tests were performed assuming a two-sample unequal variance (heteroscedastic). For all analyses, $p < 0.05$ was considered statistically significant.

3.3 Results and Discussion

3.3.1 Growth of *S. aureus* is retarded by RC-101 treatment

Previous studies have shown that retrocyclin exhibits antibacterial activity against a laboratory strain of *S. aureus* (28); however, the robustness of RC-101 activity has not been well characterized, nor has the efficacy of this peptide as a potential therapeutic against *S. aureus* been assessed in any great detail. As an initial assessment of the robustness of RC-101's anti-*S. aureus* activity, turbidity assays were performed. A sampling of five nasal carriage and three clinical isolates were treated with RC-101 and their growth kinetics were monitored over a 16 hour period. In all assays, RC-101 retarded bacterial growth in a concentration-dependent manner. Shown in Figure 3.1 are turbidity data for *S. aureus* nasal carriage strains D20-7, D535-6 and D30, as well as the clinical isolate, USA300. Data have been transformed to reflect growth retardation time (delta G_t) in RC-101 treated samples as compared to vehicle. RC-101 (2.5 μM)

was sufficient to retard *S. aureus* growth for up to five hours beyond vehicle treated bacteria, while 20 µM concentrations were observed to retard growth beyond 10 hours in nasal carriage strains, and five hours in the hypervirulent USA300 strain. Clones of D20-7 and USA300 surviving the 16-hour treatment of 20 µM RC-101 were reassessed by CFU assay for evidence of enhanced resistance to RC-101. No indication of resistance toward RC-101 was observed after this initial round of passaging (data not shown). In addition to the four *S. aureus* strains shown in Figure 3.1, four more strains were also examined using this assay. Included in these additional strains were two nasal carriage isolates, D547-4 and D566-5, and two clinical isolates, N315 and COL. Upon treatment with RC-101, the additional strains were found to exhibit similar growth kinetics to those shown in Figure 3.1 (data not shown). Collectively, RC-101 exhibited robust anti-*S. aureus* activity against all strains tested in this study.

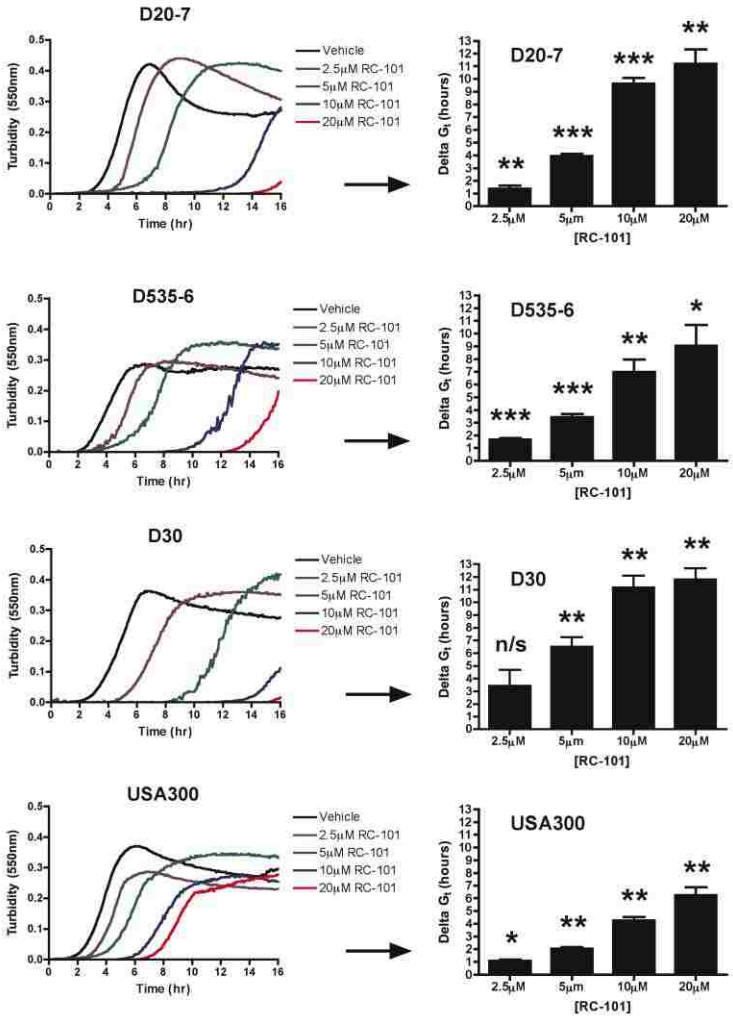

Figure 3.1. RC-101 retards the growth of nasal carriage and clinical strains of *S. aureus*. Shown are growth curves for three representative nasal carriage strains (names beginning with "D") and one representative clinical strain (USA300) treated with RC-101. Delta G_t represents the time difference between RC-101 and vehicle treated samples to reach their respective G_t values (onset of logarithmic growth phase). Left panels, representative growth curves for one of three assays. Right panels, data representative of three assays. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$. n/s, non-significant. P-values indicate statistical significance as compared to vehicle treated samples.

3.3.2 RC-101 exhibits bactericidal effects against *S. aureus*

To elucidate further the effects of RC-101 on *S. aureus* growth, CFU assays were performed. A total of eight *S. aureus* strains were tested, five of which were nasal carriage isolates and three of which were strains of clinical origin. As shown in Figure 3.2 (top panels), RC-101 exhibited a concentration dependent inhibition of growth in nasal carriage strains of *S. aureus* (D20-7, D535-6 and D30) and the clinical isolate, USA300. Importantly, bactericidal effects were observed within 15 minutes of RC-101 treatment. RC-101 concentrations as low as 5 μM resulted in significant reductions ($p=0.017-0.0002$ for all strains) in bacterial growth within 15 minutes as compared to vehicle treated bacteria. Peptide concentrations of 10 μM were observed to be almost completely bactericidal over the same 15-minute timeframe ($p\leq0.0014$ for all strains compared to their respective vehicle treatments) with complete growth inhibition observed after 30 minutes. Within three hours of treatment with RC-101 significant reductions in bacterial growth were observed for all peptide concentrations (Figure 3.2, top panels).

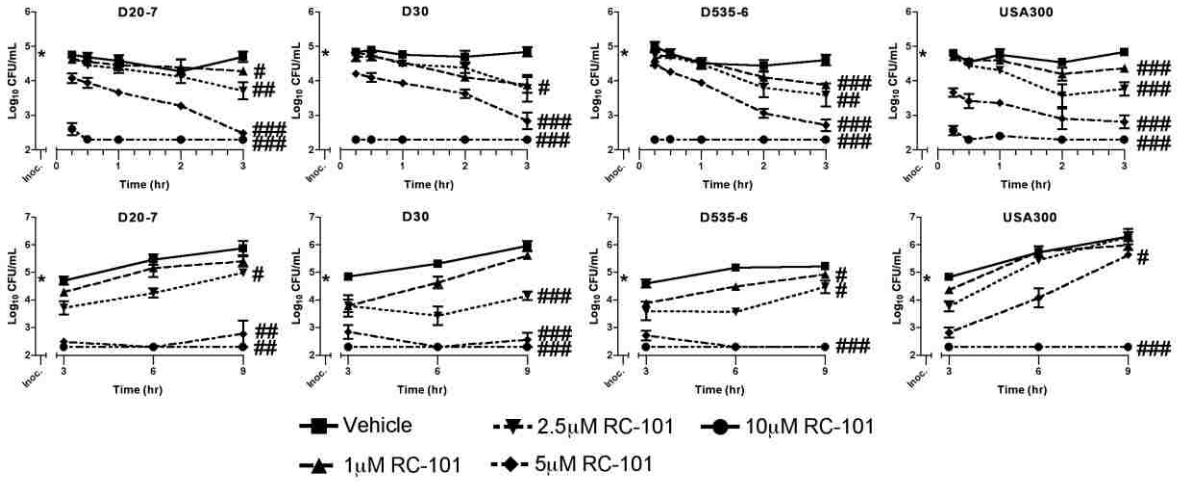

Figure 3.2. RC-101 is bactericidal toward both nasal carriage and clinical isolates of *S. aureus*. Shown are growth rate data from three nasal carriage (strain names beginning with "D") and one clinical isolate (USA300) of *S. aureus* in the presence and absence of RC-101. Top panels show RC-101 treatment for 0-3 hours while bottom panels show RC-101 treatment for 3-9 hours. Limit of detection equals 200 (i.e. $\log_{10}$=2.3) CFU/mL. $N \geq 3$ for all strains. Error bars represent the mean ± SEM. *Indicates the starting inoculum. #p < 0.05. ##p < 0.01. ###p < 0.001. P-values indicate statistical significance as compared to vehicle treated samples. For clarity of presentation, p-values are shown for three- and nine-hour treatments only.

To monitor the propensity for recovery among RC-101 treated strains, growth was monitored over an extended time course of nine hours. A RC-101 concentration of 5 μM significantly inhibited bacterial growth in all nasal carriage strains after 9 hours while a peptide concentration of 10 μM completely prevented bacterial growth from being observed in these same strains over 9 hours (Figure 3.2, bottom panels). The clinical isolate, USA300 also experienced a significant reduction in growth after nine hours of 5 μM RC-101 treatment with a 10 μM concentration completely preventing growth from being observed over the same nine hour timeframe. In addition to the four *S. aureus* strains shown in Figure 3.2, four more strains were also treated with RC-101, all of which revealed similar growth kinetics (data not shown). Among the additional strains were two nasal carriage isolates, D547-4, and D566-5 and two clinical isolates, N315 and COL.

The starting inocula used for CFU assays (Figure 3.2) approximate the physiological concentrations of *S. aureus* in nasal fluid from carriers (27, 30, 104); however, the effect of increased starting inocula on RC-101 activity was also assessed. As shown in Figure 3.3, 10 μM

RC-101 remained bactericidal to *S. aureus* with starting bacterial concentrations of approximately 2.5 million CFU/mL. Within 30 minutes of treating the 2.5 million CFU/mL starting concentration of D20-7 and USA300 with 10 µM RC-101, significant reductions ($p=0.04$ and $p=0.01$, respectively) in viable bacteria were observed **(**Figure 3.3**)**. Regardless of starting bacteria concentration, a continued reduction in CFU/mL was observed until undetectable levels of bacteria remained. No recovery was observed among RC-101 treated bacteria for up to nine hours.

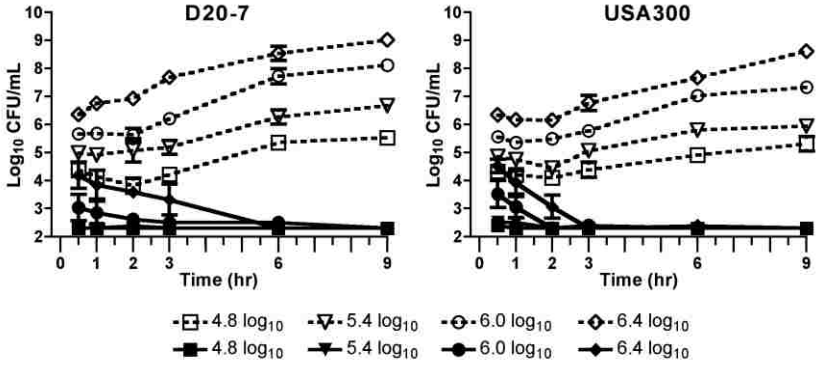

Figure 3.3. RC-101 exhibits robust anti-*S. aureus* activity in CFU assays with increasing starting inocula. Shown are bacterial growth data for *S. aureus* strains D20-7 and USA300 treated with 10 µM RC-101 under increasing bacterial starting concentrations. Starting inocula concentrations are indicated in the legend. Open symbols with dotted lines represent vehicle treated samples while closed symbols with solid lines represent RC-101 treated samples. Limit of detection equals 200 (i.e. $\log_{10}=2.3$) CFU/mL. N = 3. Error bars represent the mean ± SEM.

3.3.3 RC-101 prevents adherence of *S. aureus* to human nasal epithelia

To test the capacity for RC-101 to prevent adherence of *S. aureus* to human nasal epithelia, *ex vivo* adhesion assays were performed. In all strains analyzed, 1 µg to 2 µg of RC-

101 per tissue was sufficient to reduce *S. aureus* adherence to nasal epithelia. RC-101 concentrations of 4 µg per tissue yielded significant reductions in *S. aureus* adherence to nasal epithelial cells while 10 µg completely prevented attachment (Figure 3.4, left panels). For these assays, three nasal carriage strains were tested along with the hypervirulent USA300 strain. USA300 was analyzed as a measure of the effectiveness of RC-101 in preventing adhesion of strains frequently encountered in the clinical setting. To analyze whether RC-101 was inhibiting *S. aureus* growth on human nasal epithelial cells or was merely preventing adherence, the wash fraction was also analyzed. As can be seen in Figure 3.4 (right panels), a significant reduction in *S. aureus* growth was apparent in all isolates, similar in trend to that observed in the adhered fraction. Collectively, treatment with RC-101 exhibited a robust inhibition in human nasal epithelial cell attachment and survival of all strains of *S. aureus* tested. The simultaneous reductions in total CFUs in both the adhere and wash fractions suggests that RC-101 is exhibiting anti-*S. aureus* activity as opposed to an anti-adhesive property. Anti-adhesive property would have been expected to result in an increase in wash fraction CFUs with a simultaneous reduction in adhere fraction CFUs. Therefore, the reduction in total CFUs in both fractions suggests a more antibacterial activity by RC-101.

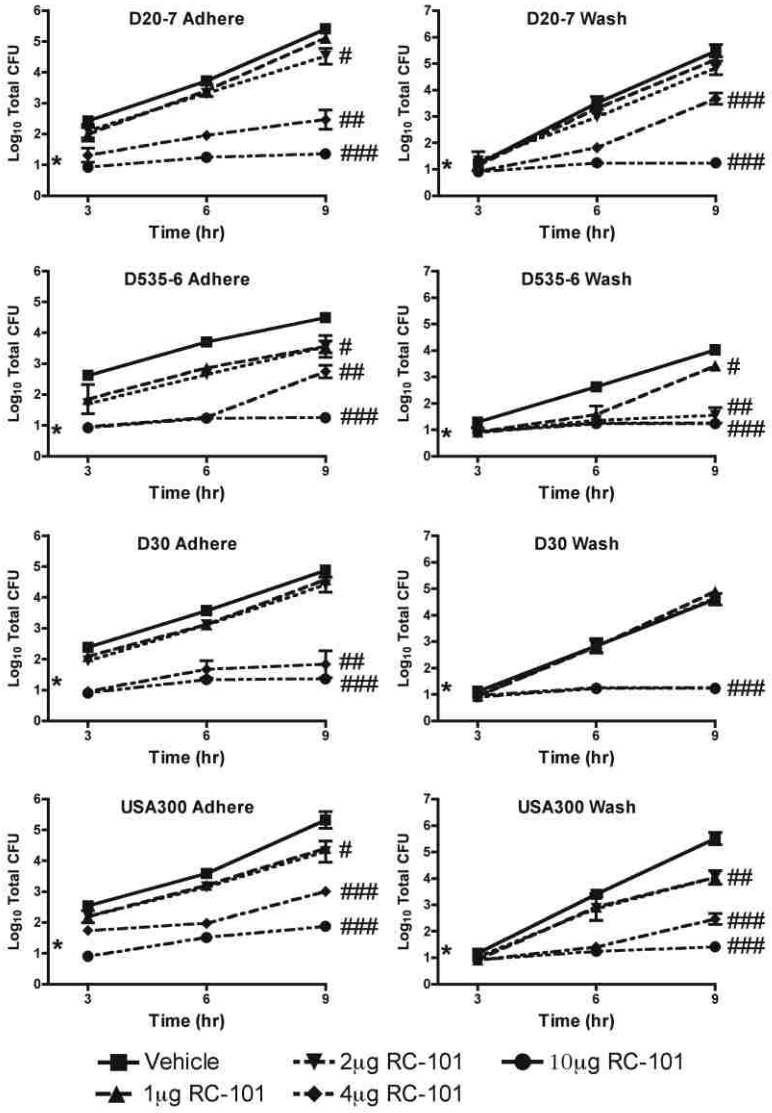

Figure 3.4. RC-101 prevents adherence and survival of *S. aureus* on human nasal epithelia. Shown are growth curve data for adhered and wash fractions from human nasal epithelia co-cultured with *S. aureus* in the presence and absence of RC-101. Nasal carriage strains are represented by names beginning with "D". USA300 represents a clinical isolate. Limit of detection for the three-hour time point is nine (i.e. $\log_{10}$=0.95) total CFU while the limit of detection for six- and nine-hour time points equals 18 (i.e. $\log_{10}$=1.26) total CFU. $N \geq 3$ for all

strains. Error bars represent the mean ± SEM. *Indicates the starting inoculum. #p < 0.05. ##p < 0.01. ###p < 0.001. P-values indicate statistical significance as compared to vehicle treated samples. For clarity of presentation, p-values are shown for nine-hour treatments only.

3.3.4 RC-101 does not exhibit cytotoxic effects to human nasal epithelia or induce inflammation

Though many agents exhibit potent antimicrobial activity, it is important that high levels of cytotoxicity do not accompany this activity. While RC-101 is effective at preventing *S. aureus* adherence and survival on human nasal epithelia, we also analyzed the cytotoxicity inflicted upon the nasal epithelia by RC-101 treatment. Nasal epithelia, under identical conditions to those used during adhesion assays, were subjected to either vehicle or 10 µg of RC-101 for a 24-hour period, after which time MTT reduction assays were performed. As shown in Figure 3.5A, no significant reduction in nasal epithelial cell viability was observed. Trypan blue dye exclusion assays were also carried out to visualize the number of viable epithelial cells after 24 hours of RC-101 treatment. As with MTT reduction assays, no cellular toxicity was observed as the result of RC-101 treatment (Figure 3.5B).

To analyze further the possibility of RC-101 cytotoxicity to nasal epithelia, increasing peptide concentrations were incubated with human nasal epithelial cells for 24 hours, after which time the viability of nasal epithelia was measured as a function of metabolic activity (indicated by MTT reduction) along with cellular apoptosis (indicated by protease release using the CytoTox-GLO™ assay). As can be seen in Figure 3.5C and Figure 3.5D, RC-101 concentrations of up to 200 µM were not cytotoxic to nasal epithelia. The notable lack of cytotoxicity to nasal epithelia may be owed to the fact that RC-101 is an analogue of a once-functional primate gene.

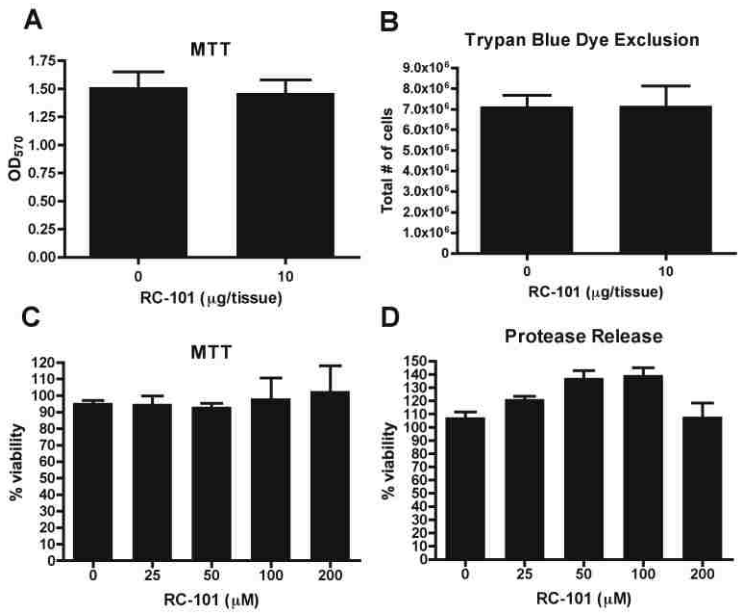

Figure 3.5. RC-101 is not cytotoxic to human nasal epithelia. Transwell inserts containing human nasal epithelia were incubated in the presence of RC-101 or vehicle for 24 hours and assayed for cellular viability using A) MTT reduction and B) trypan blue dye exclusion. Nasal epithelia (in 96-well plate format) were treated with increasing concentrations of RC-101 for 24 hours and assayed for cellular viability using C) MTT reduction and D) protease release (CytoTox-GLO™ Cytotoxicity Assay). Note that RC-101 does not exhibit cytotoxicity to nasal epithelia under any of the tested conditions. N=3 for all assays. Error bars represent the mean ± SEM.

To assess whether RC-101 promotes an inflammatory response in human nasal epithelia, expression levels of 27 proinflammatory cytokines were analyzed. Nasal epithelia treated for 24 hours with either vehicle or RC-101 (10 μg/tissue) revealed no significant difference in cytokine expression as compared to vehicle treated samples. Expression profiles for nine representative cytokines are shown in Figure 3.6. The remaining 18 cytokines tested also exhibited similar

trends to those shown in Figure 3.6, with no significant difference in expression profile observed as the result of RC-101 treatment (data not shown). Collectively, the lack of both cytotoxicity and an inflammatory response by RC-101 underscores the safety of this peptide in human nasal epithelia.

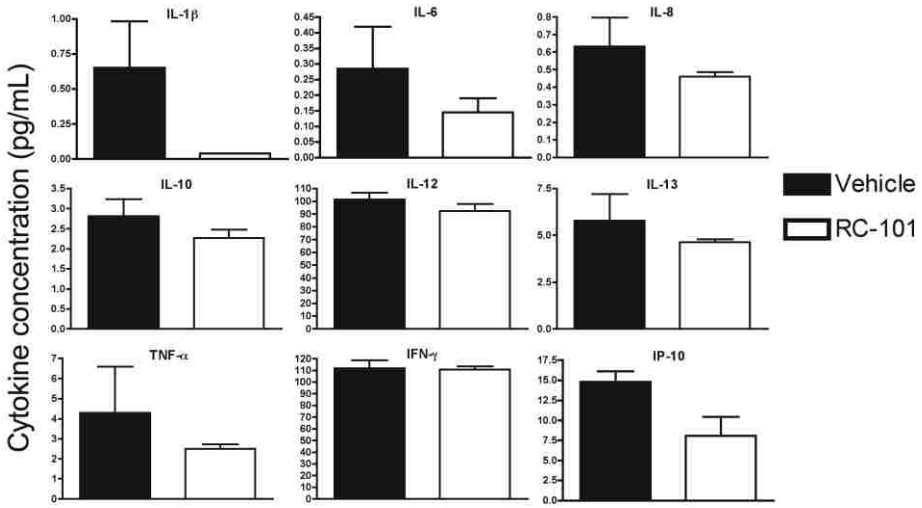

Figure 3.6. RC-101 does not stimulate a proinflammatory response in human nasal epithelial cells. Nasal epithelia were incubated with either vehicle or RC-101 (10 μg/tissue) for 24 hours and assayed for the production of human proinflammatory cytokines. Shown are nine representative plots from 27 cytokines analyzed. IL, interleukin; TNF-α, tumor necrosis factor alpha; IFN-γ, gamma interferon; IP-10, gamma interferon induced protein 10. Note that no significant difference in cytokine expression was observed as the result of RC-101 treatment. N=3. Error bars represent the mean ± SEM.

3.3.5 RC-101 prevents adherence of *S. aureus* to organotypic human airway epithelial tissues

To mimic better the physiologic state, anti-*S. aureus* assays with RC-101 were also carried out using an organotypic model of human airway epithelia. Bacterial adhesion assays using two carrier strains of *S. aureus* (D20-7 and D30) were performed on these tissues for nine hours in the presence of 4 µg and 20 µg of RC-101 per tissue. As can be seen in Figure 3.7A, 20 µg of peptide per tissue completely prevented the attachment of carrier strain D20-7, reducing adhesion by more than 2.5 $\log_{10}$ when compared to vehicle treated tissues alone. Similar data were also observed for strain D30 where 20 µg of peptide per tissue completely prevented attachment to airway epithelia, again reducing attachment by approximately 2.5 $\log_{10}$ when compared to vehicle treated tissues (data not shown).

Organotypic airway epithelia were also treated with RC-101 (20 µg/tissue) for 24- and 72-hour periods, after which time cell viability was measured by MTT reduction. As with the nasal epithelia, no reduction in cell viability was observed in the organotypic model as a result of RC-101 treatment (Figure 3.7B and Figure 3.7C). In addition to measuring the cytotoxic potential of RC-101 in organotypic airway epithelia, a panel of 27 proinflammatory cytokines were also analyzed after 24 and 72 hours of treatment to assess whether this peptide promotes an inflammatory response. As shown in Figure 3.7D, RC-101 (20 µg/tissue) did not promote inflammation in these tissues after 72 hours of treatment, reinforcing the safety of this peptide to human epithelia. Similar expression profiles were also observed for the additional 18 cytokines tested, as well as for samples treated for 24 hours with RC-101 (data not shown). The lack of both cytotoxicity and an inflammatory response imparted by RC-101 on these tissues, as well as the ability of RC-101 to exhibit antibacterial activities in these tissues support the potential of

RC-101 as a therapeutic to combat respiratory infections. Toward this end, additional research detailing the antimicrobial activity of RC-101 on a wide range of microbes will first be necessary. Previous studies have shown retrocyclin is effective against a multitude of bacteria, including *Pseudomonas aeruginosa* (28), which supports the evaluation of RC-101 in antibacterial applications other than nasal decolonization of *S. aureus* (e.g. treatment of cystic fibrosis).

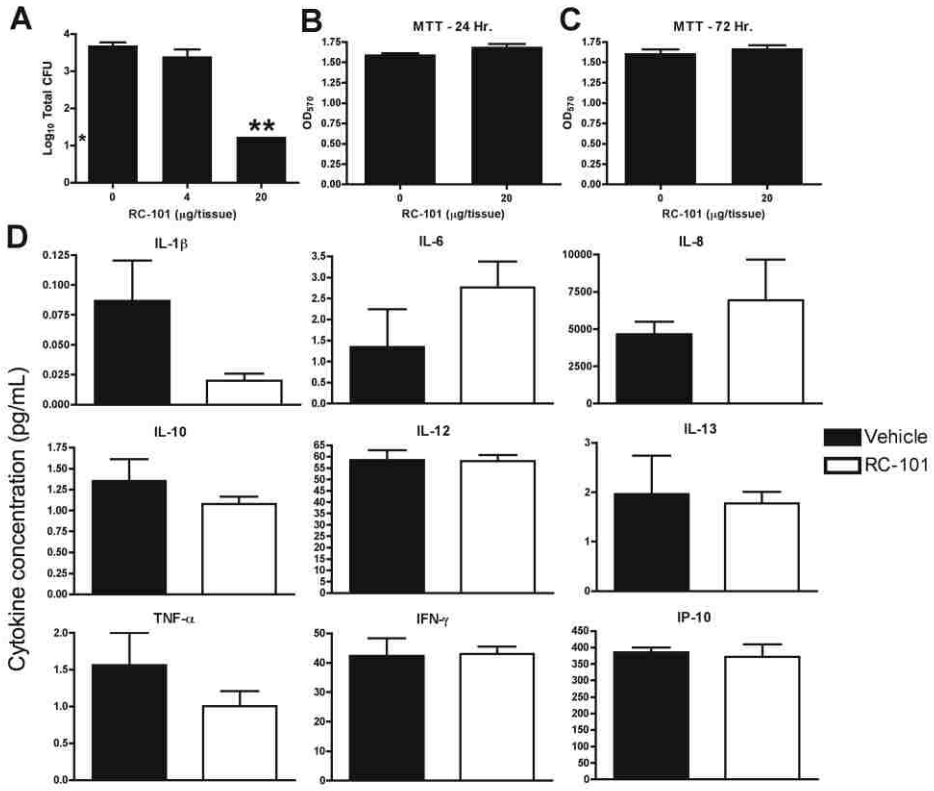

Figure 3.7. RC-101 prevents bacterial adherence to organotypic airway epithelial tissue, but does not exhibit cytotoxicity or induce a proinflammatory response. Shown are A) total CFU counts when nasal carriage strain D20-7 was inoculated for nine hours on organotypic human airway epithelia in the presence of RC-101. Organotypic airway epithelia were also treated with RC-

101 or vehicle for B) 24 and C) 72 hours and assayed for cellular viability using MTT reduction. D) Following treatment of airway epithelia with RC-101 (20 μg/tissue) for 72 hours, cytokine production was assessed. Shown are nine representative plots of the 27 cytokines assayed. Note that RC-101 does not significantly alter expression profiles for any of the tested cytokines. Limit of detection in panel A) is 18 (i.e. $\log_{10}$=1.26) total CFU. $N \geq 3$. Error bars represent the mean ± SEM. *Indicates the starting inoculum. **$p < 0.001$. P-values indicate statistical significance as compared to vehicle treated samples.

Collectively, our studies have shown the retrocyclin analogue RC-101 is a potential treatment option for the prevention and decolonization of *S. aureus* nasal carriage, and warrants further investigation in this capacity. Future studies will be instrumental in identifying the mechanism of the anti-*S. aureus* action of RC-101, as well as the efficacy of this peptide as a treatment option, or preventative measure, for *S. aureus* nasal colonization.

4. PHYLOGENETIC RELATIONSHIPS AMONG *STAPHYLOCOCCUS* SPECIES INFERRED FROM MULTILOCUS DATA

4.1 Introduction

The genus *Staphylococcus* currently contains more than 60 taxa. Many are of clinical, agricultural, and economic interest because they lead to high levels of infection among human populations or agricultural loss within the dairy, swine, and poultry industries. Moreover, multiple species within this genus are common pathogens in non-human animals and thus should be monitored with concern as these animals provide reservoirs for pathogenic bacteria (70, 141, 165). Although seemingly uncommon, host "jumping" is an important consideration of species evolution, much like what has been observed to occur with *S. aureus* and human, avian (120), and ungulate (165) hosts. As such, a thorough understanding of species relatedness is a necessity for understanding host-pathogen and pathogen-pathogen relationships within this genus (57, 62, 97).

Many previous estimates of the staphylococcal phylogeny have been based on single locus gene trees, which in many cases, exhibit marked discord. As such, robust species tree estimations have proved to be difficult. Also, due to the public health impact of staphylococcal species, many previous studies have primarily been concerned with those species responsible for human infection and as such, species that have not previously been found to directly colonize humans have been underrepresented. Thus, a robust and comprehensive assessment of staphylococcal evolution is a necessity for understanding better the evolutionary context and diversity of this important genus. The frequency with which novel species are being discovered

(six new species in 2010 alone) also heightens the necessity for a thorough understanding of staphylococcal phylogenetics.

Historically, staphylococcal species identification has been a laborious task, requiring multiple biochemical and genotypic methodologies (53). Fortunately, PCR-based analyses have become commonplace as part of the identification process of novel species (and differentiating closely related species), potentially relieving the necessity for many of the phenotypic analyses. While DNA sequencing has improved the understanding of staphylococcal (and many other microbial) phylogenies, molecular data are frequently analyzed as single genetic loci where multiple gene trees are generated using different gene fragments in isolation. The inherent caveats in this approach are that marked discord is observed between gene trees, and the lack of resolution and support, prevents any reliable estimation of the overall species phylogeny. Adding to this complication is the widespread use of the 16S rDNA fragment that exhibits only limited variability for phylogenetic reconstruction.

As with most bacterial systems, 16S rDNA continues to be the most common method for staphylococcal species identification, although its utility is limited due to high sequence similarity among different staphylococcal species (101, 186). For this reason, increased emphasis has recently been devoted towards identifying additional genes for use in species identification that offer greater taxonomic resolution between closely related species, while also limiting the incidence of misidentification. Such genes as *rpoB* (β-subunit of RNA polymerase), *tuf* (elongation factor Tu), and *dnaJ* (heat shock protein 40), have been found useful for the identification of staphylococcal species, although with the exception of one study where *dnaJ*

and *rpoB* were concatenated and assessed under a single evolutionary model (71), each has only been analyzed singularly in a phylogenetic context.

The central goal in this study was to infer a robust and comprehensive estimate of the phylogeny among staphylococcal species by utilizing evidence from multiple loci simultaneously. Here, we analyzed a large multilocus *Staphylococcus* dataset in multiple ways to thoroughly explore the phylogenetic signal in the data, and provide robust confirmatory evidence for the relationships among species. We first analyzed the combined four-gene dataset using partitioned Bayesian and maximum likelihood analyses, in which a single species tree was inferred. Such probabilistic methods of phylogeny are particularly powerful, but their accuracy can be dependent on the complexity and biological realism of the models of sequence evolution used.

There is a tradeoff between having enough parameters to accurately capture the complexity of sequence evolution in a multilocus dataset, while not having more parameters than can be accurately estimated from the data (13, 18-20). We therefore tested multiple differently partitioned model schemes to identify which best fit the multilocus dataset. Generally, we expect such partitioned model analysis of the combined (concatenated) dataset will have the best power for inferring the phylogeny of *Staphylococcus*, as long as basic assumptions of the approach are met. The most important of these assumptions is that all the underlying gene trees are the same as the species tree. There are, however, plausible scenarios whereby the gene trees and species tree are not the same (40, 41), or where systematic error in gene-tree estimation may lead to overconfidence in an incorrect species tree (17). There is some indication, however, that in such cases, maximum likelihood bootstrap support values may be more sensitive to conflicting

phylogenetic signals in the data than Bayesian posterior probability support for nodes, although both concatenated data analysis approaches are likely to experience some error (48, 79, 205).

Therefore, we also used an alternative approach to estimate relationships among species of *Staphylococcus* in which gene trees are estimated separately, and jointly considered to estimate an underlying species tree. This approach, called Bayesian Estimation of Species Trees analysis (45), thereby avoids concatenation of multiple loci, and estimates a species tree based on a model that accounts for deep coalescence of gene trees. Although this approach does not specifically model all possible scenarios that may violate the assumptions of the concatenated analysis, comparisons of results between this approach and concatenated analyses provides added perspective on the relative robustness of species-level phylogenetic inferences.

4.2 Methods and Materials

4.2.1 DNA sequence acquisition and alignment

DNA sequences for a total of four genes from 57 staphylococcal species, and two outgroup species (*Micrococcus caseolyticus* - strain JCSJ5402, and *Bacillus subtilis* - strain 168) were downloaded from NCBI's GenBank. For each species included in the analysis, sequences were specifically downloaded from the type strain. The four loci collected included the non-coding 16S rRNA gene (16S rDNA), and the three protein coding genes: *dnaJ*, *rpoB*, and *tuf*. The list of all species analyzed in this study with the accession numbers for each of the four gene fragments is given in Appendix B, Table B.1.

Nucleotide sequences were aligned using ClustalW in MEGA 4.1 (184), with manual adjustment to ensure that complete codons remained intact for downstream analyses. The

concatenated alignment totaled 3 521 nucleotides for each species. Regions of high variability were omitted from the alignments because assessment of homology was questionable (18). Secondary structure predictions (i.e. stem and loop regions) for 16S rDNA fragments were estimated using the RNAalifold approach (8, 76). Nucleotide diversities and species divergence calculations were performed using MEGA 4.1 (184) and DnaSP v5 (116).

4.2.2 Nucleotide model selection

Models of nucleotide evolution for each gene and nominal partition of the data were estimated using jModelTest v0.1.1 (64, 151) based on Akaike Information Criterion (AIC). For the purpose of model testing (and later partitioned Bayesian analyses) we divided the dataset by gene, and into biologically relevant subsets: coding versus non-coding gene fragments, codon position, and stem versus loop secondary structures (for 16S rDNA). These individual partitions, and the best-fit evolutionary model selected for each partition, are shown in Appendix B, Table B.2.

For analyses of the combined data with partitioned models, we formulated nine different partitioning schemes. These were designed to provide a hierarchical spectrum of model complexity, and parameter richness, with increasing partitioning of biologically reasonable sets of the data (Table 4.1). The simplest model (MB1) was a single evolutionary model (GTR + ΓI) fit to the entire dataset followed by additional models (MB2-MB9) that were created by the addition of dataset partitions among and within non-coding and coding gene fragments (Table 4.1).

Table 4.1. Description of alternative model partitioning strategies tested for fit to the combined nucleotide data.

Model name	# of partitions	# of free model parameters	Description of model partitions
MB1	1	10	Single model for concatenated dataset
MB2	2	13	16S; All protein coding gene fragments (*dnaJ; rpoB; tuf*)
MB3	4	29	Independent partition for each gene fragment (*16S; dnaJ; rpoB; tuf*)
MB4	7	48	16S; two partitions for each gene fragment (codon positions 1 and 2; codon position 3)
MB5	8	62	16S, stems; 16S, loops; two partitions for each gene fragment (codon positions 1 and 2; codon position 3)
MB6	10	78	16S; three partitions for each gene fragment (codon positions 1, 2, and 3, separately)
MB7	11	92	16S, stems; 16S, loops; three partitions for each gene fragment (codon positions 1, 2 and 3, separately)
MB8	3	26	16S, stems; 16S, loops; All protein coding gene fragments (*dnaJ; rpoB; tuf*)
MB9	5	43	16S, stems; 16S, loops; Independent partition for each protein coding gene fragment (*dnaJ; rpoB; tuf*)

4.2.3 Bayesian phylogenetic analysis

Bayesian inference (BI) was carried out using the Metropolis-Hastings coupled Markov chain Monte Carlo method in MrBayes v3.1.2 (83, 160) and BEST v2.3.1 (118). All Bayesian phylogenetic analyses performed in this study were carried out using the STOKES IBM High Performance Computing Cluster at the University of Central Florida. MPI-enabled versions of MrBayes v3.1.2 and BEST v2.3.1 were compiled and run in parallel (3). For each BI run, gaps in alignments were treated as missing data. For each analysis, two independent BI runs were carried out using random starting trees with one cold chain and three heated chains (following program defaults). Each model was assessed in triplicate with summary statistics being estimated from all runs.

In addition to performing BI runs on the unpartitioned multilocus dataset (using the evolutionary model specified by AIC), eight additional models were assessed where independent models of evolution were applied to different nucleotide regions within the combined dataset (refer to nucleotide model selection section). This was achieved by using the "unlink" command in MrBayes v3.1.2. Each BI run consisted of 4 million generations with every 100 steps being sampled. As verified using Tracer v1.5 (157), stationarity was reached in all BI runs prior to 500 000 generations and a conservative burn-in of 1 million (25%) generations was performed.

In addition to reconstructing phylogenies using MrBayes v3.1.2, Bayesian phylogenetic reconstruction was also performed using BEST v2.3.1, which is a modified version of MrBayes. In BEST, each gene was assigned a single model of nucleotide substitution (based on AIC, estimated in jModelTest). BI runs using BEST v2.3.1 were performed using 4 million generations with sampling every 100 steps and a burn-in of 1 million generations.

4.2.4 Assessment of BI runs

All partitioning strategies were run in triplicate to verify reproducibility. Subsequently, BI runs under each model were assessed using multiple criteria to determine the success of each model and the overall best-fit model. Bayes factors (BF; $2\Delta\ln B_{10}$) were calculated from estimates of the harmonic mean of the posterior distribution of cold chain likelihoods. Consistent with previous reports (13, 19, 143), we set a cutoff of BF>10 to support one model over another. Akaike weights (Aw) (1) were also used to identify best-fit partitioned models (20). Initially AIC values were calculated by the equation AIC = -2lnL + 2k where k equals the total number of free parameters within the model. For small samples sets, where the sample size (n) to free parameter (k) ratio is <40, it has been suggested that a small-sample bias adjustment

be applied to the AIC calculation, thus calculating AIC_c instead (89, 194). The sample size of the staphylococcal dataset (with outgroups) is 59 and the minimum number of free parameters was 10 for model MB1. As such, the n/k ratio was always <40, so we calculated the AIC_c instead. The equation for $AIC_c = -2\ln L + 2k + 2k(k+1)/n-k-1$. The ΔAIC_c was then calculated by subtracting the model with the minimum AIC_c (AIC_{cmin}) (i.e. highest lnL) from the ith model using the equation $\Delta AIC_{ci} = AIC_{ci} - AIC_{cmin}$. Following calculations of the ΔAIC_c for each model, Aw were calculated using the equation $Aw = e^{(-\Delta AICci/2)} / \sum e^{(-\Delta AICci/2)}$. By this equation, the relative likelihood of a model given the data is normalized over all models and thus, the greater the Aw for a given model, the greater the relative support for that model (19).

Further assessment of model performance was based on examining the output of model parameters and carried out by analyses of multiple additional features. Posterior distributions of parameters and analysis of trace plots were assessed for failed convergence and stationarity using Tracer v1.5 (157). Also, because model overparameterization has been linked to estimates of tree length in partitioned Bayesian analyses (123), we also compared tree length estimates among runs.

4.2.5 Maximum likelihood analysis

Phylogenetic reconstruction using maximum likelihood (ML) analysis was carried out using the program GARLI v.2.0 (211), using default parameters except where specified. Phylogenetic estimates using ML were performed using both the combined, unpartitioned dataset as well as the combined dataset partitioned by locus (Appendix B; Table B.2). Five ML search replicates were run for each dataset using random starting trees, and up to five million generations were employed for each run unless the scoring topology lnL did not improve by ≥

0.01 for 20 000 generations, in which case the run was terminated prematurely and the next bootstrap replicate was begun. Two hundred bootstrap replicates were conducted for each run and consensus trees were generated using the SumTrees v.3.0 software which is part of the DendroPy v.3.7 phylogenetic computing library (178). Likelihood ratio tests (LRTs) (51, 82) were performed to compare competing model partitioning schemes, M0 and M1. Statistical support for model M0 over M1 (or vice versa) was assessed using the Chi-square distribution for *q* degrees of freedom (df) where *q* equals the difference in the number of free parameters between model M0 and M1 (df = 19 in this study) (82).

4.3 Results

4.3.1 Gene fragments used for analyses contain differing degrees of variability

Among the four gene fragments analyzed in this study, 3 521 nucleotides were included (1 481 from 16S rDNA, 816 from *dnaJ*, 474 from *rpoB*, and 750 from *tuf*) for 59 different taxa. The dataset contained 1 016 parsimony-informative sites and 2 142 conserved sites. The nucleotide diversity of the 16S rDNA fragment was 0.029 substitutions (subs.) per site, while that for *dnaJ, rpoB,* and *tuf* was 0.241, 0.147, and 0.097 subs. per site, respectively. The lowest interspecies divergence was between *S. pseudintermedius* and *S. delphini* (0.014 subs. per site). The highest estimated evolutionary divergence within the complete dataset was between *S. piscifermentans* and the outgroup species, *B. subtilis* (0.266 subs. per site), while the highest level among staphylococcal taxa was between *S. piscifermentans* and *S. vitulinus* (0.182 subs. per site).

4.3.2 Dataset partitioning improves likelihood estimates of Bayesian phylogenetic analyses

Regardless of partitioning strategy employed, all Bayesian inference (BI) runs yielded highly reproducible phylogenetic inferences (Appendix B; Figure B.1). Within BI runs, log-likelihood (lnL) estimates rapidly reached stationarity and convergence. Log-likelihoods ranged from -38830.66 (MB1) to -37421.36 (MB7) with intermediate lnL generally increasing with partition complexity (Figure 4.1). Increased dataset partitioning improved posterior lnL values, except where less partitioning occurred within protein coding regions, but 16S rDNA was partitioned by stems and loops (Figure 4.1, compare MB5 and MB6).

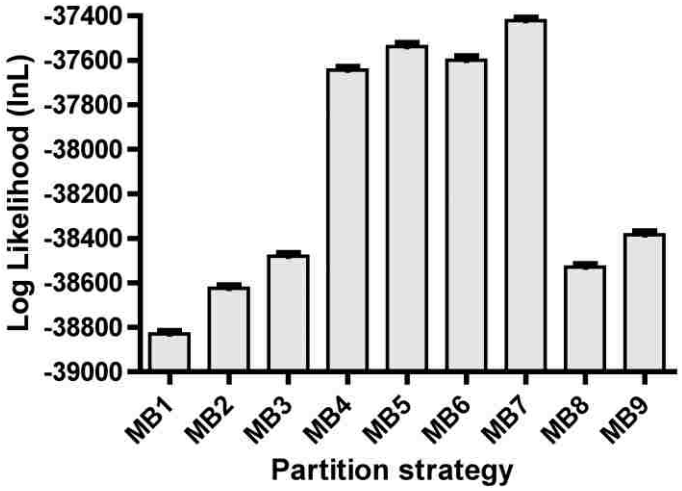

Figure 4.1. Increasing model complexity improves posterior likelihood estimates of phylogeny. Shown are log-likelihood plots comparing each partitioning strategy assessed in this study. Note that as model complexity increases so do posterior lnL. Error bars represent the mean ± 95% confidence interval.

Dataset partitioning for BI runs ranged from the most simple (unpartitioned) to highly complex (11 partitions; Table 4.1). Initial assessments of Bayes factors (BF; $2\Delta\ln B_{10}$) were used to compare topological likelihoods across each different model. As shown in Table 4.2, a large disparity between the lnL from various partitioning strategies was observed. Partitioning strategy MB7 yielded the highest lnL (Figure 4.1) with a BF>230 that of the next best model (MB5) and >2800 compared to the unpartitioned model (MB1). Model MB7 was the most complex strategy (11 different partitons) with a separate model for each codon position of each protein-coding gene, as well as stem versus loop regions of 16S rDNA (Table 4.1). The model with the second highest likelihood was MB5 whereby the 16S rDNA fragment was again partitioned by stem and loop position, however, only two independent partitions were applied to each individual protein coding gene fragment (codon positions 1 & 2; and codon position 3). Using AIC_c for the Aw calculation identified model MB5 as the best-fit model (Aw=1.000; Table 4.2). Thus, based on lnL-centric criteria, models MB5 and MB7 are the preferred models.

Inspection of TL identified that the more highly partitioned models (MB4-MB7) yielded TLs ≥15 as compared to the less partitioned models (MB1-3; MB8-9) where TLs were ≤ 7 (Appendix B; Figures B.2 and B.3). The more highly-partitioned model runs with high TLs also tended to show very high TL variance among generations, resulting in quite broad TL posteriors (Appendix B; Figures B.2 and B.3). Considering this evidence for unreliability in the more highly partitioned model runs, we tempered our choice of partitioning scheme. A combination of lnL (BF and Aw) and TL reliability criteria suggest that MB8 is the preferred partitioned model, since it had better lnL than other models (e.g., MB1-2) while resulting TL estimates were apparently uninflated and of low variance (Appendix B; Figures B.2 and B.3). Hereafter, we

discuss results based on the BI runs from model MB8, and identify any notable differences between this model and others (particularly MB5 and MB7).

Table 4.2. Bayes factors and Akaike weights reveal differences in model fitness for the different partitioning strategies applied to the multilocus dataset.

M1↓	[a]$2\Delta\ln B_{10}$									Akaike
M0→	MB1	MB2	MB3	MB4	MB5	MB6	MB7	MB8	MB9	Weight
MB1	---	**412.11**	**700.90**	**2371.08**	**2584.86**	**2462.04**	**2818.59**	**603.91**	**891.39**	0.000
MB2	-412.11	---	**288.79**	**1958.97**	**2172.74**	**2049.93**	**2406.48**	**191.80**	**479.28**	0.000
MB3	-700.90	-288.79	---	**1670.18**	**1883.96**	**1761.14**	**2117.70**	-96.99	**190.49**	0.000
MB4	-2371.08	-1958.97	-1670.18	---	**213.77**	**90.96**	**447.51**	-1767.17	-1479.69	0.000
MB5	-2584.86	-2172.74	-1883.96	-213.77	---	-122.82	**233.74**	-1980.95	-1693.46	1.000
MB6	-2462.04	-2049.93	-1761.14	-90.96	**122.82**	---	**356.56**	-1858.13	-1570.65	0.000
MB7	-2818.59	-2406.48	-2117.70	-447.51	-233.74	-356.56	---	-2214.68	-1927.20	0.000
MB8	-603.91	-191.80	**96.99**	**1767.17**	**1980.95**	**1858.13**	**2214.68**	---	**287.48**	0.000
MB9	-891.39	-479.28	-190.49	**1479.69**	**1693.46**	**1570.65**	**1927.20**	-287.48	---	0.000

[a]Positive Bayes factors ($2\Delta\ln B_{10}$) support model M0 over model M1 and negative values support model M1 over model M0. Bayes factor support values >10 are shown in bold.

4.3.3 Bayesian inference of partitioned datasets reveals highly supported relationships among staphylococci

Regardless of the model under which the staphylococcal dataset was analyzed, high overall nodal support was observed for nearly all nodes in the tree. Tree topologies were highly concordant between different partitioned model schemes, with only a single topological inconsistency between models. Beside this single topological difference, nodal support differed by very little among models ($Pp \leq 0.02$), with only two cases (MB1 and MB6) in which a single node differed by a Pp=0.05. In addition to the placement of *S. devriesei* in Figure 4.2A, this species was estimated to form a clade with *S. lugdunensis* under four models (MB2-4, and MB6) while also being estimated to diverge after *S. lugdunensis*, forming a single species sister lineage

to a clade containing *S. haemolyticus* and *S. hominis* under two models (MB5 and MB7; data not shown). Nodal support for these alternative relationships was quite low (avg. Pp=~0.64), however, in comparison to the support of *S. devriesei* forming a clade with *S. haemolyticus* (Pp=0.85; Figure 4.2A).

Figure 4.2. Bayesian MCMC analysis estimates a strongly supported staphylococcal phylogeny. Shown is A) a 50% majority rule phylogram from BI runs under the combined, partitioned dataset in MrBayes and B) a consensus cladogram from BI runs analyzing the unconcatenated

dataset using BEST. Both inferences of topology are highly concordant with only minor variations. Red boxes highlight topological differences in B). Branch lengths in B) are not informative. Numbers represent posterior probabilities with grey-filled circles representing a posterior support of 1.00.

Bayesian concatenated phylogenetic estimates supported strongly (Pp=1.00) the separation of staphylococcal species into two deeply-diverging major clades (Figure 4.2A). One of the two clades contained all of the oxidase positive staphylococcal species (frequently referred to as the Sciuri group), with the second group containing all other oxidase negative staphylococcal species (Figure 4.2A). The single lineage *S. auricularis* formed the sister group to all other members of this second group, with the next most basally-diverging lineage in this clade including the following species: *S. simulans, S. condimenti, S. carnosus* (both subspecies), and *S. piscifermentans* (Pp=1.00). The subspecies of *S. carnosus* proved to cluster tightly together, as expected, and formed the sister group to *S. condimenti*.

The next major divergence within the staphylococcal tree was that of a strongly supported clade (Pp=1.00) containing the pathogenic species *S. saprophyticus* (Figure 4.2A). This clade contained many members of the polyphyletic group of coagulase negative, novobiocin resistant species, and included the recently described species *S. massiliensis* (2) and *S. pettenkoferi* (193). Following this divergence, species of heightened clinical significance diverged, including *S. aureus, S. epidermidis, S. warneri, S. haemolyticus* and *S. lugdunensis,* which formed a well-supported clade (Pp=1.00) (Figure 4.2A). We also found that the most recently discovered *Staphylococcus* species, *S. agnetis* (187) formed a strongly supported clade (Pp=1.00) with *S. hyicus*, for which *S. chromogenes* was the sister lineage.

4.3.4 Broad agreement between concatenated and unconcatenated analyses

Estimation of staphylococcal phylogeny was also performed on the unconcatenated dataset using Bayesian Estimation of Species Trees (BEST) analysis (118). The BEST species tree estimate (Figure 4.2B) was nearly identical in topology to the BI concatenated data result, with overall high nodal support for all but four nodes which received Pp<70. Overall, the nodal Pp support values from MrBayes and BEST were generally quite similar, with a clear linear trend between Pp for one method versus the other (Figure 4.3).

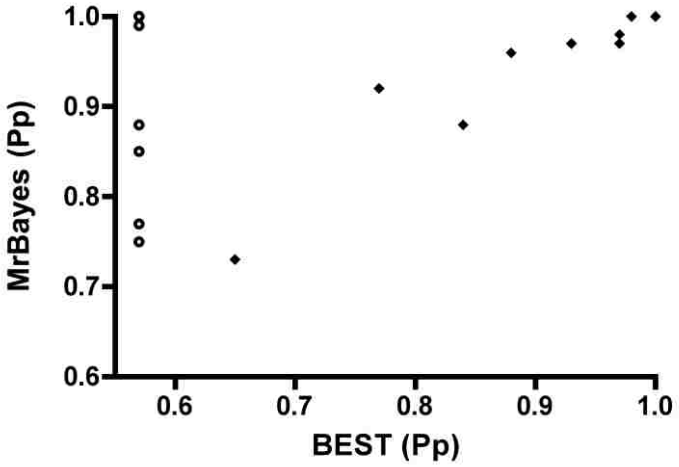

Figure 4.3. Phylogenetic relationships and nodal support are highly similar between MrBayes and BEST. Shown is a scatter plot comparing the differences in posterior probabilities (Pp) between MrBayes and BEST for identical nodes (Figure 4.2). Open circles represent discordant nodes between MrBayes and BEST where MrBayes has been taken as the reference. Posterior probabilities from discordant nodes could not be matched between methodologies and thus, BEST Pp for those nodes are not relevant. Open circles are only present to reveal the MrBayes Pp at discordant nodes. Note that the overall nodal support between MrBayes and BEST are in strong agreement.

There were, however, several alternative relationships resolved in the BEST tree (Figure 4.2B, red-indicated nodes) relative to the partitioned-model BI tree (Figure 4.2A). These included the earlier (more basal) divergence of *S. felis* and the clade containing *S. chromogenes, S. hyicus,* and *S. agnetis* in the BEST results, although this was weakly supported (Pp<0.50; Figure 4.2). BEST analysis also resolved (with weak support) a slightly different arrangement with *S. devriesei* and *S. lugdunensis* sharing a clade (Pp=0.56), contrasting the combined data BI analysis that suggested *S. devriesei* and *S. haemolyticus* form an exclusive clade (Figure 4.2). An additional discordant node between Bayesian methodologies involved the relationship between *S. condimenti* and *S. carnosus* (both subspecies) (Figure 4.2B). This relationship inferred by BEST is the only difference compared to the combined data BI tree in which both results have conflicting relationships with strong Pp support. While the combined BI tree inferred a clade containing both subspecies of *S. carnosus* (Pp=1.0), the BEST tree inferred a clade containing *S. condimenti* and *S. carnosus carnosus,* with *S. carnosus utilis* as its sister lineage (Pp=1.0 for both clades; Figure 4.2).

The concatenated-data maximum likelihood estimation of the staphylococcal phylogeny was consistent with reconstructions from concatenated BI and BEST methods (Figure 4.4). Maximum likelihood inference under a single evolutionary model yielded a lnL of -39186.39 while partitioning the concatenated dataset by individual gene yielded a lnL = -36632.34. The likelihood-ratio test supported the partitioned dataset as the best-fit model (p<0.0001; likelihood-ratio (-2ΔlnL) = 5 108; degrees of freedom (df) = 19). Topologies estimated under both models were identical except for a single discordant node: *S. devriesei* formed a single-species sister taxon to *S. haemolyticus* and *S. hominis* in the unpartitioned dataset (Bootstrap support

(BS)=59%), while in the dataset partitioned by individual gene, *S. devriesei* shared a clade with *S. haemolyticus* (BS=72%). Within the ML topology (Figure 4.4), the clade containing *S. muscae, S. rostri, and S. microti* diverged more deeply from the larger clade containing *S. felis, S. hyicus,* and *S. intermedius* (similar to the BI concatenated analysis; Figure 4.2A) as opposed to *S. felis* being the most divergent species from this clade and forming a sister lineage to the remaining clades containing *S. microti, S. hyicus,* and *S. intermedius* (as was estimated by BEST; Figure 4.2B). Among the oxidase containing species clade, ML estimated a more basal divergence of *S. lentus* and *S. stepanovicii* than was estimated under either of the Bayesian methodologies (Figure 4.4).

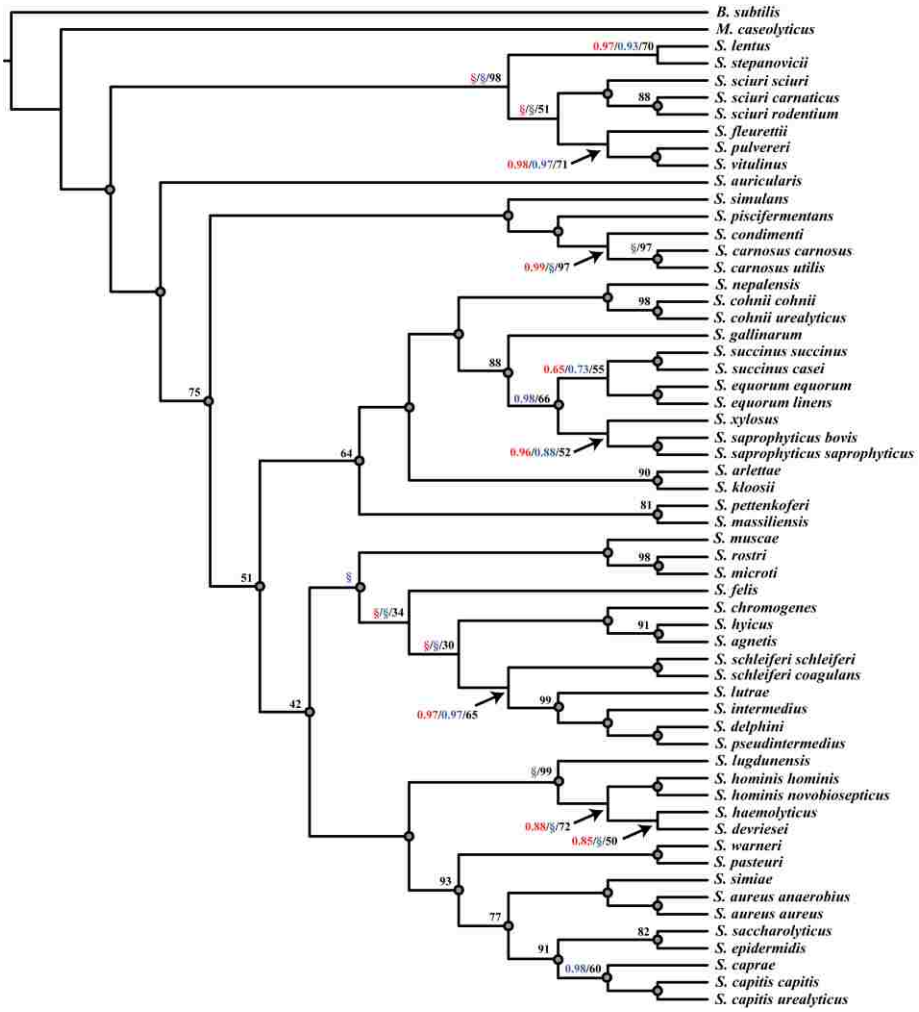

Figure 4.4. Maximum likelihood cladogram of staphylococcal species yields a highly unified topology, similar to that estimated in BI runs. Shown is a ML cladogram obtained from the assessment of the locus-partitioned dataset (similar to MB3) using GARLI v.2.0 (211). The consensus cladogram was generated from 200 bootstrap replicates with five ML search replicates per bootstrap. Nodes receiving Pp=1.00 or BS=100% are indicated by grey-filled circles; otherwise, MrBayes support is shown in red text (Pp), BEST support is shown in blue text (Pp), and ML support is shown in black text (BS). Clades that were not present in MrBayes or BEST are indicated by a red or blue §, respectively.

4.4 Discussion

4.4.1 Using multilocus data to infer the *Staphylococcus* phylogeny

Staphylococcus is a species-rich genus of importance from both a human health and economic perspective. Greater than 60 taxa of *Staphylococcus* exist, although a comprehensive study of species phylogeny within this genus was lacking. Most assessments of the staphylococcal phylogeny are provided when novel species are identified, and are often based on trees estimated from a single locus (16, 119). The predominant locus of choice for staphylococcal phylogenetics, as in most other studies of bacterial phylogenetics, is 16S rDNA (177). When used by itself, however, this locus does not provide adequate resolution for determining species relatedness (9, 57). Difficulties often arise in identifying species relationships because 16S rDNA sequences can be nearly identical between staphylococcal species (101, 186). This has led to recent reports utilizing sequence data from more variable loci to allow better species identification and phylogenic reconstructions for the group (9, 44, 57, 125, 152, 169). Interestingly, however, only rarely have multiple loci been used together in joint analyses for phylogenetic inference within this genus (71).

We have found in this study that Bayesian and maximum likelihood analysis of multilocus data yields high-resolution species trees with overall strong nodal support values for relationships among *Staphylococcus* species. We also found that partitioned-model analysis of the combined dataset, versus the concatenation-free analysis using BEST, produced near-identical estimates of phylogeny. Collectively, the multiple methodologies employed provide confirmatory evidence for the robustness of our estimated *Staphylococcus* phylogeny. To extract

as much accurate phylogenetic signal out of our multilocus dataset, we applied and tested various partitioned-model schemes for analysis of the concatenated data. Despite likelihood-based statistical evidence favoring highly partitioned models, we observed hallmarks of tree length estimate inconsistency, as is known to occasionally occur among parameter rich models (123, 124). Thus, we focused on partitioned models that appeared to have more reasonable tree length estimates with lower variance. It is notable that we found model partitioning (especially within 16S rDNA fragment stem and loop regions) to result in a marked increase in model fit with some non-trivial changes in topological support. These findings suggest that partitioned models may be of heightened use in other microbial phylogenetic studies, particularly ones utilizing 16S rDNA.

Our phylogenetic reconstructions based on the multilocus staphylococcal data confirmed many previous hypotheses of relationships, while also suggesting some novel relationships among members of the group. Historically, staphylococcal species have been clustered into between four and eleven species groups (57, 102, 103, 112, 152, 182). Most of these groupings, however, were inferred based on a single locus with a small number of staphylococcal taxa. Phylogenetic estimates from this study supported the separation of staphylococcal species into six major staphylococcal species groups comprised of 15 cluster groups (Figure 4.5). We use our Bayesian, partitioned-model concatenated data estimate (i.e., Figure 4.2A) as the phylogeny for illustrating evolutionary groupings of *Staphylococcus*, and indicate on this tree where BEST and ML concatenated inferences differed (Figure 4.5). Wherever possible, we have attempted to name cluster groups and species groups following the original nomenclature put forth by

Takahashi *et al*. (182), while recognizing only evolutionarily distinct, monophyletic groupings based on our estimates of phylogeny.

4.4.2 The phylogeny and classification of *Staphylococcus*

Consistent with previous studies (2, 44, 112, 181, 182), our analyses identified the monophyletic group containing the novobiocin-resistant, oxidase positive species (Sciuri group; Figure 4.5, blue cluster group) as the sister group to all other *Staphylococcus*. This cluster group also contains the recently discovered species, *S. stepanovicii* (71). Within this group, we inferred a close relationship, with little sequence divergence, between *S. vitulinus* and *S. pulvereri* (BI and BEST Pp=1.00; BS=100%), potentially supporting the reclassification of *S. pulvereri* as a later synonym of *S. vitulinus* (181). After the basal divergence of the Sciuri group, the second lineage to diverge from the remaining staphylococcal lineages was the oxidase negative Auricularis group, containing only *S. auricularis* (Figure 4.5). Our phylogeny therefore suggests that cytochrome C oxidase was lost in *Staphylococcus* sometime in the common ancestor of *S. auricularis* and the remaining *Staphylococcus* species, after their divergence from the Sciuri group (Figure 4.5, red star).

Our phylogenetic placement of *S. auricularis* as the sister lineage to all non-Sciuri group staphylococci is unique to our study, and we find strong unilateral support for this inference across all of our analyses (Pp=1.00 for both Bayesian analyses and ML BS=99%). Based on 16S rDNA alone, Takahashi *et al.* (182) estimated that *S. auricularis* shared a common ancestor with the *S. saprophyticus, S. lugdunensis, S. haemolyticus, S. warneri, S. epidermidis* and *S. aureus* cluster groups. More recently, Ghebremedhin *et al.* (57) estimated a similar relationship to that of Takahashi *et al.* based on 16S rDNA alone. Analyses of subsequent gene fragments, however,

yielded varying relationship estimates for *S. auricularis,* and no previous studies have found particularly strong support for the placement of this lineage. For example, Ghebremedhin *et al.* (57) recovered bootstrap support of 31% for a clade containing *S. auricularis* and *S. kloosii* based on 16S rDNA, although average BS support across their tree was particularly low, at BS=52.1%. Similarly, *S. auricularis* was placed as the sister lineage to *S. kloosii* plus the *S. saprophyticus* group, with BS=25% based on analysis of 16S rDNA by Takahashi *et al.* (182).

We inferred that the next lineage of *Staphylococcus* to diverge was the Simulans species group (Figure 4.5), which contains four species that are all novobiocin susceptible and coagulase negative. For consistency with previous nomenclature (57, 182), we refer to this clade as the Simulans-Carnosus cluster group and the species group as the Simulans group (Figure 4.5). Our estimate of relationships among species of this group agree with previous studies, although the inclusion of *S. condimenti* in our trees is novel (57, 182). We inferred a single clade (Simulans-Carnosus cluster) containing the novobiocin susceptible, coagulase negative species, *S. simulans, S. condimenti, S. carnosus* and *S. piscifermentans.* It is notable that while *S. carnosus carnosus* and *S. carnosus utilis* formed a well supported clade (with *S. condimenti* as its sister taxon) in our concatenated analyses, our BEST analysis resolved *S. carnosus carnosus* and *S. condimenti* as forming a clade, with *S. carnosus utilis* as the sister taxon. Based on this result, we conducted independent gene analyses (not shown) that suggested that 16S rDNA seemed to place the two subspecies of *S. carnosus* distantly from one another, while other genes clustered them together. This highlights a strength in using multiple approaches with differential sensitivity, and suggests that in the past, processes such as horizontal transfer, introgression, or incomplete lineage sorting may have occurred that account for this apparent discrepancy among genes.

Following the split of these three early-diverging lineages, the remaining *Staphylococcus* species diverged into three large species groups. The first of these to diverge from the remaining was the Saprophyticus species group (Figure 4.5), which we inferred consists of four cluster groups. Within this species group, the Pettenkoferi-Massiliensis cluster group contains novobiocin susceptible species while all of the remaining members of the Saprophyticus group are novobiocin resistant. Thus, it seems that an alternative gyrase B gene conferring novobiocin resistance may have been acquired in this clade sometime after the Pettenkoferi-Massiliensis cluster group diverged from the rest of the Saprophyticus species group. Based on analysis of 16S rDNA, Al Masalma *et al.* (2) reported the newly discovered species *S. massiliensis* to be a member of the Simulans group, although they failed to recover this relationship in analyses of the *dnaJ, rpoB,* and *tuf* genes, where they instead placed it with *S. pettenkoferi* as we have here. It is also notable that the close relationship between these coagulase-negative species was also suggested based on their phenotypic similarities across a range of biochemical tests (2). Additionally, in the Saprophyticus cluster group, we inferred a close relationship between *S. equorum, S. succinus, S. saprophyticus,* and *S. xylosus* with *S. gallinarum* as the sister lineage to these four species. The placement of *S. gallinarum* in other studies is variable, but on multiple occasions has clustered with the Arlettae-Kloosii group (57, 112, 152, 182, 187). This alternative placement of *S. gallinarum* seems reasonable as we find the Arlettae-Kloosii cluster group to be closely related to the Saprophyticus cluster group (Figure 4.5).

The Epidermidis-Aureus species group contained five cluster groups, including the most common taxa of heightened clinical significance (57). In general, our estimates of relationships among these species are consistent with previous reconstructions (182, 187). Relationships

within the Haemolyticus cluster group also agree with previous estimates (182), although the placement of the recently discovered coagulase-negative bovine strain, *S. devriesei*, remains an open question (179). The original report of *S. devriesei* agrees with our concatenated BI and ML results placing it in a clade with *S. haemolyticus*, albeit with weak support (BI Pp=0.85, ML BS=72%). Our BEST analysis, however, inferred *S. devriesei* forms a clade with *S. lugdunensis*, although again with weak support (Pp=0.56).

Lastly, the Hyicus-Intermedius species group contained species of the "*S. hyicus-S. intermedius* cluster group" originally proposed by Takahashi *et al*. (182) based on a 16S rDNA dataset, and additional studies have found similar estimates of relationships based on analyses of other loci (57, 103, 112, 141, 152, 158). The limited number of taxa assessed in these studies has, however, prevented a more detailed understanding of species relationships within this species group prior to our analysis here. Moreover, recent novel species discovery (in particular *S. rostri* (158), *S. microti* (141), and *S. agnetis* (187)) has also contributed to the enhanced diversity of the Hyicus-Intermedius group. We have divided this species group into three cluster groups based on their phylogenetic relationships, which is also supported by their phenotypic diversities (Figure 4.5). Species among the Intermedius cluster group are all coagulase positive, excepting *S. schleiferi schleiferi*. Interestingly, *S. schleiferi coagulans* is coagulase positive, consistent with the other members of this cluster group, implying a recent loss in *S. schleiferi schleiferi*. In contrast, the Muscae cluster group contains only coagulase negative species (*S. muscae, S. rostri,* and *S. microti*). Within the last year, both *S. rostri* (158) and *S. microti* (141) were discovered and found to cluster with *S. muscae*, thus altering previously known relationships within this species group. The Hyicus cluster group is coagulase-variable,

including coagulase positive (*S. hyicus*), negative (*S. chromogenes, S. felis*), and variable (*S. agnetis*) species (Figure 4.5, red cluster group).

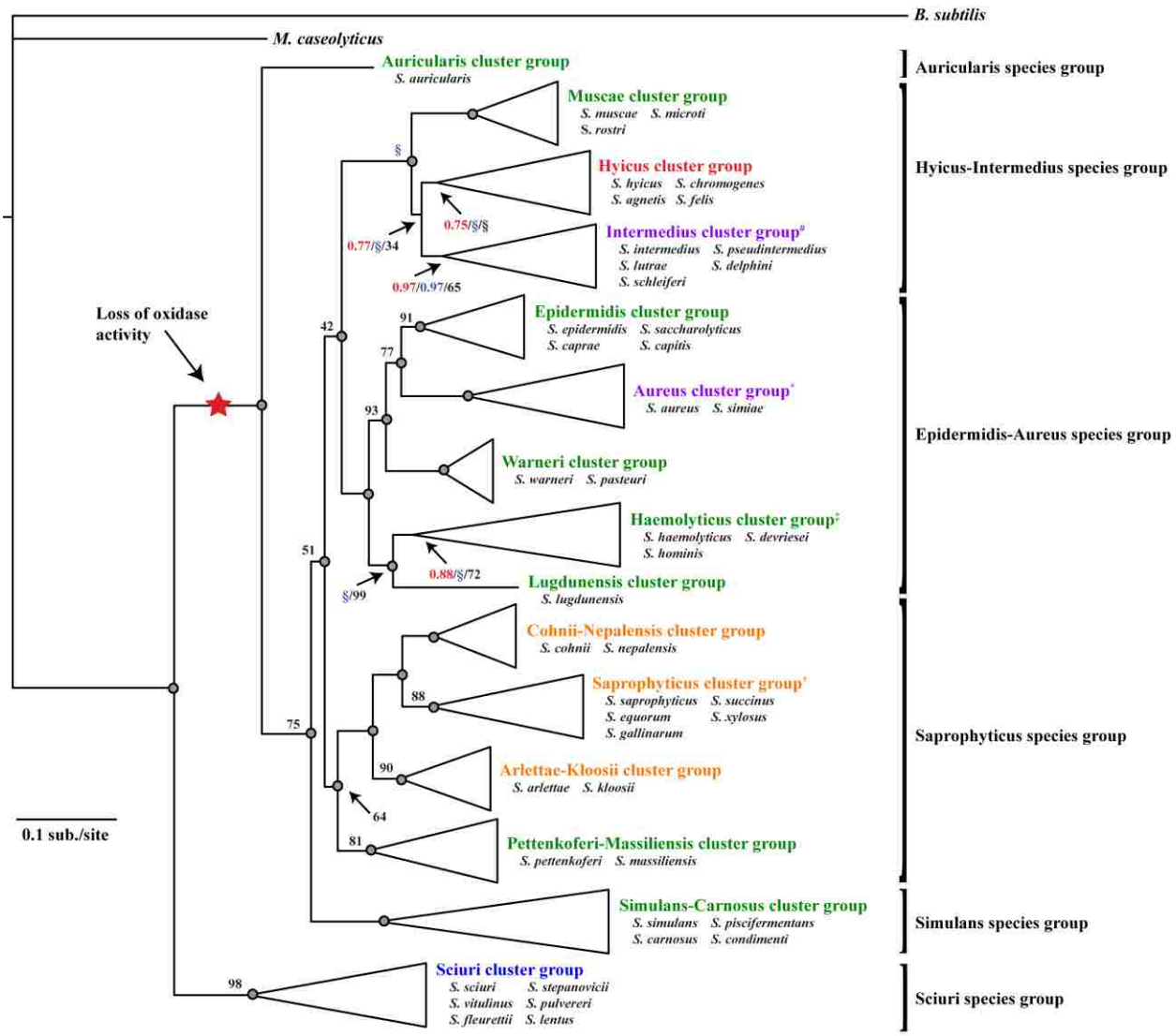

Figure 4.5. Staphylococcal species can be combined into six species groups and 15 cluster groups. Shown is a summary phylogram adapted from Figure 4.2A with clades collapsed to represent staphylococcal groupings. Whenever possible, cluster and species group names were kept consistent with (182). Cluster groups have been color-coded to represent: **blue**, species that

are novobiocin resistant, coagulase negative, and oxidase positive; **green**, species that are novobiocin susceptible, coagulase negative, and oxidase negative; **orange**, species that are novobiocin resistant, coagulase negative, and oxidase negative; **purple**, species that are novobiocin susceptible, coagulase positive, and oxidase negative; and **red**, species that are novobiocin susceptible, coagulase variable, and oxidase negative. Color scheme exceptions are: #*S. schleiferi schleiferi* is coagulase negative; **S. simiae* is coagulase negative; ‡*S. hominis novobiosepticus* is novobiocin resistant; and †*S. equorum linens* is novobiocin susceptible. Members of each cluster group are listed below the cluster group name. Nodes receiving Pp=1.00 or BS=100% are indicated by grey-filled circles; otherwise, MrBayes support is shown in red text (Pp), BEST support is shown in blue text (Pp), and ML support is shown in black text (BS). Clades that were not present in BEST or ML are indicated by a blue or black §, respectively.

Through the analysis of multiple loci under a variety of phylogenetic methods, we achieved our primary goal of inferring a robust and comprehensive estimate of the phylogeny of *Staphylococcus*. Additionally, we have used this estimate to revise the systematics and nomenclature of phylogenetic groupings for this important genus. The availability of such a robust and comprehensive estimate of the evolutionary origins of, and relationships among, staphylococci provides an important context for understanding patterns of gain and loss of genetic and physiological attributes. This is of particular relevance considering the clinical and economical significance of some *Staphylococcus* species. Additionally, our apparent success in interpreting multiple loci, using multiple types of analysis, to build this robust estimate argues strongly for the utility of doing so in other microbial groups. Approaches such as this will

provide a more natural classification of species based on phylogenetic inferences and lend support to future evolutionarily-informed studies of microbial diversity and physiology.

5. GENERAL DISCUSSION, CONCLUSIONS, AND FUTURE CONSIDERATIONS

5.1 Asymptomatic nasal carriage of clinical *Staphylococcus aureus* isolates

Staphylococcus aureus and other coagulase-negative staphylococci are becoming an increasing concern to public health worldwide. Of particular concern is the ease and frequency with which *S. aureus* causes human disease. Central to this concern is the fact that nasal colonization of *S. aureus* increases the risk of autoinfection. Moreover, nasal carriage of *S. aureus* has been considered a major vector for transmission of virulent strains throughout the community. This hypothesis, however, had not been well studied prior to this dissertation. In Chapter Two, we reported on our investigative findings of the evolutionary relationships among and between nasal carriage strains of *S. aureus,* and clinical isolates. Importantly, we have observed strong supporting evidence that nasal carriage strains and clinical isolates have both evolved from the same genetic background and are genetically (near-) identical within the hypervariable virulence related genes, *clfA, clfB, fnbA,* and *fnbB.* These findings support that these two types of strains are in fact genetically indistinguishable. The implication of this finding is that strains being carried asymptomatically within the nares of healthy individuals are also the strains responsible for high levels of infection and death. Recognizing this fact, efforts to understand better the population dynamics of *S. aureus,* as well as the interplay between the host and bacterial factors involved in nasal colonization have been the focus of a number of studies. Of the virulence genes assessed in this study, clumping factor B (*clfB*) is known to be required for *S. aureus* nasal colonization; however, the requirement for the clumping factor A gene (*clfA*) and fibronectin binding protein genes (*fnbA* and *fnbB*) remains unknown.

While our research has observed nasal carriage and clinical isolates, as genetically identical at clumping factor (*clf*) and fibronectin binding protein (*fnb*) virulence gene loci, it should be noted that additional virulence gene loci, accessory plasmids, and other genetic elements known to contribute to virulence were not assessed in this study. For example, the assessment of such mobile genetic elements as the staphylococcal cassette chromosome *mec* (SCC*mec*) or another feature common to community acquired *S. aureus* strains, the Panton-Valentine leukocidin (PVL) genetic element (172, 198) would be welcomed additions to future studies detailing the genetic similarities between nasal carriage and clinical isolates. The presence of SCC*mec* provides methicillin resistance and is present in approximately 1.3% of nasal carriage strains. Interestingly, when healthcare workers are excluded from studies of MSRA prevalence among community members, the frequency of MRSA drops to 0.2% (121). The presence of MRSA among clinical isolates is approximately 43.2% (12, 108). The PVL element is responsible for encoding two pore-forming β-toxins that are responsible for causing necrotizing lesions. This element is ubiquitous in strains responsible for community-acquired infections while it is only present in approximately 1-5% of strains responsible for Hospital-acquired infections (172). Future assessment of additional factors such as these will be important to gain a fuller understanding of the relationships between clinical *S. aureus* isolates and those being carried asymptomatically within the nares of healthy individuals throughout the community.

Collective research, including that from our laboratory, supports the hypothesis that nasal carriage of *S. aureus* is a major vector for the transmission of virulent strains throughout the community. Thus, prevention of *S. aureus* nasal carriage, particularly in clinical patients and

healthcare workers, would contribute substantially to the reduction or prevention of healthcare associated *S. aureus* infection. For successful nasal decolonization regimens to be implemented in healthcare settings, robust and potent anti-staphylococcal therapies must be used. Currently, only very few options for such practices exist, with resistance a concern.

5.2 Prevention of *Staphylococcus aureus* nasal colonization

Due to the capacity for autoinfection among persistent nasal carriers of *S. aureus* (197), nasal decolonization prior to hospitalization has become a primary consideration within healthcare settings. Nasal decolonization of *S. aureus* would be expected to substantially reduce the length of hospitalization and associated costs as well as the number of subsequent infections and deaths (132).

Toward this end, the most common means of nasal decolonization has been through the use of mupirocin ointment. Widespread use of mupirocin ointment within healthcare settings, however, has led to increased resistance among *S. aureus* isolates. As such, there is a necessity for the development of anti-*S. aureus* antibiotics that will be safe for human use while exhibiting potent antimicrobial activities. In Chapter Three, we characterized the retrocyclin analogue RC-101 as a preventative agent for *S. aureus* nasal colonization. Importantly, RC-101 exhibits robust anti-*S. aureus* activity, but does not impart cytotoxicity or inflammation to human epithelia.

Retrocyclins exhibit broad-spectrum antimicrobial properties (38) which, pending further investigation, may indicate that RC-101 is useful in treating a number of other microbial conditions as well. The safety of RC-101 has been revealed in a number of studies, including a

recent *in vivo* study (29), and may be due to the fact that RC-101 is an analogue of the once-expressed primate peptide, Retrocyclin. Retrocyclin is effective against a number of microbes in addition to *S. aureus,* including *Pseudomonas aeruginosa*. *S. aureus* and *P. aeruginosa* are predominant causes of infection in cystic fibrosis patients and detailed assessments of the extent to which RC-101 is active against *P. aeruginosa* may reveal its therapeutic potential in treating this condition as well. RC-101 may also reveal heightened efficacy toward other species of *Staphylococcus* as well. While *S. aureus* is the primary infectious species among humans, coagulase negative species are exhibiting increased virulence among the human population. An important assessment of RC-101's anti-staphylococcal capacity among other species as well as the activity of this peptide in the presence of multiple species simultaneously will be an important future consideration as this peptide is developed as a therapeutic for nasal decolonization.

5.3 Updated species phylogeny within *Staphylococcus*

The updated species phylogeny of *Staphylococcus* revealed in Chapter Four is the most comprehensive and robust assessment of phylogeny within this genus to date. The necessity for such an assessment within *Staphylococcus* (and other genera) becomes apparent when studies of species relatedness, or evolutionarily-informed studies including phenotypic and biochemical assessments are being conducted. For instance, robust evolutionary assessments of phylogeny can be used to reveal evolutionary events leading to pathogen virulence and antibiotic resistance. Within *Staphylococcus,* novobiocin resistance is frequently assessed as an indicator of the efficacy of using gyrase inhibitors as treatment options. Our phylogenetic inference of

Staphylococcus indicates that novobiocin resistance within this genus was acquired sometime after the split of the Pettenkoferi-Massiliensis cluster group, but prior to divergence of the Arlettae-Kloosii cluster group (Figure 4.5). Thus, when these or related species are discovered in the healthcare setting, treatment options should not include gyrase inhibitors. Future studies addressing the origin of novobiocin resistance among staphylococci may also indicate the pathogen-pathogen interactions leading to the acquisition of the alternative gyrase gene responsible for novobiocin resistance.

Coagulase positive staphylococci comprise only few species of *Staphylococcus*, including *S. aureus*; however, *S. aureus* is responsible for the majority of all human staphylococcal infections. Interestingly, *S. aureus* is a member of the polyphyletic clade containing coagulase positive staphylococci. More basally diverging species that share a clade with *S. aureus* are coagulase negative while other species from more distinct lineages are also coagulase positive. Based on our phylogeny, it appears as though the coagulase gene was acquired two times within *Staphylococcus*; however, the evolutionary processes behind these acquisitions remain unknown. Future assessments of coagulase acquisition among staphylococci will reveal if two independent acquisitions have occurred, or if horizontal transfer between staphylococci contributed to the spread of the coagulase gene. Processes of horizontal gene transfer within *Staphylococcus* or between *Staphylococcus* and other microbial populations remain only weakly understood, and the acquisition of the coagulase gene provides an important avenue for future evolutionary assessment in this regard.

APPENDIX A: CHAPTER TWO SUPPLEMENT

Table A.1. Genotyping details for *S. aureus* isolates analyzed in this study.

Sample[a]	Sequence Type (ST)	Genes[b]				Reference
		clfA	*clfB*	*fnbA*	*fnbB*	
MSSA476	1	1	4	1	3	(77)
MW2	1	1	4	1	3	(7)
D535-3	5	1	2	2	1	This study
D543	5	1	2	2	1	This study
D582	5	1	2	2	1	This study
D618	5	1	2	2	1	This study
D619	5	1	2	2	1	This study
D623	5	1	2	2	1	This study
D635	5	1	2	2	1	This study
N315	5	1	2	2	1	(111)
Mu50	5	1	2	2	1	(111)
Mu3	5	1	2	2	1	(136)
H6556	5	1	2	2	NA	(109)
H7920	5	1	2	2	NA	(109)
D30	8	1	2	4	3	This study
D517	8	1	2	4	3	This study
D521-3	8	1	2	4	3	This study
D554	8	1	2	4	3	This study
D637	8	1	2	4	3	This study
USA300_FR3757	8	1	2	4	3	(43)
NCTC8325	8	1	2	4	3	(59)
Newman	8	1	2	NA	NA	(5)
USA300_TCH1516	8	1	2	4	3	(73)
D540	15	2	4	4	3	This study
D566	15	2	4	4	3	This study
D597	15	2	4	4	3	This study
D627	15	2	4	4	3	This study
H13911	15	2	4	4	NA	(109)
D512	30	5	3	4	1	This study
D512-2	30	1	3	4	1	This study
D512-4	30	1	3	4	1	This study
D512-5	30	1	3	1	1	This study
D521	30	NA	3	1	1	This study
D521-2	30	2	3	1	1	This study
D524	30	NA	3	1	1	This study
D531	30	NA	3	NA	1	This study
D535-2	30	2	3	1	1	This study
D547	30	NA	3	1	1	This study
D563	30	NA	3	4	1	This study
D592	30	3	3	1	1	This study
D599	30	4	3	NA	1	This study
D607	30	NA	3	2	1	This study
D608	30	NA	3	1	1	This study
D651	30	NA	3	1	1	This study
D662	30	NA	3	2	1	This study
D710	30	NA	3	2	1	This study
D719	30	1	NA	4	3	This study

D574	34	1	9	4	1	This study
MRSA252	36	3	3	1	NA	(77)
D558	45	4	5	4	2	This study
D584	45	4	NA	4	2	This study
D589	45	4	5	4	2	This study
D657	45	4	5	4	2	This study
H6606	45	4	5	4	NA	(109)
H13717	45	4	5	4	NA	(109)
D553	50	6	12	NA	NA	This study
D20	59	3	1	1	2	This study
D535	59	NA	1	1	2	This study
D547-4	59	3	1	2	2	This study
D664	72	2	6	1	3	This study
H7639	80	1	2	4	NA	(109)
D714	81	1	4	1	3	This study
D565	87	2	1	1	2	This study
D613	97	1	8	1	3	This study
JH9	105	1	2	2	NA	Copeland 2007 Unpub.
JH1	105	1	2	2	NA	Copeland 2007 Unpub.
H9140	105	1	2	2	NA	(109)
H13199	105	1	2	2	NA	(109)
D628	109	1	4	4	NA	This study
D629	109	1	4	3	NA	This study
D523-5	188	1	2	1	5	This study
D594	188	1	2	1	5	This study
04-02981	225	1	2	2	1	(142)
H9502	228	1	2	2	NA	(109)
TW20	239	1	3	4	3	(78)
H7051	239	1	3	4	NA	(109)
H7951	239	1	3	4	NA	(109)
H7681	239	1	3	4	NA	(109)
COL	250	1	2	4	3	(58)
D579	398	3	11	1	1	This study
D560	508	4	5	4	2	This study
D643	508	4	5	4	1	This study
D507	582	5	4	4	3	This study
D577	672	5	7	3	3	This study
D681-2	1159	2	7	3	4	This study
D605	1181	1	2	4	3	This study
D547-2	1434	NA	6	1	1	This study
D547-3	1507	2	2	1	1	This study
D720	1657	NA	3	1	1	This study
D636	1658	1	3	4	1	This study
D636-2	1658	NA	3	2	1	This study
D20-5	1723	2	6	1	3	This study
D672-2	1724	1	2	3	5	This study
D564	NR	1	10	4	NA	This study
D717	NR	1	4	3	NA	This study
H9779	NR	1	4	1	NA	(109)

[a]Sample names beginning in "D" are nasal carriage strains while all others are clinical strains

[b]Number indicates lineage

NR; not reported, NA; no sequence obtained

Table A.2. GenBank accession numbers for nucleotide sequences utilized/generated in this study.

Isolate	**Sequence available**	**GenBank accession #**
N315	Whole genome	NC_002745
Mu50	Whole genome	NC_002758
COL	Whole genome	NC_002951
MRSA252	Whole genome	NC_002952
MSSA476	Whole genome	NC_002953
MW2	Whole genome	NC_003923
USA300_FPR3757	Whole genome	NC_007793
NCTC8325	Whole genome	NC_007795
JH1	Whole genome	NC_009632
JH9	Whole genome	NC_009487
Newman	Whole genome	NC_009641
Mu3	Whole genome	NC_009782
USA300_TCH1516	Whole genome	NC_010079
04-02981	Whole genome	CP001844
TW20	Whole genome	FN433596
H6556	Partial *clfA*, SD repeats	AM406905
H7920	Partial *clfA*, SD repeats	AM406930
H13911	Partial *clfA*, SD repeats	AM406870
H6606	Partial *clfA*, SD repeats	AM406906
H13717	Partial *clfA*, SD repeats	AM406861
H7639	Partial *clfA*, SD repeats	AM406923
H9140	Partial *clfA*, SD repeats	AM406950
H13199	Partial *clfA*, SD repeats	AM406847
H9502	Partial *clfA*, SD repeats	AM406958
H7051	Partial *clfA*, SD repeats	AM406914
H7951	Partial *clfA*, SD repeats	AM406931
H7681	Partial *clfA*, SD repeats	AM406924
H9779	Partial *clfA*, SD repeats	AM406961
H6556	Partial *clfB*, SD repeats	AM407049
H7920	Partial *clfB*, SD repeats	AM407074
H13911	Partial *clfB*, SD repeats	AM407014
H6606	Partial *clfB*, SD repeats	AM407050
H13717	Partial *clfB*, SD repeats	AM407005
H7639	Partial *clfB*, SD repeats	AM407067
H9140	Partial *clfB*, SD repeats	AM407094
H13199	Partial *clfB*, SD repeats	AM406991
H9502	Partial *clfB*, SD repeats	AM407102
H7051	Partial *clfB*, SD repeats	AM407058
H7951	Partial *clfB*, SD repeats	AM407075
H7681	Partial *clfB*, SD repeats	AM407068
H9779	Partial *clfB*, SD repeats	AM407105
H6556	Partial *fnbA*, D and W domains	AM407190
H7920	Partial *fnbA*, D and W domains	AM407215
H13911	Partial *fnbA*, D and W domains	AM407158
H6606	Partial *fnbA*, D and W domains	AM407191
H13717	Partial *fnbA*, D and W domains	AM407149

H7639	Partial *fnbA*, D and W domains	AM407208
H9140	Partial *fnbA*, D and W domains	AM407235
H13199	Partial *fnbA*, D and W domains	AM407135
H9502	Partial *fnbA*, D and W domains	AM407243
H7051	Partial *fnbA*, D and W domains	AM407199
H7951	Partial *fnbA*, D and W domains	AM407216
H7681	Partial *fnbA*, D and W domains	AM407209
H9779	Partial *fnbA*, D and W domains	AM407246
D20	Partial *clfA*, SD repeats	HQ325854
D20-5	Partial *clfA*, SD repeats	HQ325855
D30	Partial *clfA*, SD repeats	HQ325856
D507	Partial *clfA*, SD repeats	HQ325857
D512	Partial *clfA*, SD repeats	HQ325858
D512-2	Partial *clfA*, SD repeats	HQ325859
D517	Partial *clfA*, SD repeats	HQ325860
D521-2	Partial *clfA*, SD repeats	HQ325861
D521-3	Partial *clfA*, SD repeats	HQ325862
D523-5	Partial *clfA*, SD repeats	HQ325863
D535-2	Partial *clfA*, SD repeats	HQ325864
D535-3	Partial *clfA*, SD repeats	HQ325865
D540	Partial *clfA*, SD repeats	HQ325866
D543	Partial *clfA*, SD repeats	HQ325867
D547-3	Partial *clfA*, SD repeats	HQ325868
D547-4	Partial *clfA*, SD repeats	HQ325869
D553	Partial *clfA*, SD repeats	HQ325870
D554	Partial *clfA*, SD repeats	HQ325871
D558	Partial *clfA*, SD repeats	HQ325872
D560	Partial *clfA*, SD repeats	HQ325873
D564	Partial *clfA*, SD repeats	HQ325874
D565	Partial *clfA*, SD repeats	HQ325875
D566	Partial *clfA*, SD repeats	HQ325876
D574	Partial *clfA*, SD repeats	HQ325877
D577	Partial *clfA*, SD repeats	HQ325878
D579	Partial *clfA*, SD repeats	HQ325879
D582	Partial *clfA*, SD repeats	HQ325880
D584	Partial *clfA*, SD repeats	HQ325881
D589	Partial *clfA*, SD repeats	HQ325882
D592	Partial *clfA*, SD repeats	HQ325883
D594	Partial *clfA*, SD repeats	HQ325884
D597	Partial *clfA*, SD repeats	HQ325885
D599	Partial *clfA*, SD repeats	HQ325886
D605	Partial *clfA*, SD repeats	HQ325887
D613	Partial *clfA*, SD repeats	HQ325888
D618	Partial *clfA*, SD repeats	HQ325889
D619	Partial *clfA*, SD repeats	HQ325890
D623	Partial *clfA*, SD repeats	HQ325891
D627	Partial *clfA*, SD repeats	HQ325892
D628	Partial *clfA*, SD repeats	HQ325893
D629	Partial *clfA*, SD repeats	HQ325894
D635	Partial *clfA*, SD repeats	HQ325895

D636	Partial *clfA*, SD repeats	HQ325896
D637	Partial *clfA*, SD repeats	HQ325897
D643	Partial *clfA*, SD repeats	HQ325898
D657	Partial *clfA*, SD repeats	HQ325899
D664	Partial *clfA*, SD repeats	HQ325900
D672-2	Partial *clfA*, SD repeats	HQ325901
D681-2	Partial *clfA*, SD repeats	HQ325902
D714	Partial *clfA*, SD repeats	HQ325903
D717	Partial *clfA*, SD repeats	HQ325904
D719	Partial *clfA*, SD repeats	HQ325905
D20	Partial *clfB*, SD repeats	HQ325906
D20-5	Partial *clfB*, SD repeats	HQ325907
D30	Partial *clfB*, SD repeats	HQ325908
D507	Partial *clfB*, SD repeats	HQ325909
D512	Partial *clfB*, SD repeats	HQ325910
D517	Partial *clfB*, SD repeats	HQ325911
D521	Partial *clfB*, SD repeats	HQ325912
D521-3	Partial *clfB*, SD repeats	HQ325913
D523-5	Partial *clfB*, SD repeats	HQ325914
D524	Partial *clfB*, SD repeats	HQ325915
D531	Partial *clfB*, SD repeats	HQ325916
D535	Partial *clfB*, SD repeats	HQ325917
D535-2	Partial *clfB*, SD repeats	HQ325918
D535-3	Partial *clfB*, SD repeats	HQ325919
D540	Partial *clfB*, SD repeats	HQ325920
D543	Partial *clfB*, SD repeats	HQ325921
D547	Partial *clfB*, SD repeats	HQ325922
D547-2	Partial *clfB*, SD repeats	HQ325923
D547-3	Partial *clfB*, SD repeats	HQ325924
D547-4	Partial *clfB*, SD repeats	HQ325925
D553	Partial *clfB*, SD repeats	HQ325926
D554	Partial *clfB*, SD repeats	HQ325927
D558	Partial *clfB*, SD repeats	HQ325928
D560	Partial *clfB*, SD repeats	HQ325929
D563	Partial *clfB*, SD repeats	HQ325930
D564	Partial *clfB*, SD repeats	HQ325931
D565	Partial *clfB*, SD repeats	HQ325932
D566	Partial *clfB*, SD repeats	HQ325933
D574	Partial *clfB*, SD repeats	HQ325934
D577	Partial *clfB*, SD repeats	HQ325935
D579	Partial *clfB*, SD repeats	HQ325936
D582	Partial *clfB*, SD repeats	HQ325937
D589	Partial *clfB*, SD repeats	HQ325938
D592	Partial *clfB*, SD repeats	HQ325939
D594	Partial *clfB*, SD repeats	HQ325940
D597	Partial *clfB*, SD repeats	HQ325941
D599	Partial *clfB*, SD repeats	HQ325942
D605	Partial *clfB*, SD repeats	HQ325943
D607	Partial *clfB*, SD repeats	HQ325944
D608	Partial *clfB*, SD repeats	HQ325945

D613	Partial *clfB*, SD repeats	HQ325946
D618	Partial *clfB*, SD repeats	HQ325947
D619	Partial *clfB*, SD repeats	HQ325948
D623	Partial *clfB*, SD repeats	HQ325949
D627	Partial *clfB*, SD repeats	HQ325950
D628	Partial *clfB*, SD repeats	HQ325951
D629	Partial *clfB*, SD repeats	HQ325952
D635	Partial *clfB*, SD repeats	HQ325953
D636	Partial *clfB*, SD repeats	HQ325954
D637	Partial *clfB*, SD repeats	HQ325955
D643	Partial *clfB*, SD repeats	HQ325956
D651	Partial *clfB*, SD repeats	HQ325957
D657	Partial *clfB*, SD repeats	HQ325958
D662	Partial *clfB*, SD repeats	HQ325959
D664	Partial *clfB*, SD repeats	HQ325960
D672-2	Partial *clfB*, SD repeats	HQ325961
D681-2	Partial *clfB*, SD repeats	HQ325962
D710	Partial *clfB*, SD repeats	HQ325963
D714	Partial *clfB*, SD repeats	HQ325964
D717	Partial *clfB*, SD repeats	HQ325965
D720	Partial *clfB*, SD repeats	HQ325966
D20	Partial *fnbA*, D and W domains	HQ325967
D523-5	Partial *fnbA*, D and W domains	HQ325968
D521	Partial *fnbA*, D and W domains	HQ325969
D524	Partial *fnbA*, D and W domains	HQ325970
D535	Partial *fnbA*, D and W domains	HQ325971
D565	Partial *fnbA*, D and W domains	HQ325972
D594	Partial *fnbA*, D and W domains	HQ325973
D608	Partial *fnbA*, D and W domains	HQ325974
D535-2	Partial *fnbA*, D and W domains	HQ325975
D547	Partial *fnbA*, D and W domains	HQ325976
D592	Partial *fnbA*, D and W domains	HQ325977
D512-5	Partial *fnbA*, D and W domains	HQ325978
D720	Partial *fnbA*, D and W domains	HQ325979
D579	Partial *fnbA*, D and W domains	HQ325980
D521-2	Partial *fnbA*, D and W domains	HQ325981
D20-5	Partial *fnbA*, D and W domains	HQ325982
D664	Partial *fnbA*, D and W domains	HQ325983
D547-2	Partial *fnbA*, D and W domains	HQ325984
D613	Partial *fnbA*, D and W domains	HQ325985
D651	Partial *fnbA*, D and W domains	HQ325986
D714	Partial *fnbA*, D and W domains	HQ325987
D710	Partial *fnbA*, D and W domains	HQ325988
D582	Partial *fnbA*, D and W domains	HQ325989
D635	Partial *fnbA*, D and W domains	HQ325990
D543	Partial *fnbA*, D and W domains	HQ325991
D636-2	Partial *fnbA*, D and W domains	HQ325992
D618	Partial *fnbA*, D and W domains	HQ325993
D607	Partial *fnbA*, D and W domains	HQ325994
D547-4	Partial *fnbA*, D and W domains	HQ325995

D535-3	Partial *fnbA*, D and W domains	HQ325996
D623	Partial *fnbA*, D and W domains	HQ325997
D662	Partial *fnbA*, D and W domains	HQ325998
D619	Partial *fnbA*, D and W domains	HQ325999
D629	Partial *fnbA*, D and W domains	HQ326000
D717	Partial *fnbA*, D and W domains	HQ326001
D577	Partial *fnbA*, D and W domains	HQ326002
D672-2	Partial *fnbA*, D and W domains	HQ326003
D681-2	Partial *fnbA*, D and W domains	HQ326004
D560	Partial *fnbA*, D and W domains	HQ326005
D657	Partial *fnbA*, D and W domains	HQ326006
D563	Partial *fnbA*, D and W domains	HQ326007
D643	Partial *fnbA*, D and W domains	HQ326008
D558	Partial *fnbA*, D and W domains	HQ326009
D589	Partial *fnbA*, D and W domains	HQ326010
D507	Partial *fnbA*, D and W domains	HQ326011
D540	Partial *fnbA*, D and W domains	HQ326012
D597	Partial *fnbA*, D and W domains	HQ326013
D627	Partial *fnbA*, D and W domains	HQ326014
D512	Partial *fnbA*, D and W domains	HQ326015
D566	Partial *fnbA*, D and W domains	HQ326016
D719	Partial *fnbA*, D and W domains	HQ326017
D512-2	Partial *fnbA*, D and W domains	HQ326018
D30	Partial *fnbA*, D and W domains	HQ326019
D636	Partial *fnbA*, D and W domains	HQ326020
D521-3	Partial *fnbA*, D and W domains	HQ326021
D517	Partial *fnbA*, D and W domains	HQ326022
D584	Partial *fnbA*, D and W domains	HQ326023
D637	Partial *fnbA*, D and W domains	HQ326024
D574	Partial *fnbA*, D and W domains	HQ326025
D628	Partial *fnbA*, D and W domains	HQ326026
D605	Partial *fnbA*, D and W domains	HQ326027
D554	Partial *fnbA*, D and W domains	HQ326028
D564	Partial *fnbA*, D and W domains	HQ326029
D662	Partial *fnbB*, D and W domains	HQ326030
D592	Partial *fnbB*, D and W domains	HQ326031
D531	Partial *fnbB*, D and W domains	HQ326032
D710	Partial *fnbB*, D and W domains	HQ326033
D651	Partial *fnbB*, D and W domains	HQ326034
D599	Partial *fnbB*, D and W domains	HQ326035
D608	Partial *fnbB*, D and W domains	HQ326036
D574	Partial *fnbB*, D and W domains	HQ326037
D636	Partial *fnbB*, D and W domains	HQ326038
D512	Partial *fnbB*, D and W domains	HQ326039
D521	Partial *fnbB*, D and W domains	HQ326040
D607	Partial *fnbB*, D and W domains	HQ326041
D563	Partial *fnbB*, D and W domains	HQ326042
D524	Partial *fnbB*, D and W domains	HQ326043
D720	Partial *fnbB*, D and W domains	HQ326044
D535-2	Partial *fnbB*, D and W domains	HQ326045

D643	Partial *fnbB*, D and W domains	HQ326046
D579	Partial *fnbB*, D and W domains	HQ326047
D618	Partial *fnbB*, D and W domains	HQ326048
D619	Partial *fnbB*, D and W domains	HQ326049
D547-2	Partial *fnbB*, D and W domains	HQ326050
D623	Partial *fnbB*, D and W domains	HQ326051
D582	Partial *fnbB*, D and W domains	HQ326052
D543	Partial *fnbB*, D and W domains	HQ326053
D535-3	Partial *fnbB*, D and W domains	HQ326054
D635	Partial *fnbB*, D and W domains	HQ326055
D20	Partial *fnbB*, D and W domains	HQ326056
D565	Partial *fnbB*, D and W domains	HQ326057
D535	Partial *fnbB*, D and W domains	HQ326058
D547-4	Partial *fnbB*, D and W domains	HQ326059
D657	Partial *fnbB*, D and W domains	HQ326060
D589	Partial *fnbB*, D and W domains	HQ326061
D584	Partial *fnbB*, D and W domains	HQ326062
D558	Partial *fnbB*, D and W domains	HQ326063
D560	Partial *fnbB*, D and W domains	HQ326064
D681-2	Partial *fnbB*, D and W domains	HQ326065
D597	Partial *fnbB*, D and W domains	HQ326066
D613	Partial *fnbB*, D and W domains	HQ326067
D20-5	Partial *fnbB*, D and W domains	HQ326068
D507	Partial *fnbB*, D and W domains	HQ326069
D540	Partial *fnbB*, D and W domains	HQ326070
D664	Partial *fnbB*, D and W domains	HQ326071
D566	Partial *fnbB*, D and W domains	HQ326072
D577	Partial *fnbB*, D and W domains	HQ326073
D627	Partial *fnbB*, D and W domains	HQ326074
D714	Partial *fnbB*, D and W domains	HQ326075
D554	Partial *fnbB*, D and W domains	HQ326076
D719	Partial *fnbB*, D and W domains	HQ326077
D637	Partial *fnbB*, D and W domains	HQ326078
D605	Partial *fnbB*, D and W domains	HQ326079
D521-3	Partial *fnbB*, D and W domains	HQ326080
D30	Partial *fnbB*, D and W domains	HQ326081
D517	Partial *fnbB*, D and W domains	HQ326082
D594	Partial *fnbB*, D and W domains	HQ326083
D523-5	Partial *fnbB*, D and W domains	HQ326084
D672-2	Partial *fnbB*, D and W domains	HQ326085

Table A.3. Nucleotide sequences for SD repeats at *clfA*.

Repeat numbers and sequences		Repeat numbers and sequences	
1	TCAGATTCTGACCCAGGT	94	TCAGATTCCGACAGCGAT
2	TCAGATAGTGGT	95	TCAGACAGCGAT
3	TCAGATTCTGGCAGCGAT	96	TCTGACTCAGATAGTGAC
4	TCTAATTCAGATAGCGGT	97	TCCGACTTAGACAGCGAC
5	TCAGATTCGGGTAGTGAT	98	TCCGAGTCAGAT
6	TCTACATCAGATAGTGAT	99	TCAGATTCTGGCAGTGAT
7	TCAGATTCAGATAGTGAT	100	TCAGATTCAGACCCAGGT
8	TCAGATTCAGCAAGCGAT	101	TCAGACTCAGTGAGCGAT
9	TCAGATTCAGCGAGCGAT	102	TCCTACTCAGATAGCGAC
10	TCAGATTCAGCAAGTGAT	103	TCAGACTCGGATAGCGAT
11	TCAGATTCAGCGAGTGAT	104	TCAGAATCAGATAATGAC
12	TCCGACTCCGACAGTGAC	105	TCTGACTCAGGTAGTGAC
13	TCCGACTCAGATAACGAT	106	TCGGATTCAGATAGCGAA
14	TCTGACTCAGACAGTGAC	107	TTAGATTCAGACAGCGAC
15	TCAGACTCAGATAGCGAT	108	TCAGATTCAGGTAGCGAT
16	TCAGATTCAGAGAGCGAT	109	TCAGATTCAGAC
17	TCGGATTCAGATAGTGAT	110	TCCGATTCTGAC
18	TCTGACTCAGACAGCGAC	111	TCCGATTCAGATAGCGGT
19	TCAGACTCAGACAGCGAC	112	TCCGATTCAGCAAGTGAT
20	TCAGACTCAGACAGTGAT	113	TCAGACTCAGAAAGTGAC
21	TCAGATTCCGACAGTGAT	114	TCAAATTCCGATAGCGAT
22	TTAGACTCAGACAGTGAC	115	TCAGATTCCGAC
23	TTAGACTCAGACAGCGAC	116	TCAGGTAGTGCC
24	TCAGACTCAGACAGTGAC	117	TCCGACTCAGACAGTGAT
25	TCAGATTCCGACAGTGAC	118	TCAGACTCAGGTAGTGCC
26	TCGGATTCCGATAGCGAT	119	TCTGATTCAGATAGTGAC
27	TCCGACTCAGACAGCGAC	120	TCAACGAGTGACAAAGAA
28	TCCGACTCAGACAGCGAT	121	TCAGACAATGAC
29	TCCGACTCAGATAGCGAC	122	TCAATAGCGATTCCGAGT
30	TCAGACTCAGACAGCGAT	123	TCAGACTCAAACAGCGAT
31	TCAGATTCAGACAGTGAT	124	TCAGATTTAGCAAGCGAT
32	TCAGATTCCGATAGCGAT	125	TCCGATTCAGCGAGTGAC
33	TCAGAATCAGATAGCGAC	126	TCAGATTCCAACAGTGAC
34	TCCGACTCAGTTAGCGAT	127	TCAGACTCAGATAATGAC
35	TCAGATTCAGATAGCAAT	128	TCAGATTCATCAAGTGAT
36	TCAGAATCAGATAGTGAT	129	TCAGATTTGGGTAGTGAT
37	TCAGATTCCGACAGCGAC	130	TCCGATTCAGCGAGCGAT
38	TCCGACTCAGGTAGTGAC	131	TCAGACTCAGCGAGCGAT
39	TCCGACTCAGATAGTGAT	132	TCAGATTTAGACAGCGAC
40	TCAGATTCAACGAGTGAT	133	TCAGACTCACGTAGTGAC
41	TCCGATTCTGAT	134	TCCGAGTCAGTT
42	TCAACGAGTGACACAGGA	135	TCAGATTCAGTGAGTGAT
43	TCAGACAACGAC	136	TCAGACTCAGAC
44	TCTGACTCAGAAAGTGAT	137	TCAGAATCGGATAGCGAC
45	TCAAATAGCGAT	138	TCAGACAGCGAC
46	TCCGACTCAGGT	139	TCAGAATCAGAAAGCGAC
47	TCAGATAGCGGT	140	TCCGATTCAGACAGTGAC
48	TCCGACTCAGCGAGCGAT	141	TCCGACTCAGACAGTGCC

49	TCAGACTCAGATAGTGAC	142	TCGGATTCAACGAGTGAC
50	TCCGATAGCGAT	143	ACAGGATCAGACAACGAC
51	TCAGATTCAGACAGCGAT	144	TCTGAGTCAGGT
52	TCCGACTCAGATAGCGAT	145	TCAGACTCAGGTAGTGGC
53	TCAGATTCAGACAACGAT	146	TCCGATTCAGCAAGCGAT
54	TCTGACTCAGACAGCGAT	147	TCAGACTCAGAAAGCGAC
55	TCCGACTCAGACAGTGAC	148	TCAGACAGTGTT
56	TCGGATTCAGACAGCGAT	149	TCAGACTCGGATAGTGAA
57	TCGGATTCCGACAGTGAT	150	TCCGACTCGGATAGCGAT
58	TCAGATTCCGATAGTGAC	151	TCGGATTCCGACAGCGAT
59	TCGGATTCAGCGAGTGAT	152	TCCGACTCAGATAGTGCC
60	TCCGATTCATCAAGTGAT	153	TCCGATTCAGAT
61	TCCGACTCAGAAAGTGAT	154	TCAGATAACGAC
62	TCCGAGTCAGGT	155	TCAGACTCAGAAAGTGAT
63	TCTACATCAGATAGTGGT	156	TCGAATAGCGAT
64	TCAGACTCAGCGAGTGAT	157	TCCGATTCAGGT
65	TCCGACTCAGACAATGAC	158	TCAGATTCGGGTAGAGGT
66	TCGGATTCAGATAGCGAT	159	TCAGACTCTGGCAGCGAT
67	TCAGATTCAGATAGCGAT	160	TCAGACTCAGAT
68	TCTGACTCCGACAGTGAT	161	TCTGACTCAGAT
69	TCGGATTCAGATAGCGAC	162	TCTGACTCAGACAGTGAT
70	TCAGACTCGGATAGCGAC	163	GCAGACTCAGACAGTGAC
71	TCGGACTCAGATAGCGAT	164	TCAGATTCACGTAGCGAT
72	TCAGAATCAGACAGCGAT	165	TCCGACTCAGATAGTGAC
73	TCAGATTCAGACAGCGAC	166	TCCGACTCAGCAAGTGAT
74	TCAGACAGTGAC	167	TCTAATTCAGATAGCGGC
75	TCAGATTCAGATAGTGAC	168	TCAGACTCAGCAAGCGAT
76	TCAGACTCAGGTAGTGAC	169	TCTGACTCAGAC
77	TCAGATTCAGGCAGCGAT	170	TCAGACTCAGGTAGTGAT
78	TCTACATCAGATAGCGAT	171	TCCGACTCAGGTAGTGAT
79	TCTGACTCAGATAGCGAT	172	TCCGACTCAGGTAGTGCC
80	TCAGATTCAGATAGCGAC	173	TCGGATTCAACCAGTGAC
81	TCAGACTCAGATAGCGAC	174	ACAGGATCAGATAACGAC
82	TCAGATTCGGATAGCGAT	175	TCAGATTCTGACAGTGCC
83	TCAGATTCAGACAGTGAC	176	TCGGAATCAGCGAGTGAT
84	TCAGAATCAGATAGTGAC	177	TCAGATTCTGAT
85	TCCGATTCAGACAGCGAT	178	TCGGAGTCAGGT
86	TCCGATTCAGATAGCGAT	179	TCCGACTCGGATAGCGAC
87	TCAGATTCCGAT	180	TCAGATTCCAATAGCGAT
88	TCAGACAGTGAT	181	TCAGATTCAGCGAGTGGT
89	TCAGATTCGGACCCAGGT	182	TCTACATCAGATAGCGAC
90	TCAGATAGCGAT	183	TCGGATTCCGAC
91	TCAGATTCGGGTAGTGAC	184	TCAGACTCAGATAACGAT
92	TCAGACTCAGCAAGTGAT	185	TCAAATTCTGGCAGTGAT
93	TCGGATTCAGATAGTGAC		

Table A.4. Nucleotide sequences of SD repeats at *clfB*.

Repeat numbers and sequences		Repeat numbers and sequences	
1	TCGGATTCGGACAGTGAC	56	TCAGGTTCAGACAGTGAG
2	TCAGGCTCAGACAGCGAC	57	TCGGACTCAGATAGCAAC
3	TCAGGTTCAGACAGTGAC	58	TCGGATTCGGACAGCGAC
4	TCGGACTCAGACAGCGAC	59	ACAGATTCAGATAGTGAC
5	TCAGATTCAGATAGTGAC	60	ACAGATTCAGACAGCGAC
6	TCAGACTCAGATAGTGAC	61	TCTGATTCAGACAGCGAC
7	TCAGATTCAGACAGCGAT	62	TCCGATTCAGATAGTGAT
8	TCGGATTTAGACAGCGAT	63	TCAGACTCAGGTAGCGAT
9	TCGGATTCAGACAGCGAC	64	TCAGACTCAGATAGTGAG
10	TCAGATTCAGATAGTGAT	65	TCAGATTCCGATAGTGAC
11	TCAGATTCAGACAGCGAC	66	TCCGACTCCGAC
12	TCAGACTCAGATAGTGAT	67	TCCGACAGCGAT
13	TCAGACTCAGACAGTGAG	68	TCCGATTCAGACAGCGAT
14	TCAGATTCAGATAGCGAT	69	TCCGACTCCGACAGCGAT
15	TCAGACTCAGACAGTGAC	70	TCAGATTCAGACAGCGAG
16	TCCGATTCAGATAGCGAT	71	TCCGACACGGACAGCGAC
17	TCGGACTCAGATAGCGAC	72	TCAGATTCAGAAAGTGAC
18	TCCGATTCAGATAGCGAG	73	TCTGATTCAGACAGCGAT
19	TCAGACTCAGACAGTGAT	74	TCAGATTCAGAGAGCGAT
20	TCGGATTCAGACAGCGAT	75	TCCGACTCAGACAGCGAC
21	TCGGATTCAGACAGTGAC	76	TCCGGTTCAGATAGTGAT
22	TCAGAATCAGACAGTGAT	77	TCAGATTCCGACAGCGAT
23	TCAGACTCAGACAGCGAC	78	TCGGATTCCGACAGCGAC
24	TCAGGTTCAGATAGCGAT	79	TCAGATTCCGACAGTGAT
25	TCAGACTCAGATAGCGAT	80	TCCGACTCAGACAGCGAT
26	TCAGAATCAGATAGTGAG	81	TCAGATTCCGACAGCGAC
27	TCAGATTCAGACAGTGAC	82	TCCGATTCAGATAATGAC
28	TCGGACTCAGACAGTGAT	83	TCCGATTCTGATAGTGAC
29	TCAGACTCAGACAGCGAT	84	TCCGACTCTGATAGTGAC
30	TCAGATTCAGATAGCGAC	85	TCTGATTCAGATAGTGAT
31	TCAGAATCAGACAGCGAC	86	TCCGATTCAGACAGTGAC
32	TCAGACTCAGATAGCGAC	87	TCAGACTCAGAAAGCGAT
33	TCAGAATCAGACAGTGAC	88	TCGGACTCAGATAGTGAT
34	TCAGGTTCAGATAGCGAC	89	TCGGATTCAGACAGTGAG
35	TCAGAATCAGATAGCGAT	90	TCCGATTCAGATAGTGAC
36	TCGGATTCAGACAGTGAT	91	TCCGATTCAGACAGTGAG
37	TCAGAATCAGATAGCGAC	92	TCAGGCTCAGACAGCGAT
38	TCGGACTCAGACAGCGAT	93	TCGGATTCAGACAAAGAT
39	TCAGACTCGGATAGCGAT	94	TCAGACTCAGAC
40	TCAGACTCGGATAGCGAC	95	TCAGATAGCGAT
41	TCGGATTCAGATAGCGAC	96	TCAGGCTCAGACAGTGAC
42	TCAGAATCAGACAGTGAG	97	TCAGACTCAGAGAGTGAC
43	TCAGATTCAGATAGTGAG	98	TCAGATTCGGACAGTGAC
44	TCGGACTCAGATAGCGAT	99	TCAGACAGTGAC
45	TCGGATTCAGATAGTGAC	100	TCAGACTTAGACAGTGAC
46	TCAAACTCAGACAGTGAG	101	TCGGACTCAGAGAGTGAC
47	TCGGACTCAGATAGTGAC	102	TCAGATTTAGATAGCGAC
48	TCGGACTCAGACAGTGAG	103	TCAGATTCGGACAGCGAT

49	TCGGATTCAAACAGCGAT	104	TCAGATTCAGATAGCAAC
50	TCGGACTCAGACAGTGAC	105	TTAGATTCAGATAGCGAT
51	TCAAACTCAGATAGTGAC	106	TCGGATTCAGACAACGAT
52	TCGGATTCAGATAGCGAT	107	TCGGAGTCAGAGAGTGAC
53	TCAGAATCAGACAGCGAT	108	TCAGATAGCGAC
54	TCAGACCCAGACAGTGAG	109	TCAGACCCAGATCCGGAT
55	TCAGATTCAGACAGTGAG		

Table A.5. Repeat profiles for *clfA*.

Lineage	Haplotype	Sample	Numeric Profile
1	1	714	1-3-4-5-63-11-11-10-64-8-9-11-11-11-8-65-66-14-32-14-67-67-21-28-68-28-21-52-29-51-51-67-126-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-62
	40	MW2	1-3-4-5-63-11-11-10-64-8-9-11-11-8-65-66-14-32-14-67-67-21-28-68-28-21-52-29-51-51-67-126-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-62
	41	H7051	1-3-4-5-63-181-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
		H7951	1-3-4-5-63-181-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
	42	H9779	1-3-4-5-63-11-10-64-8-9-11-11-8-65-66-14-32-14-67-67-21-28-68-28-21-52-29-51-51-67-126-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-62
	2	574	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-66-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-132-74-75-59-133-60-61-45-134
	3	636	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-66-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-132-74-75-59-76-60-61-45-134
	4	605	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-51-72-73-74-75-59-76-60-61-45-62
	5	30	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
		637	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
		COL	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
		USA300_FPR3757	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
		Newman	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
		USA300_TCH1516	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
	6	719	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-178
	43	NCTC8325	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-66-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-73-74-75-59-76-60-61-45-134
	7	512-2	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-62
		517	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-

			25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-62
		521-3	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-62
	44	TW20	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-21-52-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-73-74-75-59-76-60-61-45-62
	45	MSSA476	1-3-4-5-63-11-11-10-64-8-9-11-11-8-65-66-14-32-21-28-68-28-21-52-29-51-51-7-126-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-75-59-76-60-61-45-62
	46	H7681	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-67-21-28-68-28-29-51-51-67-25-25-66-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
	8	564	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-21-28-68-28-21-52-29-51-51-67-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-45-62
	9	554	1-3-4-5-63-11-10-64-8-48-65-66-14-25-67-14-67-21-28-68-28-21-52-29-51-51-67-25-25-20-59-17-12-69-70-66-71-72-72-73-74-75-59-76-60-61-122
	10	628	1-3-4-5-63-11-10-10-92-11-8-9-11-48-49-32-50-29-51-28-24-32-55-32-24-32-21-48-52-29-51-51-24-32-31-15-51-25-30-55-66-14-59-24-75-59-165-60-61-45-62
	11	629	1-3-4-5-63-11-10-10-92-11-8-9-11-48-49-32-50-29-51-28-24-32-55-32-24-21-48-52-29-51-51-24-32-31-15-51-25-30-55-66-14-59-24-75-59-49-60-61-45-62
	12	717	1-3-4-5-63-11-10-10-92-11-8-9-11-48-49-32-50-29-51-28-24-32-55-32-24-21-48-52-29-51-51-24-32-31-15-51-25-30-55-66-14-59-24-75-59-49-60-61-45-98
	13	613	1-3-4-158-159-4-99-63-11-10-64-11-10-64-10-10-10-48-24-32-117-67-21-29-51-51-160-90-161-90-21-28-162-28-21-29-51-51-51-55-67-31-15-163-19-30-55-66-14-57-55-75-59-49-60-61-45-62
	47	H7920	1-3-4-99-63-11-10-64-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-180-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
	14	582	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-145-60-87-42-43-88-45-62
	48	H9502	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-24-67-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
	15	535-3	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		618	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		619	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-

			76-60-87-42-43-88-45-62
		623	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		635	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		N315	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		H6556	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-55-32-55-32-21-52-29-51-51-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
	49	Mu50	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		Mu3	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-14-32-31-15-25-30-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
	50	H13199	1-3-4-185-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-28-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
	16	543	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-28-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		JH9	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-28-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		JH1	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-28-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		04-02981	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-28-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		H9140	1-3-4-99-63-11-10-64-10-10-10-92-10-8-9-9-11-48-49-32-29-51-54-14-32-28-32-25-58-59-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
	51	H7639	1-3-4-91-63-8-10-11-11-10-92-10-92-9-11-11-51-24-58-81-58-30-7-67-93-80-67-25-94-25-20-66-24-66-55-94-25-20-66-24-66-55-66-79-24-30-32-25-95-24-66-96-97-67-25-82-84-46
	17	523-5	1-3-4-91-63-8-10-11-11-10-92-10-92-9-11-51-24-58-81-58-30-7-67-93-80-67-25-94-25-20-66-24-66-55-94-25-20-66-24-66-55-66-79-24-30-32-25-95-24-66-96-97-67-25-82-84-46
		594	1-3-4-91-63-8-10-11-11-10-92-10-92-9-11-51-24-58-81-58-30-7-67-93-80-67-25-94-25-20-66-24-66-55-94-25-20-66-24-66-55-66-79-24-30-32-25-95-24-66-96-97-67-25-82-84-46
	18	672-2	1-3-4-91-63-8-10-11-11-10-92-10-92-9-11-51-24-58-81-58-30-7-67-93-80-67-25-94-

			25-20-66-24-66-55-66-79-24-30-32-25-95-24-66-96-97-67-25-82-84-46
2	19	547-3	1-47-5-4-11-10-48-49-32-50-29-51-30-14-32-50-52-29-109-110-95-14-32-29-51-54-24-32-55-32-24-32-31-52-52-55-56-55-57-55-58-59-49-60-61-45-62
	20	521-2	89-47-5-4-11-10-48-49-32-50-29-51-30-14-32-50-52-29-53-54-14-32-29-51-54-24-32-55-32-24-32-31-52-52-55-56-55-57-55-58-59-49-60-61-90-62
	21	565	1-47-5-4-11-10-48-49-32-50-29-51-30-14-32-50-52-29-53-54-14-32-29-51-30-24-32-55-32-24-32-31-52-52-55-56-55-57-55-58-59-49-60-61-45-62
	22	535-2	1-47-5-4-11-10-48-49-32-50-29-51-30-14-32-50-52-29-53-54-14-32-29-51-54-24-32-55-32-24-32-31-52-52-55-56-55-57-55-58-59-49-60-61-45-98
	23	20-5	1-47-5-4-11-10-48-49-32-50-29-51-30-14-32-50-52-29-53-54-14-32-29-51-54-24-32-55-32-24-32-31-52-52-55-56-55-57-55-58-59-49-60-61-45-62
	24	540	100-47-5-4-11-10-10-92-9-11-101-9-9-11-101-49-32-102-51-30-24-30-25-103-25-24-80-24-32-21-55-67-21-28-104-30-84-30-11-105-7-55-106-83-7-55-24-107-108-84-61-45-62
	25	566	1-47-5-4-11-10-10-92-92-10-8-9-9-11-131-49-32-102-51-30-24-30-25-103-25-24-80-24-32-21-55-67-21-28-104-30-84-30-11-105-7-55-106-83-7-55-24-107-108-84-61-45-62
		597	1-47-5-4-11-10-10-92-92-10-8-9-9-11-131-49-32-102-51-30-24-30-25-103-25-24-80-24-32-21-55-67-21-28-104-30-84-30-11-105-7-55-106-83-7-55-24-107-108-84-61-45-62
		H13911	1-47-5-4-11-10-10-92-92-10-8-9-9-11-131-49-32-102-51-30-24-30-25-103-25-24-80-24-32-21-55-67-21-28-104-30-84-30-11-105-7-55-106-83-7-55-24-107-108-84-61-45-62
	26	627	1-47-5-4-11-10-10-92-92-10-8-9-9-11-131-49-32-102-51-30-30-25-24-80-24-32-21-55-67-21-28-104-30-84-30-11-105-7-55-106-83-7-55-24-107-164-84-61-45-62
	27	681-2	1-47-5-167-11-10-168-10-11-11-11-11-11-8-10-9-11-20-51-169-95-7-83-94-25-20-71-163-19-170-55-19-170-171-172-83-173-174-74-36-37-128-118-119-59-76-128-81-175-59-76-128-81-175-176-177-42-43-88-45-46
	28	664	1-47-5-4-11-10-8-10-48-10-8-10-48-49-32-50-29-51-30-14-32-50-52-29-53-54-14-32-29-51-54-24-32-55-32-24-32-31-52-52-24-56-55-57-55-58-59-49-60-166-45-62
3	52	MRSA252	1-2-3-4-5-63-9-10-11-10-11-10-11-40-40-146-75-30-75-15-25-66-55-80-147-51-39-73-148-83-149-19-73-149-24-19-81-67-19-15-179-15-150-84-67-29-94-31-22-66-55-151-31-22-66-152-80-30-84-67-141-75-59-153-42-154-155-156-157
	29	592	1-2-3-4-5-63-9-10-11-10-11-10-11-40-40-146-75-30-75-15-25-66-55-80-147-51-39-73-148-83-149-27-73-149-24-19-81-19-15-150-84-67-29-94-31-22-66-55-151-31-22-66-152-80-30-84-67-141-75-59-153-42-154-155-156-157
	30	20	1-2-3-4-5-6-7-8-9-10-6-10-11-8-12-13-14-15-16-12-17-18-19-20-21-22-13-14-15-16-12-17-18-19-20-21-23-24-25-26-27-15-28-29-30-29-11-31-28-32-33-34-14-35-36-37-11-38-32-39-37-40-41-42-43-44-45-46

		547-4	1-2-3-4-5-6-7-8-9-10-6-10-11-8-12-13-14-15-16-12-17-18-19-20-21-22-13-14-15-16-12-17-18-19-20-21-23-24-25-26-27-15-28-29-30-29-11-31-28-32-33-34-14-35-36-37-11-38-32-39-37-40-41-42-43-44-45-46
	31	579	1-2-3-4-5-6-9-9-11-40-40-40-40-11-11-40-40-135-8-75-136-95-75-136-95-67-25-66-55-67-51-137-138-139-25-75-85-140-24-15-141-15-140-20-127-30-59-38-60-141-54-59-38-60-141-119-142-143-61-45-144
4	32	584	1-2-3-4-5-6-11-40-11-8-75-123-75-15-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-130-10-81-32-28-124-11-125-86-28-126-15-19-94-82-51-127-30-59-38-60-55-64-41-42-43-128-117-27-62
		589	1-2-3-4-5-6-11-40-11-8-75-123-75-15-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-130-10-81-32-28-124-11-125-86-28-126-15-19-94-82-51-127-30-59-38-60-55-64-41-42-43-128-117-27-62
		599	1-2-3-4-5-6-11-40-11-8-75-123-75-15-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-130-10-81-32-28-124-11-125-86-28-126-15-19-94-82-51-127-30-59-38-60-55-64-41-42-43-128-117-27-62
	33	643	1-2-3-4-5-6-11-40-11-8-75-123-75-15-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-130-10-81-32-28-124-11-125-86-28-126-15-19-94-82-51-104-30-59-38-60-55-64-41-42-43-128-117-27-62
	34	558	1-2-3-4-5-6-11-40-11-8-75-123-75-15-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-9-10-81-32-28-124-11-125-86-28-126-15-19-94-82-51-127-30-59-38-60-55-64-41-42-43-128-117-27-62
	53	H6606	1-2-3-4-5-6-11-75-123-75-15-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-130-10-81-32-30-124-11-125-86-28-126-15-27-94-82-51-127-30-59-38-60-55-64-41-42-43-128-117-27-62
	35	560	1-2-3-4-129-6-11-40-11-8-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-130-10-81-32-28-124-11-125-86-28-19-94-82-51-104-30-59-38-60-55-64-41-42-43-128-117-27-62
	36	657	1-2-3-4-5-6-11-40-11-8-75-123-75-15-83-15-25-66-55-67-73-24-83-67-73-24-31-19-15-130-10-81-32-28-124-11-125-86-28-126-15-19-94-82-51-127-30-59-38-60-117-27-62
	54	H13717	1-2-3-4-5-182-32-28-124-11-125-86-29-28-126-15-19-94-82-51-127-30-59-38-60-55-64-41-42-43-128-117-183-74-184-112-27-62
5	37	507	1-77-4-5-78-8-8-8-11-11-11-67-79-17-80-73-73-51-24-15-25-24-67-24-67-15-19-80-30-81-67-24-66-24-82-83-58-11-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
		512	1-77-4-5-78-8-8-8-11-11-11-67-79-17-80-73-73-51-24-15-25-24-67-24-67-15-19-80-30-81-67-24-66-24-82-83-58-11-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-45-62
	38	577	1-77-4-5-78-8-8-8-11-11-11-67-15-17-80-73-73-51-24-15-25-24-67-24-67-15-19-80-30-81-67-24-66-24-82-83-58-11-67-84-20-75-30-84-85-84-86-59-76-60-87-42-43-88-

			45-62
6	39	553	100-47-5-8-111-5-10-8-112-5-8-112-5-32-55-113-114-27-7-31-28-32-52-55-19-95-115-116-21-52-10-117-66-36-37-11-118-119-59-76-10-117-56-59-41-120-121-46

Table A.6. Repeat profiles for *clfB*.

Lineage	Haplotype	Sample	Numeric Profile
1	1	20	1-2-3-4-5-5-6-6-7-8-4-5-6-7-8-7-9-10-6-11-11-10-12-7-13-14-15-12-13-16-17-18-7-15-10-19-20-14-17
		535	1-2-3-4-5-5-6-6-7-8-4-5-6-7-8-7-9-10-6-11-11-10-12-7-13-14-15-12-13-16-17-18-7-15-10-19-20-14-17
		547-4	1-2-3-4-5-5-6-6-7-8-4-5-6-7-8-7-9-10-6-11-11-10-12-7-13-14-15-12-13-16-17-18-7-15-10-19-20-14-17
	2	565	1-2-3-4-5-5-6-6-6-7-8-4-10-12-7-13-14-15-12-13-16-17-18-7-15-10-19-20-14-17
2	3	672-2	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-29-11-11-32-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
	4	30	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-29-11-11-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
		USA300_FPR3757	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-29-11-11-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
		USA300_TCH1516	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-29-11-11-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
	5	554	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-29-11-11-32-23-35-23-23-32-14-23-14-38-11-39-14-36-11-40-29-40-41-6
	45	COL	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-11-11-32-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
		Newman	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-11-11-32-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
	6	637	21-2-34-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-29-11-11-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
	7	517	21-2-34-35-36-11-35-37-25-14-14-36-11-37-26-27-28-25-25-29-11-11-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
		521-3	21-2-34-35-36-11-35-37-25-14-14-36-11-37-26-27-28-25-25-29-11-11-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
	8	523-5	21-2-34-37-14-14-14-36-11-37-26-27-28-25-25-29-11-11-32-23-14-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
		594	21-2-34-37-14-14-14-36-11-37-26-27-28-25-25-29-11-11-32-23-14-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
	46	H6556	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-7-11-31-32-23-14-23-31-25-23-14-38-11-39-11-40-45
	9	605	21-2-35-36-11-35-37-14-14-14-36-11-37-26-27-28-25-25-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-40-41-6

	10	535-3	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-23-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		618	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-23-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		619	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-23-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		635	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-23-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		H7920	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-23-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		H9502	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-23-32-23-14-23-23-25-23-14-38-11-39-11-40-45
	11	543	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		N315	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		Mu50	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		Mu3	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		04-02981	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		H9140	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		H13199	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
	47	JH9	21-2-34-35-36-11-37-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
		JH1	21-2-34-35-36-11-37-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-23-25-23-14-38-11-39-11-40-45
	12	547-3	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-37-23-14-23-23-25-23-14-38-11-39-11-40-45
		582	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-37-23-14-23-23-25-23-14-38-11-39-11-40-45
	48	NCTC8325	21-2-34-35-36-11-37-26-27-28-25-25-29-11-11-32-23-14-23-23-32-14-23-14-38-11-39-14-36-11-40-29-32-40-41-6
	49	H7639	21-2-34-35-36-11-35-21-14-37-43-27-44-106-23-30-31-32-23-14-29-29-29-32-31-32-14-29-23-25-19-7-33-14-17
	13	623	21-2-34-35-36-11-35-37-14-14-35-36-11-37-26-27-28-25-25-29-11-31-32-23-14-23-

			23-25-23-14-38-11-40-45
3	50	TW20	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-46-27-17-44-21-56-7-45-45-47-7-52-7-53-21-41-20-50-16-42-14-33-33-7-29-13-55-7-17
	14	535-2	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-46-27-17-44-21-56-7-41-45-47-7-52-7-53-21-41-20-50-16-42-14-33-13-7-29-13-55-7-17
		547	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-46-27-17-44-21-56-7-41-45-47-7-52-7-53-21-41-20-50-16-42-14-33-13-7-29-13-55-7-17
	51	H7681	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-46-27-17-44-21-56-7-45-45-47-7-52-7-53-21-41-20-50-16-42-14-33-53-29-13-55-7-17
	15	720	45-24-45-6-46-27-47-14-7-48-30-49-50-46-27-17-44-21-56-7-41-45-47-7-52-7-53-21-41-20-50-16-42-14-33-15-7-29-13-55-7-17
	52	H7051	45-2-24-45-6-46-27-47-47-47-46-27-17-44-21-56-7-45-45-47-7-52-7-53-21-41-20-50-16-42-14-33-33-7-29-13-55-7-17
		H7951	45-2-24-45-6-46-27-47-47-47-46-27-17-44-21-56-7-45-45-47-7-52-7-53-21-41-20-50-16-42-14-33-33-7-29-13-55-7-17
	16	524	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-42-14-33-54-7-29-13-55-7-17
	17	592	45-2-24-45-6-46-27-47-5-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
	18	512	45-2-24-45-6-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
	19	521	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
		563	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
		636	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
		662	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
	20	599	45-2-24-45-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
	21	651	45-2-24-45-6-46-27-47-14-53-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-42-7-17
	53	MRSA252	45-2-24-45-6-46-27-47-14-7-48-14-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-13-55-7-17
	22	531	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-42-7-17
		607	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-42-7-17

		608	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-52-7-53-21-41-20-50-16-42-14-33-54-7-29-42-7-17
	23	710	45-2-24-45-6-46-27-47-14-7-48-30-49-50-47-47-51-7-53-21-41-20-50-16-42-14-33-54-7-29-42-7-17
4	24	540	9-24-20-23-24-9-42-14-37-43-27-44-29-23-29-37-43-5-44-29-23-32-43-27-32-14-29-7-23-23-35-29-30-29-30-23-35-29-11-40-45
	25	627	9-24-20-23-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-32-14-29-7-23-23-35-29-30-29-30-23-11-40-45
	26	507	9-24-20-23-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-32-14-29-7-29-30-29-30-23-35-29-11-40-45
	27	597	9-24-20-23-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-32-14-29-7-23-23-35-29-11-40-45
	28	628	9-24-20-23-24-9-42-14-29-23-32-43-27-32-14-29-7-23-23-35-29-11-40-45
		629	9-24-20-23-24-9-42-14-29-23-32-43-27-32-14-29-7-23-23-35-29-11-40-45
		717	9-24-20-23-24-9-42-14-29-23-32-43-27-32-14-29-7-23-23-35-29-11-40-45
	29	566	9-24-20-23-24-9-42-14-29-7-11-35-29-30-29-30-23-25-29-11-40-45
	54	H13911	9-24-20-23-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-32-14-29-7-23-23-35-29-30-29-30-23-35-29-11-40-45
	30	714	9-24-20-23-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-29-29-7-23-23-35-29-30-29-30-23-35-29-11-40-107
	55	MSSA476	9-24-20-23-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-25-29-7-23-23-35-29-30-29-108-23-35-29-11-40-45
	56	MW2	9-24-20-23-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-25-29-7-23-23-35-29-30-29-30-23-35-29-11-40-45
	57	H9779	9-24-9-42-14-37-43-27-44-29-23-30-29-37-43-5-44-29-23-32-43-27-25-29-7-23-23-35-29-30-29-30-23-35-29-11-40-45
5	31	558	7-63-64-7-63-64-7-21-16-17-65-66-67-14-20-20-15-14-48-7-9-68-13-14-48-7-7-20-15-16-20-15-14-9-68-13-14-48-7-20-15-14-52-48-7-20-13-14-41-15-38-17
	32	560	7-63-64-7-63-64-7-21-16-17-65-69-14-20-20-15-14-48-7-9-68-13-14-48-70-7-20-15-14-9-68-17
	33	657	7-63-64-7-63-64-7-21-16-17-65-69-14-20-20-15-14-48-7-9-68-13-14-48-70-7-20-15-16-20-15-14-9-68-13-105-48-7-20-15-14-52-48-7-20-13-14-41-15-38-17
	58	H6606	7-63-64-7-63-64-7-21-16-17-65-69-14-20-20-15-14-48-7-9-68-13-14-48-7-7-20-15-16-20-15-14-9-68-13-14-48-7-20-15-14-52-48-7-20-13-14-41-15-38-17
	34	589	7-63-64-7-63-64-7-21-68-15-14-48-7-9-68-13-14-48-70-7-20-15-16-20-15-16-20-15-14-9-68-13-14-48-7-20-15-14-52-48-7-20-13-14-41-15-38-17
	59	H13717	7-63-64-7-21-14-44-20-20-15-14-48-7-9-68-13-14-48-70-7-20-15-16-52-48-7-20-15-16-41-13-7-20-15-16-52-48-7-20-15-16-41-109-7-32
	35	643	7-63-64-7-63-64-7-21-16-17-65-69-14-20-15-14-48-7-9-68-13-14-48-70-7-20-15-14-

			20-15-14-9-68-17
6	36	664	21-22-20-23-24-9-25-26-27-28-25-25-29-11-11-32-23-14-23-30-31-32-23-14-29-29-32-23-32-23-25-23-14-29-23-25-19-7-33-14-17
	37	20-5	21-22-20-23-24-9-25-26-27-28-25-25-29-11-30-14-23-30-31-32-23-7-29-29-29-29-32-23-32-23-25-23-14-29-23-25-19-7-33-14-17
		547-2	21-22-20-23-24-9-25-26-27-28-25-25-29-11-30-14-23-30-31-32-23-7-29-29-29-29-32-23-32-23-25-23-14-29-23-25-19-7-33-14-17
7	38	577	21-2-34-35-36-11-35-37-14-14-14-36-14-37-43-11-24-20-21-92-21-14-37-43-27-44-93-23-30-31-32-23-14-29-29-29-94-95-29-32-31-32-23-25-23-14-29-23-25-19-7-33-14-17
	39	681-2	43-11-24-20-21-92-36-33-14-37-43-27-44-106-23-31-25-23-14-29-32-29-29-32-23-32-23-32-23-25-23-14-29-23-25-19-7-33-14-17
8	40	613	27-2-24-52-103-32-104-21-22-53-15-14-37-43-23-41-35-36-37-13-14-37-43-14-29-7-29-23-35-35-11-13-14-37-43-27-17-23-25-31-14-29-25-19-31-25-30-52-47-32-23-25-17
9	41	574	1-2-24-41-52-9-88-31-52-21-14-42-11-14-17-88-31-52-17-88-31-52-21-45-46-27-14-42-14-33-13-7-41-13-7-41-16-89-29-17-90-91-30
10	42	564	71-72-24-41-68-44-47-36-73-17-74-75-76-77-17-52-75-14-78-30-52-75-79-25-80-75-81-82-83-32-81-84-27-32-85-11-86-11-87
11	43	579	1-96-20-50-52-33-97-33-47-5-6-46-98-14-99-29-17-15-3-21-6-100-101-20-52-33-102-9-6-14-20-33-6-7-22-14-9-38-17
12	44	553	1-2-3-57-4-7-58-30-59-30-60-11-13-61-44-41-62-12-7-13-14-13-11-15-13-16-17-18-7-13-62-19-20-14-17

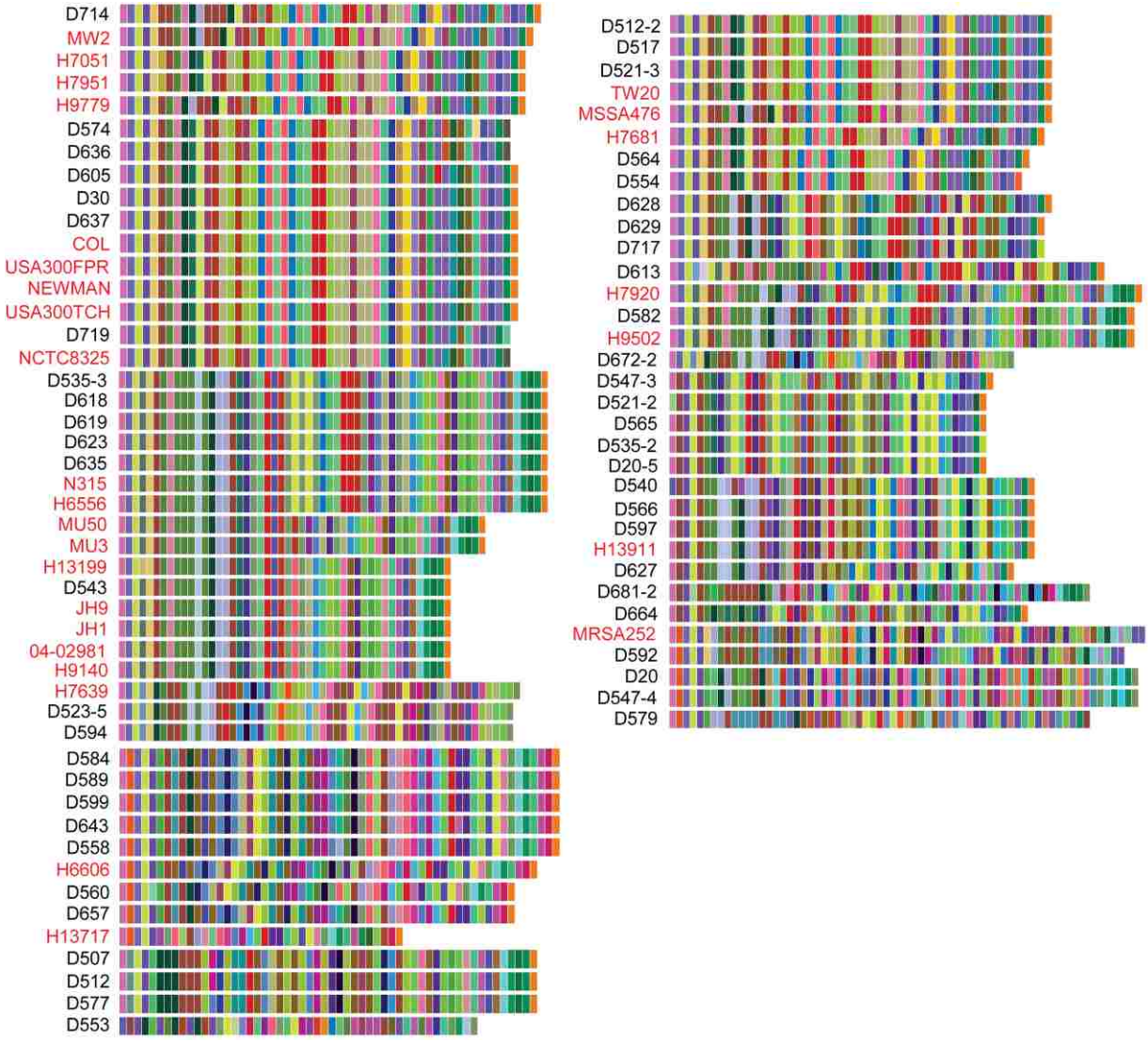

Figure A.1. Color-coded repeats of *clfA* R domains.

Figure A.2. Color-coded repeats of *clfB* R domains.

A.1 Repeat profiling program for *clfA*

The *clf* repeat profiling program, gensh, was written using the C++ programming language and designed to convert DNA sequences in the serine-aspartic acid repeat domain (R domain) into numeric profiles. When a series of DNA sequences are entered into the program, the R domain start site is identified and used as the beginning of the first repeat unit. Beginning with the first repeat, consecutive 18 bp segments of DNA are then analyzed. All unique repeat units are given a different number in the sequence in which they are encountered (all identical repeats are given the same number). When identifying repeat units, 18 bp segments are analyzed; however, 12 bp repeats are identified by the presence of "TCN" at bases 13-15. When "TCN" is present at bases 13-15, the program treats this a the start of a new repeat unit and identifies the previous repeat as only 12 bp. This process continues until the stop sequence, which identifies the end of the R domain, is encountered. Note that because the program arbitrarily numbers repeat units in the order in which they are encountered, no inference can be made as to the degree of nucleotide similarity between different numbers (e.g. repeat number 1 is not necessarily more similar to repeat number 2 than any other repeat number). After the program is complete, two files will be generated. One contains a log of the numbered repeat sequences that were identified throughout the complete dataset while the other is a file containing the sample names and their numeric repeat profiles. Note that the source code below is only for *clfA* DNA sequences. DNA sequences from *clfA* and *clfB* cannot be combined and analyzed in the same data set. For DNA sequences of *clfB* R domains, refer to Supplementary Text A2.

To use the program, gensh:

compile using g++ example: g++ gensh.cpp -o gensh

This creates a gensh file that can be run by: ./gensh INFILE OUTFILE1 OUTFILE2 where the INFILE is a fasta formatted DNA file containing *clf* R domain DNA sequences and OUTFILEs are the files that the program produces. OUTFILE1 is a tab delimited file containing all repeat units identified in the dataset along with their repeat numbers. OUTFILE2 is also a tab delimited file and contains the sample names and their numeric repeat.

```
#include <iostream> #include <fstream> #include <list> #include <vector>
using namespace std;
//Global Variables fstream INFILE; fstream OUTFILE1; fstream OUTFILE2;
typedef struct{ string Segment; int Number;
}TChunk; vector<string> SeqName;
//Function Declarations void ReadInput(list<string> &SeqList); void CreateOutput(list<string> SeqList); int
FindChunk(vector<TChunk> list, string chunk);
int main(int argc, char* argv[]) {
/*cout << "argc = " << argc << endl; for(int i = 0; i < argc; i++)
cout << "argv[" << i << "] = " << argv[i] <<
profiles.
endl;*/
if(argc < 4 || argc > 4){ cout << "Program input as follows:" << endl; cout << "\t./gensh INFILE OUTFILE1
OUTFILE2\n" << endl; return 0;
} INFILE.open(argv[1], ios::in);
}
}
OUTFILE1.open(argv[2], OUTFILE2.open(argv[3],
string FirstSeq; list<string> SeqList;
ReadInput(SeqList); CreateOutput(SeqList);
return 0;
ios::out); ios::out);
void ReadInput(list<string> &SeqList){ //Hold the sequence name just in case we need it later... string Line; string
CompleteSeq; bool SequenceStored = false;
if (INFILE.is_open()) { /* ok, proceed with output */ cout << "WOHO, lets collect some strings!" << endl;
}
while( !INFILE.eof() ){
//Read in the Sequence Name getline(INFILE, Line);
switch (Line[0]){ case '>':
}
//Handle the Line as a name SeqName.push_back(Line.substr(1, Line.size()-1)); if(SequenceStored){
//cout << "SEQ = " << CompleteSeq << endl; SeqList.push_back(CompleteSeq); SequenceStored = false;
} CompleteSeq.clear(); break;
default: //Keep adding up the String... SequenceStored = true; if(Line[Line.size()-1] == '\n')
CompleteSeq += Line.substr(0, Line.size()-1); else
CompleteSeq += Line.substr(0, Line.size()); break;
} SeqList.push_back(CompleteSeq); INFILE.close();
void CreateOutput(list<string> SeqList) {
bool terminated = false; vector<TChunk> UniqueSeq; vector<TChunk> Outfile; //Temporary Chunk that will be
pushed onto the list it is not a duplicate TChunk tmp;
string chunk; int count = 0; for (list<string>::iterator it = SeqList.begin(); it != SeqList.end(); it++) {
string seq = *it; //cout << seq << endl;
for(int i=0; i<seq.size()-2; i++) {
```

```
if(seq.substr(i,3) == "GAT")
{
{
OUTFILE2
cout << "Found a MATCH!" << endl;
//From the newly found chunk to the end of the complete seq. seq = seq.substr(i+3, seq.size());
//Find the terminating seq, if not found skip out of everything. terminated = false; for(int j=0; j<seq.size(); j++) {
if(seq.substr(j,2) == "TC" && seq.substr(j+3, 9) == "AACAATAAT") {
terminated = true; seq = seq.substr(0,j);
}
} if(terminated) { << SeqName.at(count) << '\t';
//cout << "TERMINATED" << endl; //cout << seq << endl; while(seq.size() >= 12) {
//location of the chunk, -1 if not found in the list. int chunk_location;
if(seq.substr(i+3, 2) == "TC" && (seq.substr(i+6, 3) == "GAC" || seq.substr(i+6, 3) == "GAT"))
}
}
//Grab the 18 character chunk if(seq.size() < 18) {
chunk = seq; seq.clear();
} else {
and check to see if it matches specs
chunk = seq.substr(0, 18); //cout << "CHUNK = " << chunk //If the chunk has "TC" in pos 13-14, copy only the first 12
and //restart chunks from TC... if(chunk.substr(12, 2) == "TC") {
//cout << "TCN found :" << endl; chunk = chunk.substr(0,12); seq = seq.substr(12, seq.size());
} else {
//cout << "TCN Not found : " << endl; seq = seq.substr(18, seq.size());
} } chunk_location = FindChunk(UniqueSeq, chunk);
if(chunk_location >= 0) {
OUTFILE2 << (chunk_location+1); //UniqueSeq.at(chunk_location).Number += 1;
} else {
TChunk tmp = {chunk, 1}; UniqueSeq.push_back(tmp); OUTFILE2 << UniqueSeq.size();
} if(seq.size() >= 12)
OUTFILE2 << '-';
<< endl;
/*
}
} if(terminated)
OUTFILE2 << '\n'; //Output everything to a file yo...
OUTFILE1 << SeqName.at(count) << '\n'; OUTFILE2 << SeqName.at(count) << '\t'; for(int i=0; i<UniqueSeq.size();
i++) {
OUTFILE1 << (i+1) << ' ' << UniqueSeq.at(i).Segment << '\n'; for(int j=0; j<UniqueSeq.at(i).Number; j++)
OUTFILE2 << (i+1) << '-';
} OUTFILE1 << '\n'; OUTFILE2 << '\n';*/ count++;
}
}
else {
OUTFILE2 << SeqName.at(count) << '\t' << "ERROR" << '\n'; break;
}
for(int i=0; i<UniqueSeq.size(); i++) {
OUTFILE1 << (i+1) << ' ' << UniqueSeq.at(i).Segment << '\n';
}
int FindChunk(vector<TChunk> list, string chunk) {
}
}
for(int i=0; i<list.size(); i++) {
string tmp = list.at(i).Segment; if(tmp == chunk)
return i; return -1;
}
```

A.2 Repeat profiling program for *clfB*

This *clf* repeat profiling program is intended to be used with *clfB* R domain DNA sequences only. It is similar to that for *clfA* repeat profiling; however, the stop site has been changed for use with locus *clfB*. Refer to Text A1 for a detailed description of what the program does and how it works.

To use the program, gensh:

compile using g++ example: g++ gensh.cpp -o gensh

This creates a gensh file that can be run by: ./gensh INFILE OUTFILE1 OUTFILE2 where the INFILE is a fasta formatted DNA file containing clf R domain DNA sequences and OUTFILEs are the files that the program produces. OUTFILE1 is a tab delimited file containing all repeat units identified in the dataset along with their repeat numbers. OUTFILE2 is also a tab delimited file and contains the sample names and their numeric repeat profiles.

```
#include <iostream> #include <fstream> #include <list> #include <vector>
using namespace std;
//Global Variables fstream INFILE; fstream OUTFILE1; fstream OUTFILE2;
typedef struct{ string Segment; int Number;
}TChunk; vector<string> SeqName;
//Function Declarations void ReadInput(list<string> &SeqList); void CreateOutput(list<string> SeqList); int
FindChunk(vector<TChunk> list, string chunk);
int main(int argc, char* argv[]) {
}
/*cout << "argc = " << argc << endl; for(int i = 0; i < argc; i++)
cout << "argv[" << i << "] = " << argv[i] << endl;*/
if(argc < 4 || argc > 4){ cout << "Program input as follows:" << endl; cout << "\t./gensh INFILE OUTFILE1
OUTFILE2\n" << endl; return 0;
} INFILE.open(argv[1], ios::in);
OUTFILE1.open(argv[2], OUTFILE2.open(argv[3],
string FirstSeq; list<string> SeqList;
ReadInput(SeqList); CreateOutput(SeqList);
return 0;
ios::out); ios::out);
void ReadInput(list<string> &SeqList){ //Hold the sequence name just in case we need it later...
}
string Line; string CompleteSeq; bool SequenceStored = false;
if (INFILE.is_open()) { /* ok, proceed with output */ cout << "WOHO, lets collect some strings!" << endl;
```

```
}
while( !INFILE.eof() ){
//Read in the Sequence Name getline(INFILE, Line);
switch (Line[0]){ case '>':
}
//Handle the Line as a name SeqName.push_back(Line.substr(1, Line.size()-1)); if(SequenceStored){
//cout << "SEQ = " << CompleteSeq << endl; SeqList.push_back(CompleteSeq); SequenceStored = false;
} CompleteSeq.clear(); break;
default: //Keep adding up the String... SequenceStored = true; if(Line[Line.size()-1] == '\n')
CompleteSeq += Line.substr(0, Line.size()-1); else
CompleteSeq += Line.substr(0, Line.size()); break;
} SeqList.push_back(CompleteSeq); INFILE.close();
void CreateOutput(list<string> SeqList) {
bool terminated = false; vector<TChunk> UniqueSeq; vector<TChunk> Outfile; //Temporary Chunk that will be
pushed onto the list it is not a duplicate TChunk tmp;
string chunk; int count = 0; for (list<string>::iterator it = SeqList.begin(); it != SeqList.end(); it++) {
string seq = *it; //cout << seq << endl;
for(int i=0; i<seq.size()-2; i++) {
if(seq.substr(i,3) == "GAT") {
if(seq.substr(i+3, 2) == "TC" && (seq.substr(i+6, 3) == "GAC" || seq.substr(i+6, 3) == "GAT"))
{
cout << "Found a MATCH!" << endl;
//From the newly found chunk to the end of the complete seq. seq = seq.substr(i+3, seq.size());
//Find the terminating seq, if not found skip out of everything. terminated = false; for(int j=0; j<seq.size(); j++) {
if(seq.substr(j,2) == "TC" && seq.substr(j+3, 9) == "GATTCAAGA")
/*
}
} else {
{
}
terminated = true; seq = seq.substr(0,j);
} if(terminated) { << SeqName.at(count) << '\t';
//cout << "TERMINATED" << endl; //cout << seq << endl; while(seq.size() >= 12) {
OUTFILE2
}
}
} if(terminated)
OUTFILE2 << '\n'; //Output everything to a file yo...
OUTFILE1 << SeqName.at(count) << '\n'; OUTFILE2 << SeqName.at(count) << '\t'; for(int i=0; i<UniqueSeq.size();
i++)
//location of the chunk, -1 if not found in the list. int chunk_location;
//Grab the 18 character chunk if(seq.size() < 18) {
chunk = seq; seq.clear();
} else {
and check to see if it matches specs
chunk = seq.substr(0, 18); //cout << "CHUNK = " << chunk //If the chunk has "TC" in pos 13-14, copy only the first 12
and //restart chunks from TC... if(chunk.substr(12, 2) == "TC") {
//cout << "TCN found :" << endl; chunk = chunk.substr(0,12); seq = seq.substr(12, seq.size());
} else {
//cout << "TCN Not found : " << endl; seq = seq.substr(18, seq.size());
} } chunk_location = FindChunk(UniqueSeq, chunk);
if(chunk_location >= 0) {
OUTFILE2 << (chunk_location+1); //UniqueSeq.at(chunk_location).Number += 1;
} else {
TChunk tmp = {chunk, 1}; UniqueSeq.push_back(tmp); OUTFILE2 << UniqueSeq.size();
} if(seq.size() >= 12)
OUTFILE2 << '-';
OUTFILE2 << SeqName.at(count) << '\t' << "ERROR" << '\n'; break;
}
```

```
<< endl;
}
}
}
{
OUTFILE1 << (i+1) << ' ' << UniqueSeq.at(i).Segment << '\n'; for(int j=0; j<UniqueSeq.at(i).Number; j++)
OUTFILE2 << (i+1) << '-';
} OUTFILE1 << '\n'; OUTFILE2 << '\n';*/ count++;
for(int i=0; i<UniqueSeq.size(); i++) {
OUTFILE1 << (i+1) << ' ' << UniqueSeq.at(i).Segment << '\n';
}
int FindChunk(vector<TChunk> list, string chunk) {
for(int i=0; i<list.size(); i++) {
string tmp = list.at(i).Segment; if(tmp == chunk)
return i; return -1;
}
```

A.3 *clf* color-coded repeat generator

The program, wgraph, was written using the C++ programming language and designed to convert numeric clf R domain profiles (such as those created by the accompanying program, gensh) into color-coded representations. When a file containing numeric repeat profiles is entered into the program, a color coded visual representation is generated. Identical numbers are colored the same throughout the data set so repeats containing identical DNA sequence are colored the same. When initiated, the program automatically reads an input file containing hexadecimal colors that it then uses to generate the output. The output can be saved by using the print screen command. The size of the individual boxes on the graph can be adjusted as per the user's preference by changing the S_WIDTH and S_HEIGHT parameters.

To use the program, wgraph:

```
compile using g++ example: g++ wgraph.cpp -o wgraph
This creates a wgraph ./wgraph INFILE.dat
where the INFILE is a
#include <GL/glut.h> #include <iostream> #include <fstream> #include <string> #include <stdlib.h> #include
<string.h> #include <stdio.h> #include <vector>
using namespace std;
#define WINDOW_X 1000 #define WINDOW_Y 1000 #define S_WIDTH 12 #define S_HEIGHT 44
file that can be run by: .dat file containing numeric profiles for clf R domain DNA sequences.
void displayCB(void); void SetPenColorHex(unsigned long color); void DrawFillBox(double x0, double y0, double x1,
double y1); int xtoi(const char* xs, unsigned long* result);
int ypos = WINDOW_Y - 10; int xpos = 10;
void displayCB(void) /* function called whenever redisplay needed */ {
glClear(GL_COLOR_BUFFER_BIT);
//Open and read the hex datafile... //unsigned long hex; string line; vector<string> HexList;
ifstream myfile ("hex.dat"); if (myfile.is_open()) {
while (! myfile.eof() ) {
getline (myfile,line); HexList.push_back(line); /*xtoi(line.c_str(), &hex); //cout << hex << endl; SetPenColorHex(hex);
/* clear the display */
//cout << line << endl; DrawFillBox(xpos, ypos-S_HEIGHT, xpos+S_WIDTH, ypos); xpos += (S_WIDTH+1); if(xpos >
(WINDOW_X - 20)) {
xpos = 10; cout << "BEFORE: " << S_HEIGHT << endl; ypos = ypos - S_HEIGHT - 10;
cout << "Y POS: " << ypos << endl; }*/
} myfile.close();
} else cout << "Unable to open hex file" << endl; //Read the Datfile and print this... myfile.open("infile.dat"); if
(myfile.is_open()) {
}
```

```
int pos; string Name; while (! myfile.eof() ) {
getline(myfile,line); if(line.length() > 0) {
pos = line.find("\t"); Name = line.substr(0,pos); cout << "Name: " << Name << " Size=" << line.length() << endl; line =
line.substr(pos+1, line.length()); cout << "Line: " << line << endl;
int found=line.find_first_of("-"); unsigned long hex; while (found!=string::npos) {
int val = atoi(line.substr(0,found).c_str()); xtoi(HexList.at(val).c_str(), &hex); SetPenColorHex(hex); DrawFillBox(xpos,
ypos-S_HEIGHT, xpos+S_WIDTH, ypos); xpos += (S_WIDTH+1);
line = line.substr(found+1, line.length()); found=line.find_first_of("-");
} int val = atoi(line.c_str()); xtoi(HexList.at(val).c_str(), &hex); SetPenColorHex(hex); DrawFillBox(xpos, ypos-
S_HEIGHT, xpos+S_WIDTH, ypos);
xpos = 10; ypos = ypos - S_HEIGHT - 10;
} myfile.close();
}
//DrawFillBox(20, 20, 40, 70); glFlush();    /* Complete any pending operations */
}
void keyCB(unsigned char key, int x, int y) {
if( key == 'q' ) exit(0);
}
int main(int argc, char *argv[]) {
/* called on key press */
int win; glutInit(&argc, argv);/* initialize GLUT system */
glutInitDisplayMode(GLUT_RGB); glutInitWindowSize(WINDOW_X,WINDOW_Y);    /* width=400pixels
height=500pixels */ win = glutCreateWindow("Triangle");/* create window */
/* from this point on the current window is win */
}
}
glClearColor(1.0,1.0,1.0,1.0); gluOrtho2D(0,WINDOW_X,0,WINDOW_Y); glutDisplayFunc(displayCB);
glutKeyboardFunc(keyCB);
/* set background to black */ /* how object is mapped to window */
/* set window's display callback */ /* set window's key callback */
glutMainLoop(); /* execution never reaches this point */ return 0;
void SetPenColor(double red, double green, {
glColor3d(red,green,blue);
}
void SetPenColorHex(unsigned long color) {
double blue)
SetPenColor((color >> 16) / 256.0, (color >> 8 & 0xFF) / 256.0,
(color & 0xFF) / 256.0);
void DrawFillBox(double x0, double y0, double x1, double y1) {
}
glBegin(GL_POLYGON); glVertex2d(x0,y0); glVertex2d(x1,y0); glVertex2d(x1,y1); glVertex2d(x0,y1);
glVertex2d(x0,y0);
glEnd();
// Draw a connected line from // corner to
// Converts a hexadecimal string to integer int xtoi(const char* xs, unsigned long* result) {
size_t szlen = strlen(xs); int i, xv, fact;
if (szlen > 0) {
// Converting more than 32bit hexadecimal value? if (szlen>8) return 2; // exit
// Begin conversion here *result = 0; fact = 1;
// Run until no more character to convert for(i=szlen-1; i>=0 ;i--) {
if (isxdigit(*(xs+i))) {
if (*(xs+i)>=97) {
xv = ( *(xs+i) - 97) + 10;
} else if ( *(xs+i) >= 65) {
xv = (*(xs+i) - 65) + 10;
} else {
xv = *(xs+i) - 48;
} *result += (xv * fact); fact *= 16;
/* start processing events... */
```

```
// lcorner to // corner to // corner to // corner,
// then stop--we're finished.
}
} else { // Conversion was abnormally terminated // by non hexadecimal digit, hence // returning only the converted
with // an error value 4 (illegal hex character)
return 4;
}
// Nothing to convert return 1;
}
}
```

APPENDIX B: CHAPTER FOUR SUPPLEMENT

Table B.1. GenBank accession numbers for 16S rDNA, *dnaJ*, *rpoB*, and *tuf* gene fragments analyzed in this study.

No.	***Staphylococcus* Species**	**Subspecies**	**[a]Strain**	**[b]Genes**			
				16S	***dnaJ***	***rpoB***	***tuf***
1	*S. agnetis*		DSM 23656	HM484980.1	N/A	HM484993.1	HM485006.1
2	*S. arlettae*		ATCC 43957	AB009933.1	AB234056.1	AF325874.1	EU652781.1
3	*S. aureus*	*aureus*	ATCC 12600	D83357.1	AB234058.1	N/A	AB472826.1
4		*anaerobius*	ATCC 35844	D83355.1	AB234057.1	AF325894.1	HM352930.1
5	*S. auricularis*		ATCC 33753	D83358.1	AB234059.1	AF325889.1	EU652784.1
6	*S. capitis*	*capitis*	ATCC 27840	L37599.1	AB234060.1	AF325885.1	AF298798.1
7		*urealyticus*	ATCC 43926	AB009937.1	AB234061.1	DQ120729.1	EU652786.1
8	*S. caprae*		ATCC 35538	AB009935.1	AB234062.1	AF325896.1	EU652787.1
9	*S. carnosus*	*carnosus*	ATCC 51365	AB009934.1	AB234063.1	AF325880.1	EU652788.1
10		*utilis*	DSM 11676	AB233329.1	AB234064.1	DQ120730.1	EU652789.1
11	*S. chromogenes*		ATCC 43764	D83360.1	AB234065.1	AF325892.1	EU652790.1
12	*S. cohnii*	*cohnii*	ATCC 29974	D83361.1	AB234066.1	AF325893.1	EU652791.1
13		*urealyticus*	ATCC 49330	AB009936.1	AB234067.1	DQ120732.1	HM352939.1
14	*S. condimenti*		DSM 11674	Y15750.1	AB234068.1	DQ120733.1	EU652792.1
15	*S. delphini*		ATCC 49171	AB009938.1	AB234319.1	DQ120735.1	EU157611.1
16	*S. devriesei*		CCUG 58238	FJ389206.1	FJ907454.1	FJ389232.1	FJ389248.1
17	*S. epidermidis*		ATCC 14990	D83363.1	AB234069.1	AF325872.1	AF298800.1
18	*S. equorum*	*equorum*	ATCC 43958	AB009939.1	AB234070.1	AF325882.1	EU652795.1
19		*linens*	DSM 15097	AF527483.1	EU652838.1	DQ120736.1	EU652796.1
20	*S. felis*		ATCC 49168	D83364.1	AB234071.1	AF325878.1	EU652797.1
21	*S. fleurettii*		ATCC BAA274	AB233330.1	AB234072.1	DQ120737.1	HM352961.1
22	*S. gallinarum*		ATCC 35539	D83366.1	AB234073.1	AF325890.1	EU652799.1
23	*S. haemolyticus*		ATCC 29970	L37600.1	AB234074.1	AF325888.1	HM352923.1
24	*S. hominis*	*hominis*	ATCC 27844	L37601.1	AB234075.1	AF325875.1	EU652801.1
25		*novobiosepticus*	ATCC 700236	AB233326.1	AB234076.1	DQ120738.1	EU652802.1
26	*S. hyicus*		ATCC 11249	D83368.1	AB234077.1	AF325876.1	EU571080.1
27	*S. intermedius*		ATCC 29663	D83369.1	AB234078.1	AF325869.1	EU652804.1
28	*S. kloosii*		ATCC 43959	AB009940.1	AB234079.1	AF325891.1	EU652813.1
29	*S. lentus*		ATCC 29070	D83370.1	AB234080.1	AY036973.1	HM352944.1
30	*S. lugdunensis*		ATCC 43809	AB009941.1	AB234081.1	AF325870.1	AF298803.1
31	*S. lutrae*		ATCC 700373	AB233333.1	AB234082.1	DQ120739.1	EU652806.1
32	*S. massiliensis*		CCUG 55927	EU707796.1	EU652841.1	N/A	EU652827.1
33	*S. microti*		DSM 22147	EU888120.1	FN433124.1	EU888121.1	N/A
34	*S. muscae*		ATCC 49910	S83566.1	AB234083.1	AF325884.1	EU652807.1
35	*S. nepalensis*		DSM 15150	AJ517414.1	GQ222247.1	GQ222237.1	EU652808.1
36	*S. pasteuri*		ATCC 51129	AB009944.1	AB234084.1	DQ120742.1	EU652809.1
37	*S. pettenkoferi*		DSM 19554	DQ538517.1	EU652829.1	DQ120744.1	EU652810.1

38	*S. piscifermentans*		ATCC 51136	AF041359.1	AB234085.1	DQ120745.1	HM352955.1
39	*S. pseudintermedius*		CCUG 22219	AJ780976.1	EU652840.1	AM921786.1	EU157680.1
40	[c]*S. pulvereri*		ATCC 51698	AB009942.1	AB234086.1	AF325879.1	N/A
41	*S. rostri*		DSM 21968	FM242137.1	FM244714.1	FM242139.1	N/A
42	*S. saccharolyticus*		ATCC 14953	L37602.1	AB234087.1	AF325871.1	EU652814.1
43	*S. saprophyticus*	*bovis*	DSM 18669	AB233327.1	AB234088.1	DQ120746.1	HM352934.1
44		*saprophyticus*	ATCC 15305	D83371.2	AB234089.1	EF173662.1	EU571085.1
45	*S. schleiferi*	*schleiferi*	ATCC 43808	D83372.1	AB234321.1	AF325886.1	EU652818.1
46		*coagulans*	ATCC 49545	AB009945.1	AB234320.1	DQ120747.1	EU571086.1
47	*S. sciuri*	*carnaticus*	ATCC 700058	AB233331.1	AB234322.1	DQ120748.1	EU652819.1
48		*rodentium*	ATCC 700061	AB233332.1	AB234323.1	DQ120749.1	EU652820.1
49		*sciuri*	ATCC 29062	AJ421446.1	AB234324.1	HM146323.1	HM352947.1
50	*S. simiae*		DSM 17636	AY727530.2	GQ222248.1	EU888127.1	HM352931.1
51	*S. simulans*		ATCC 27848	D83373.1	AB234325.1	AF325877.1	EU571090.1
52	*S. stepanovicii*		CCM 7717	GQ222244.1	GQ222254.1	FJ906724.1	N/A
53	*S. succinus*	*succinus*	ATCC 700337	AF004220.1	AB234326.1	DQ120751.1	EU652824.1
54		*casei*	DSM 15096	AJ320272.1	EU652830.1	DQ120750.1	EU652823.1
55	*S. vitulinus*		ATCC 51145	AB009946.1	AB234327.1	DQ120752.1	EU652825.1
56	*S. warneri*		ATCC 27836	L37603.1	AB234328.1	AF325887.1	AF298806.1
57	*S. xylosus*		ATCC 29971	D83374.1	AB234329.1	AF325883.1	HM352950.1

[a]All are type strains; DSM, German Collection of Microorganisms and Cell Cultures; ATCC, American Type Culture Collection; CCUG, Culture Collection, University of Gothenburg; CCM, Czechoslovak Collection of Microorganisms.

[b]N/A, no sequence analyzed. Numbers indicated are GenBank accession numbers.

[c]Reclassified as a later synonym of *S. vitulinus* (181).

Table B.2. Evolutionary models for each partition were chosen based on AIC using jModelTest.

Partition	Model
All gene fragments	GTR + ΓI
16S, whole fragment	K80 + ΓI
16S, stem nucleotides	SYM + ΓI
16S, loop nucleotides	GTR + ΓI
All protein coding gene fragments	GTR + Γ
dnaJ, whole fragment	GTR + ΓI
dnaJ, codon positions 1 and 2	GTR + Γ
dnaJ, codon position 1	SYM + Γ
dnaJ, codon position 2	HKY + ΓI
dnaJ, codon position 3	GTR + Γ
rpoB, , whole fragment	SYM + ΓI
rpoB, codon positions 1 and 2	SYM + ΓI
rpoB, codon position 1	GTR + ΓI
rpoB, codon position 2	SYM + ΓI
rpoB, codon position 3	GTR + ΓI
tuf, whole fragment	GTR + Γ
tuf, codon positions 1 and 2	JC + Γ
tuf, codon position 1	GTR + Γ
tuf, codon position 2	GTR + Γ
tuf, codon position 3	GTR + Γ

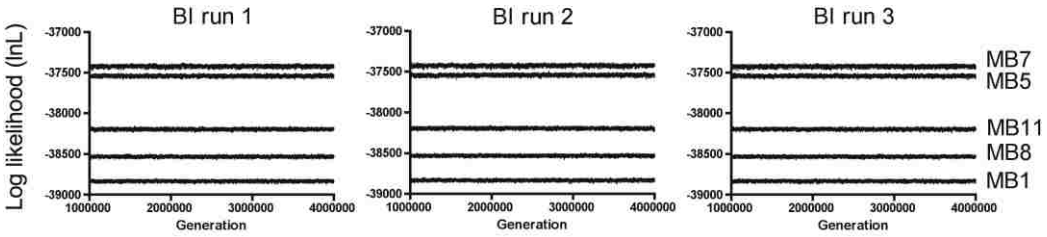

Figure B.1. Bayesian inferences of phylogeny are highly reproducible, regardless of model employed. Shown are plots of post-burnin generational log likelihoods (lnL) from five representative partitioning strategies across triplicate BI runs. All runs were highly reproducible regardless of partitioning strategy.

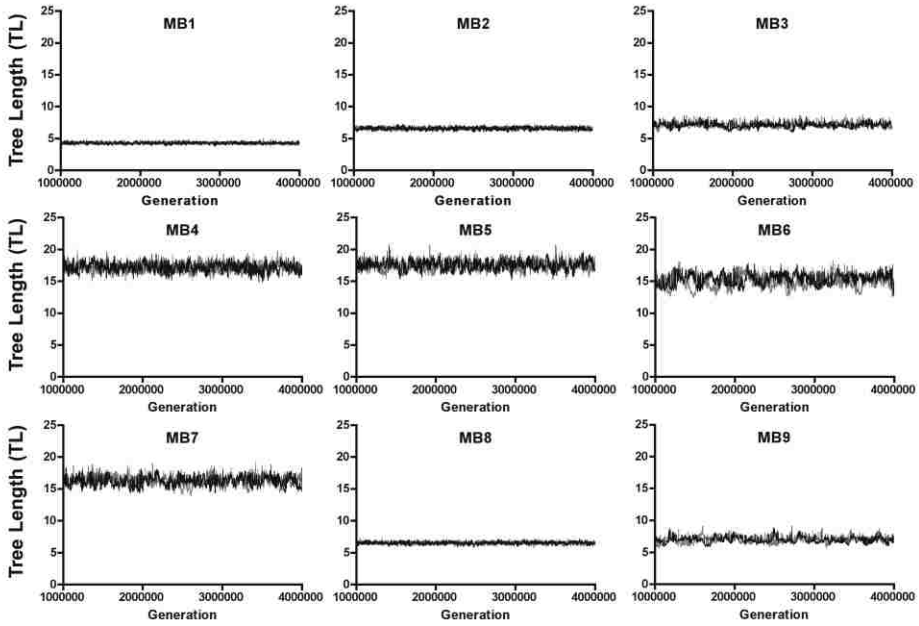

Figure B.2. Tree length (TL) analysis indicates that overparameterization may be occurring within more highly partitioned datasets. Shown are post-burnin generational TL estimates for partitioning strategies assessed in this study. Note that as the complexity of partitioning increases evidence of increased TL and failed convergence is observed.

Figure B.3. Model partitioning increases the mean tree length (TL) and run variance.

Shown is a box plot indicating the mean TL and 95% confidence interval among partitioning strategies.

REFERENCES

1. **Akaike, H.** 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control **AC-19:**716-723.

2. **Al Masalma, M., D. Raoult, and V. Roux.** 2010. *Staphylococcus massiliensis* sp. nov., isolated from a human brain abscess. Int J Syst Evol Microbiol **60:**1066-72.

3. **Altekar, G., S. Dwarkadas, J. P. Huelsenbeck, and F. Ronquist.** 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics **20:**407-15.

4. **Atkins, K. L., J. D. Burman, E. S. Chamberlain, J. E. Cooper, B. Poutrel, S. Bagby, A. T. Jenkins, E. J. Feil, and J. M. van den Elsen.** 2008. *S. aureus* IgG-binding proteins SpA and Sbi: host specificity and mechanisms of immune complex formation. Mol Immunol **45:**1600-11.

5. **Baba, T., T. Bae, O. Schneewind, F. Takeuchi, and K. Hiramatsu.** 2008. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. J Bacteriol **190:**300-10.

6. **Baba, T., K. Kuwahara-Arai, I. Uchiyama, F. Takeuchi, T. Ito, and K. Hiramatsu.** 2009. Complete genome sequence of *Macrococcus caseolyticus* strain JCSCS5402, [corrected] reflecting the ancestral genome of the human-pathogenic staphylococci. J Bacteriol **191:**1180-90.

7. **Baba, T., F. Takeuchi, M. Kuroda, H. Yuzawa, K. Aoki, A. Oguchi, Y. Nagai, N. Iwama, K. Asano, T. Naimi, H. Kuroda, L. Cui, K. Yamamoto, and K. Hiramatsu.** 2002. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet **359:**1819-27.

8. **Bernhart, S. H., I. L. Hofacker, S. Will, A. R. Gruber, and P. F. Stadler.** 2008. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics **9:**474.

9. **Blaiotta, G., V. Fusco, D. Ercolini, O. Pepe, and S. Coppola.** 2010. Diversity of *Staphylococcus* species strains based on partial *kat* (catalase) gene sequences and design of a PCR-restriction fragment length polymorphism assay for identification and differentiation of coagulase-positive species (*S. aureus, S. delphini, S. hyicus, S. intermedius, S. pseudintermedius*, and *S. schleiferi* subsp. *coagulans*). J Clin Microbiol **48:**192-201.

10. **Boelaert, J. R., H. W. Van Landuyt, C. A. Godard, R. F. Daneels, M. L. Schurgers, E. G. Matthys, Y. A. De Baere, D. W. Gheyle, B. Z. Gordts, and L. A. Herwaldt.** 1993. Nasal mupirocin ointment decreases the incidence of *Staphylococcus aureus* bacteraemias in haemodialysis patients. Nephrol Dial Transplant **8:**235-9.

11. **Bos, D. H., and D. Posada.** 2005. Using models of nucleotide evolution to build phylogenetic trees. Dev Comp Immunol **29:**211-227.

12. **Boucher, H. W., and G. R. Corey.** 2008. Epidemiology of methicillin-resistant *Staphylococcus aureus*. Clin Infect Dis **46 Suppl 5:**S344-9.

13. **Brandley, M. C., A. Schmitz, and T. W. Reeder.** 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst Biol **54:**373-90.

14. **Bukharin, O. V., O. L. Kartashova, S. B. Kirgizova, and I. V. Valysheva.** 2005. Antilactoferrin activity of microorganisms. Zh Mikrobiol Epidemiol Immunobiol**:**7-10.

15. **Burman, J. D., E. Leung, K. L. Atkins, M. N. O'Seaghdha, L. Lango, P. Bernado, S. Bagby, D. I. Svergun, T. J. Foster, D. E. Isenman, and J. M. van den Elsen.** 2008. Interaction of human complement with Sbi, a staphylococcal immunoglobulin-binding protein: indications of a novel mechanism of complement evasion by *Staphylococcus aureus*. J Biol Chem **283:**17579-93.

16. **Carstens, B. C., and L. L. Knowles.** 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from Melanoplus grasshoppers. Syst Biol **56:**400-11.

17. **Castoe, T. A., A. P. de Koning, H. M. Kim, W. Gu, B. P. Noonan, G. Naylor, Z. J. Jiang, C. L. Parkinson, and D. D. Pollock.** 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc Natl Acad Sci U S A **106:**8986-91.

18. **Castoe, T. A., T. M. Doan, and C. L. Parkinson.** 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of Gymnophthalmid lizards. Syst Biol **53:**448-69.

19. **Castoe, T. A., and C. L. Parkinson.** 2006. Bayesian mixed models and the phylogeny of pitvipers (*Viperidae: Serpentes*). Mol Phylogenet Evol **39:**91-110.

20. **Castoe, T. A., M. M. Sasa, and C. L. Parkinson.** 2005. Modeling nucleotide evolution at the mesoscale: the phylogeny of the neotropical pitvipers of the Porthidium group (*viperidae: crotalinae*). Mol Phylogenet Evol **37:**881-98.

21. **Chang, S., D. M. Sievert, J. C. Hageman, M. L. Boulton, F. C. Tenover, F. P. Downes, S. Shah, J. T. Rudrik, G. R. Pupp, W. J. Brown, D. Cardo, and S. K.**

Fridkin. 2003. Infection with vancomycin-resistant *Staphylococcus aureus* containing the *vanA* resistance gene. N Engl J Med **348:**1342-7.

22. **Clarke, S. R., and S. J. Foster.** 2006. Surface adhesins of *Staphylococcus aureus*. Adv Microb Physiol **51:**187-224.

23. **Clement, M., D. Posada, and K. A. Crandall.** 2000. TCS: a computer program to estimate gene genealogies. Mol Ecol **9:**1657-9.

24. **Coates, T., R. Bax, and A. Coates.** 2009. Nasal decolonization of *Staphylococcus aureus* with mupirocin: strengths, weaknesses and future prospects. J Antimicrob Chemother **64:**9-15.

25. **Cole, A. L., A. Herasimtschuk, P. Gupta, A. J. Waring, R. I. Lehrer, and A. M. Cole.** 2007. The retrocyclin analogue RC-101 prevents human immunodeficiency virus type 1 infection of a model human cervicovaginal tissue construct. Immunology **121:**140-5.

26. **Cole, A. L., O. O. Yang, A. D. Warren, A. J. Waring, R. I. Lehrer, and A. M. Cole.** 2006. HIV-1 adapts to a retrocyclin with cationic amino acid substitutions that reduce fusion efficiency of gp41. J Immunol **176:**6900-5.

27. **Cole, A. M., P. Dewan, and T. Ganz.** 1999. Innate antimicrobial activity of nasal secretions. Infect Immun **67:**3267-75.

28. **Cole, A. M., T. Hong, L. M. Boo, T. Nguyen, C. Zhao, G. Bristol, J. A. Zack, A. J. Waring, O. O. Yang, and R. I. Lehrer.** 2002. Retrocyclin: a primate peptide that protects cells from infection by T- and M-tropic strains of HIV-1. Proc Natl Acad Sci U S A **99:**1813-8.

29. **Cole, A. M., D. L. Patton, L. C. Rohan, A. L. Cole, Y. Cosgrove-Sweeney, N. A. Rogers, D. Ratner, A. B. Sassi, C. Lackman-Smith, P. Tarwater, B. Ramratnam, P. Ruchala, R. I. Lehrer, A. J. Waring, and P. Gupta.** 2010. The formulated microbicide RC-101 was safe and antivirally active following intravaginal application in pigtailed macaques. PLoS One **5:**e15111.

30. **Cole, A. M., S. Tahk, A. Oren, D. Yoshioka, Y. H. Kim, A. Park, and T. Ganz.** 2001. Determinants of *Staphylococcus aureus* nasal carriage. Clin Diagn Lab Immunol **8:**1064-9.

31. **Cole, A. M., M. Wu, Y. H. Kim, and T. Ganz.** 2000. Microanalysis of antimicrobial properties of human fluids. J Microbiol Methods **41:**135-43.

32. **Connolly, S., W. C. Noble, and I. Phillips.** 1993. Mupirocin resistance in coagulase-negative staphylococci. J Med Microbiol **39:**450-3.

33. **Conover, M. S., G. P. Sloan, C. F. Love, N. Sukumar, and R. Deora.** 2010. The Bps polysaccharide of *Bordetella pertussis* promotes colonization and biofilm formation in the nose by functioning as an adhesin. Mol Microbiol **77:**1439-55.

34. **Cookson, B. D.** 1990. Mupirocin resistance in staphylococci. J Antimicrob Chemother **25:**497-501.

35. **Cookson, B. D.** 1998. The emergence of mupirocin resistance: a challenge to infection control and antibiotic prescribing practice. J Antimicrob Chemother **41:**11-8.

36. **Corrigan, R. M., H. Miajlovic, and T. J. Foster.** 2009. Surface proteins that promote adherence of *Staphylococcus aureus* to human desquamated nasal epithelial cells. BMC Microbiol **9:**22.

37. **Crisostomo, M. I., H. Westh, A. Tomasz, M. Chung, D. C. Oliveira, and H. de Lencastre.** 2001. The evolution of methicillin resistance in *Staphylococcus aureus*: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. Proc Natl Acad Sci U S A **98:**9865-70.

38. **Daly, N. L., Y. K. Chen, K. J. Rosengren, U. C. Marx, M. L. Phillips, A. J. Waring, W. Wang, R. I. Lehrer, and D. J. Craik.** 2007. Retrocyclin-2: structural analysis of a potent anti-HIV theta-defensin. Biochemistry **46:**9920-8.

39. **de Lencastre, H., M. Chung, and H. Westh.** 2000. Archaic strains of methicillin-resistant *Staphylococcus aureus*: molecular and microbiological properties of isolates from the 1960s in Denmark. Microb Drug Resist **6:**1-10.

40. **Degnan, J. H., and N. A. Rosenberg.** 2006. Discordance of species trees with their most likely gene trees. PLoS Genet **2:**e68.

41. **Degnan, J. H., and N. A. Rosenberg.** 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol **24:**332-40.

42. **Deurenberg, R. H., and E. E. Stobberingh.** 2008. The evolution of *Staphylococcus aureus*. Infect Genet Evol **8:**747-63.

43. **Diep, B. A., S. R. Gill, R. F. Chang, T. H. Phan, J. H. Chen, M. G. Davidson, F. Lin, J. Lin, H. A. Carleton, E. F. Mongodin, G. F. Sensabaugh, and F. Perdreau-Remington.** 2006. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. Lancet **367:**731-9.

44. **Drancourt, M., and D. Raoult.** 2002. *rpoB* gene sequence-based identification of *Staphylococcus* species. J Clin Microbiol **40:**1333-8.

45. **Edwards, S. V., L. Liu, and D. K. Pearl.** 2007. High-resolution species trees without concatenation. Proc Natl Acad Sci U S A **104:**5936-41.

46. **Enright, M. C., N. P. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt.** 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J Clin Microbiol **38:**1008-15.

47. **Enright, M. C., D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, and B. G. Spratt.** 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc Natl Acad Sci U S A **99:**7687-92.

48. **Erixon, P., B. Svennblad, T. Britton, and B. Oxelman.** 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst Biol **52:**665-73.

49. **Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. Day.** 2003. How clonal is *Staphylococcus aureus*? J Bacteriol **185:**3307-16.

50. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol **186:**1518-30.

51. **Felsenstein, J.** 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol **17:**368-76.

52. **Fournier, B., and D. J. Philpott.** 2005. Recognition of *Staphylococcus aureus* by the innate immune system. Clin Microbiol Rev **18:**521-40.

53. **Freney, J., W. E. Kloos, V. Hajek, J. A. Webster, M. Bes, Y. Brun, and C. Vernozy-Rozand.** 1999. Recommended minimal standards for description of new staphylococcal species. Subcommittee on the taxonomy of staphylococci and streptococci of the International Committee on Systematic Bacteriology. Int J Syst Bacteriol **49 Pt 2:**489-502.

54. **Fuhrman, C. A., A. D. Warren, A. J. Waring, S. M. Dutz, S. Sharma, R. I. Lehrer, A. L. Cole, and A. M. Cole.** 2007. Retrocyclin RC-101 overcomes cationic mutations on the heptad repeat 2 region of HIV-1 gp41. Febs J **274:**6477-87.

55. **Fuller, A. T., G. Mellows, M. Woolford, G. T. Banks, K. D. Barrow, and E. B. Chain.** 1971. Pseudomonic acid: an antibiotic produced by *Pseudomonas fluorescens*. Nature **234:**416-7.

56. **Garza-Gonzalez, E., D. Lopez, C. Pezina, W. Muruet, V. Bocanegra-Garcia, I. Munoz, C. Ramirez, and J. M. Llaca-Diaz.** 2010. Diversity of staphylococcal cassette

chromosome mec structures in coagulase-negative staphylococci and relationship to drug resistance. J Med Microbiol **59:**323-9.

57. **Ghebremedhin, B., F. Layer, W. Konig, and B. Konig.** 2008. Genetic classification and distinguishing of *Staphylococcus* species based on different partial *gap*, 16S rRNA, *hsp60, rpoB, sodA,* and *tuf* gene sequences. J Clin Microbiol **46:**1019-25.

58. **Gill, S. R., D. E. Fouts, G. L. Archer, E. F. Mongodin, R. T. Deboy, J. Ravel, I. T. Paulsen, J. F. Kolonay, L. Brinkac, M. Beanan, R. J. Dodson, S. C. Daugherty, R. Madupu, S. V. Angiuoli, A. S. Durkin, D. H. Haft, J. Vamathevan, H. Khouri, T. Utterback, C. Lee, G. Dimitrov, L. Jiang, H. Qin, J. Weidman, K. Tran, K. Kang, I. R. Hance, K. E. Nelson, and C. M. Fraser.** 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J Bacteriol **187:**2426-38.

59. **Gillaspy, A. F., V. Worrell, J. Orvis, B. A. Roe, W. Dyer, and J. J. Iandolo.** 2006. The *Staphylococcus aureus* NCTC 8325 genome, p. 381-412. *In* V. A. Fischetti, R. Novick, J. Ferretti, D. Portnoy, and J. Rood (ed.), Gram positive pathogens. ASM Press, Washington, DC.

60. **Gomes, A. R., S. Vinga, M. Zavolan, and H. de Lencastre.** 2005. Analysis of the genetic variability of virulence-related loci in epidemic clones of methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother **49:**366-79.

61. **Graur, D., and W. Li.** 2000. Fundamentals of Molecular Evolution, Second ed. Sinauer Associates, Inc., Sunderland.

62. **Gribaldo, S., B. Cookson, N. Saunders, R. Marples, and J. Stanley.** 1997. Rapid identification by specific PCR of coagulase-negative staphylococcal species important in hospital infection. J Med Microbiol **46:**45-53.

63. **Grundmeier, M., M. Hussain, P. Becker, C. Heilmann, G. Peters, and B. Sinha.** 2004. Truncation of fibronectin-binding proteins in *Staphylococcus aureus* strain Newman leads to deficient adherence and host cell invasion due to loss of the cell wall anchor function. Infect Immun **72:**7155-63.

64. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:**696-704.

65. **Hacker, J., and E. Carniel.** 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep **2:**376-81.

66. **Hall, T. A.** 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids. Symp. Ser. **41:**95-98.

67. **Hanssen, A. M., G. Kjeldsen, and J. U. Sollid.** 2004. Local variants of Staphylococcal cassette chromosome mec in sporadic methicillin-resistant *Staphylococcus aureus* and methicillin-resistant coagulase-negative Staphylococci: evidence of horizontal gene transfer? Antimicrob Agents Chemother **48:**285-96.

68. **Hartford, O., P. Francois, P. Vaudaux, and T. J. Foster.** 1997. The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. Mol Microbiol **25:**1065-76.

69. **Hashimoto, M., K. Tawaratsumida, H. Kariya, K. Aoyama, T. Tamura, and Y. Suda.** 2006. Lipoprotein is a predominant Toll-like receptor 2 ligand in *Staphylococcus aureus* cell wall components. Int Immunol **18:**355-62.

70. **Hauschild, T.** 2001. Phenotypic and genotypic identification of staphylococci isolated from wild small mammals. Syst Appl Microbiol **24:**411-6.

71. **Hauschild, T., S. Stepanovic, and J. Zakrzewska-Czerwinska.** 2010. *Staphylococcus stepanovicii* sp. nov., a novel novobiocin-resistant oxidase-positive staphylococcal species isolated from wild small mammals. Syst Appl Microbiol **33:**183-7.

72. **Hidron, A. I., C. E. Low, E. G. Honig, and H. M. Blumberg.** 2009. Emergence of community-acquired meticillin-resistant *Staphylococcus aureus* strain USA300 as a cause of necrotising community-onset pneumonia. Lancet Infect Dis **9:**384-92.

73. **Highlander, S. K., K. G. Hulten, X. Qin, H. Jiang, S. Yerrapragada, E. O. Mason, Jr., Y. Shang, T. M. Williams, R. M. Fortunov, Y. Liu, O. Igboeli, J. Petrosino, M. Tirumalai, A. Uzman, G. E. Fox, A. M. Cardenas, D. M. Muzny, L. Hemphill, Y. Ding, S. Dugan, P. R. Blyth, C. J. Buhay, H. H. Dinh, A. C. Hawes, M. Holder, C. L. Kovar, S. L. Lee, W. Liu, L. V. Nazareth, Q. Wang, J. Zhou, S. L. Kaplan, and G. M. Weinstock.** 2007. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. BMC Microbiol **7:**99.

74. **Hiramatsu, K., L. Cui, M. Kuroda, and T. Ito.** 2001. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. Trends Microbiol **9:**486-93.

75. **Hodgson, J. E., S. P. Curnock, K. G. Dyke, R. Morris, D. R. Sylvester, and M. S. Gross.** 1994. Molecular characterization of the gene encoding high-level mupirocin resistance in *Staphylococcus aureus* J2870. Antimicrob Agents Chemother **38:**1205-8.

76. **Hofacker, I. L., M. Fekete, and P. F. Stadler.** 2002. Secondary structure prediction for aligned RNA sequences. J Mol Biol **319:**1059-66.

77. **Holden, M. T., E. J. Feil, J. A. Lindsay, S. J. Peacock, N. P. Day, M. C. Enright, T. J. Foster, C. E. Moore, L. Hurst, R. Atkin, A. Barron, N. Bason, S. D. Bentley, C.**

Chillingworth, T. Chillingworth, C. Churcher, L. Clark, C. Corton, A. Cronin, J. Doggett, L. Dowd, T. Feltwell, Z. Hance, B. Harris, H. Hauser, S. Holroyd, K. Jagels, K. D. James, N. Lennard, A. Line, R. Mayes, S. Moule, K. Mungall, D. Ormond, M. A. Quail, E. Rabbinowitsch, K. Rutherford, M. Sanders, S. Sharp, M. Simmonds, K. Stevens, S. Whitehead, B. G. Barrell, B. G. Spratt, and J. Parkhill. 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci U S A **101:**9786-91.

78. **Holden, M. T., J. A. Lindsay, C. Corton, M. A. Quail, J. D. Cockfield, S. Pathak, R. Batra, J. Parkhill, S. D. Bentley, and J. D. Edgeworth.** 2010. Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). J Bacteriol **192:**888-92.

79. **Holland, B., F. Delsuc, and V. Moulton.** 2005. Visualizing conflicting evolutionary hypotheses in large collections of trees: using consensus networks to study the origins of placentals and hexapods. Syst Biol **54:**66-76.

80. **Holton, D. L., L. E. Nicolle, D. Diley, and K. Bernstein.** 1991. Efficacy of mupirocin nasal ointment in eradicating *Staphylococcus aureus* nasal carriage in chronic haemodialysis patients. J Hosp Infect **17:**133-7.

81. **Howden, B. P., J. K. Davies, P. D. Johnson, T. P. Stinear, and M. L. Grayson.** 2010. Reduced vancomycin susceptibility in *Staphylococcus aureus*, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications. Clin Microbiol Rev **23:**99-139.

82. **Huelsenbeck, J. P., and B. Rannala.** 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science **276:**227-32.

83. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754-5.

84. **Hughes, J., and G. Mellows.** 1978. Inhibition of isoleucyl-transfer ribonucleic acid synthetase in *Escherichia coli* by pseudomonic acid. Biochem J **176:**305-18.

85. **Hunter, P. R.** 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol **28:**1903-5.

86. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol **26:**2465-6.

87. **Hurdle, J. G., A. J. O'Neill, E. Ingham, C. Fishwick, and I. Chopra.** 2004. Analysis of mupirocin resistance and fitness in *Staphylococcus aureus* by molecular genetic and structural modeling techniques. Antimicrob Agents Chemother **48:**4366-76.

88. **Hurdle, J. G., A. J. O'Neill, L. Mody, I. Chopra, and S. F. Bradley.** 2005. In vivo transfer of high-level mupirocin resistance from *Staphylococcus epidermidis* to methicillin-resistant *Staphylococcus aureus* associated with failure of mupirocin prophylaxis. J Antimicrob Chemother **56:**1166-8.

89. **Hurvich, C. M., and T. Chih-Ling.** 1988. Regression and time series model selection in small samples. Biometrika **76:**297-307.

90. **Ito, T., Y. Katayama, and K. Hiramatsu.** 1999. Cloning and nucleotide sequence determination of the entire mec DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. Antimicrob Agents Chemother **43:**1449-58.

91. **John, J. F., and A. M. Harvin.** 2007. History and evolution of antibiotic resistance in coagulase-negative staphylococci: Susceptibility profiles of new anti-staphylococcal agents. Ther Clin Risk Manag **3:**1143-52.

92. **Josefsson, E., J. Higgins, T. J. Foster, and A. Tarkowski.** 2008. Fibrinogen binding sites P336 and Y338 of clumping factor A are crucial for *Staphylococcus aureus* virulence. PLoS ONE **3:**e2206.

93. **Kaliner, M. A.** 1991. Human nasal respiratory secretions and host defense. Am Rev Respir Dis **144:**S52-6.

94. **Katayama, Y., T. Ito, and K. Hiramatsu.** 2000. A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother **44:**1549-55.

95. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **16:**111-20.

96. **King, M. D., B. J. Humphrey, Y. F. Wang, E. V. Kourbatova, S. M. Ray, and H. M. Blumberg.** 2006. Emergence of community-acquired methicillin-resistant *Staphylococcus aureus* USA 300 clone as the predominant cause of skin and soft-tissue infections. Ann Intern Med **144:**309-17.

97. **Kleeman, K. T., T. L. Bannerman, and W. E. Kloos.** 1993. Species distribution of coagulase-negative staphylococcal isolates at a community hospital and implications for selection of staphylococcal identification procedures. J Clin Microbiol **31:**1318-21.

98. **Klein, E., D. L. Smith, and R. Laxminarayan.** 2007. Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*, United States, 1999-2005. Emerg Infect Dis **13:**1840-6.

99. **Klevens, R. M., M. A. Morrison, J. Nadle, S. Petit, K. Gershman, S. Ray, L. H. Harrison, R. Lynfield, G. Dumyati, J. M. Townes, A. S. Craig, E. R. Zell, G. E. Fosheim, L. K. McDougal, R. B. Carey, and S. K. Fridkin.** 2007. Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. Jama **298:**1763-71.

100. **Kloos, W. E., D. N. Ballard, C. G. George, J. A. Webster, R. J. Hubner, W. Ludwig, K. H. Schleifer, F. Fiedler, and K. Schubert.** 1998. Delimiting the genus *Staphylococcus* through description of *Macrococcus caseolyticus* gen. nov., comb. nov. and *Macrococcus equipercicus* sp. nov., and *Macrococcus bovicus* sp. no. and *Macrococcus carouselicus* sp. nov. Int J Syst Bacteriol **48 Pt 3:**859-77.

101. **Kloos, W. E., and T. L. Bannerman.** 1994. Update on clinical significance of coagulase-negative staphylococci. Clin Microbiol Rev **7:**117-40.

102. **Kloos, W. E., and C. G. George.** 1991. Identification of *Staphylococcus* species and subspecies with the MicroScan Pos ID and Rapid Pos ID panel systems. J Clin Microbiol **29:**738-44.

103. **Kloos, W. E., and K. H. Schleifer.** 1986. Genus IV *Staphylococcus*., p. 1013-1035. *In* P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt (ed.), Bergey's Manual of Systematic Bacteriology, vol. 2. Williams & Wilkins, Baltimore.

104. **Kluytmans, J., A. van Belkum, and H. Verbrugh.** 1997. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. Clin Microbiol Rev **10:**505-20.

105. **Kluytmans, J. A., J. W. Mouton, M. F. VandenBergh, M. J. Manders, A. P. Maat, J. H. Wagenvoort, M. F. Michel, and H. A. Verbrugh.** 1996. Reduction of surgical-site infections in cardiothoracic surgery by elimination of nasal carriage of *Staphylococcus aureus*. Infect Control Hosp Epidemiol **17:**780-5.

106. **Koreen, L., S. V. Ramaswamy, E. A. Graviss, S. Naidich, J. M. Musser, and B. N. Kreiswirth.** 2004. *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. J Clin Microbiol **42:**792-9.

107. **Koreen, L., S. V. Ramaswamy, S. Naidich, I. V. Koreen, G. R. Graff, E. A. Graviss, and B. N. Kreiswirth.** 2005. Comparative sequencing of the serine-aspartate repeat-encoding region of the clumping factor B gene (*clfB*) for resolution within clonal groups of *Staphylococcus aureus*. J Clin Microbiol **43:**3985-94.

108. **Kuehnert, M. J., H. A. Hill, B. A. Kupronis, J. I. Tokars, S. L. Solomon, and D. B. Jernigan.** 2005. Methicillin-resistant-*Staphylococcus aureus* hospitalizations, United States. Emerg Infect Dis **11:**868-72.

109. **Kuhn, G., P. Francioli, and D. S. Blanc.** 2007. Double-locus sequence typing using *clfB* and *spa*, a fast and simple method for epidemiological typing of methicillin-resistant *Staphylococcus aureus*. J Clin Microbiol **45:**54-62.

110. **Kuhn, G., P. Francioli, and D. S. Blanc.** 2006. Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. J Bacteriol **188:**169-78.

111. **Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu.** 2001. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. Lancet **357:**1225-40.

112. **Kwok, A. Y., S. C. Su, R. P. Reynolds, S. J. Bay, Y. Av-Gay, N. J. Dovichi, and A. W. Chow.** 1999. Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus *Staphylococcus*. Int J Syst Bacteriol **49 Pt 3:**1181-92.

113. **Lamers, R. P., J. W. Stinnett, G. Muthukrishnan, C. L. Parkinson, and A. M. Cole.** 2011. Evolutionary Analyses of *Staphylococcus aureus* Identify Genetic Relationships between Nasal Carriage and Clinical Isolates. PLoS One **6:**e16426.

114. **Leonova, L., V. N. Kokryakov, G. Aleshina, T. Hong, T. Nguyen, C. Zhao, A. J. Waring, and R. I. Lehrer.** 2001. Circular minidefensins and posttranslational generation of molecular diversity. J Leukoc Biol **70:**461-4.

115. **Levine, D. P.** 2006. Vancomycin: a history. Clin Infect Dis **42 Suppl 1:**S5-12.

116. **Librado, P., and J. Rozas.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:**1451-2.

117. **Lindsay, J. A., and M. T. Holden.** 2006. Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. Funct Integr Genomics **6:**186-201.

118. **Liu, L.** 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics **24:**2542-3.

119. **Liu, L., and D. K. Pearl.** 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst Biol **56:**504-14.

120. **Lowder, B. V., C. M. Guinane, N. L. Ben Zakour, L. A. Weinert, A. Conway-Morris, R. A. Cartwright, A. J. Simpson, A. Rambaut, U. Nubel, and J. R. Fitzgerald.** 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. Proc Natl Acad Sci U S A **106:**19545-50.

121. **Mainous, A. G., 3rd, W. J. Hueston, C. J. Everett, and V. A. Diaz.** 2006. Nasal carriage of *Staphylococcus aureus* and methicillin-resistant *S. aureus* in the United States, 2001-2002. Ann Fam Med **4:**132-7.

122. **Mani, N., L. M. Baddour, D. Q. Offutt, U. Vijaranakul, M. J. Nadakavukaren, and R. K. Jayaswal.** 1994. Autolysis-defective mutant of *Staphylococcus aureus*: pathological considerations, genetic mapping, and electron microscopic studies. Infect Immun **62:**1406-9.

123. **Marshall, D. C.** 2010. Cryptic failure of partitioned Bayesian phylogenetic analyses: lost in the land of long trees. Syst Biol **59:**108-17.

124. **Marshall, D. C., C. Simon, and T. R. Buckley.** 2006. Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. Syst Biol **55:**993-1003.

125. **Martineau, F., F. J. Picard, D. Ke, S. Paradis, P. H. Roy, M. Ouellette, and M. G. Bergeron.** 2001. Development of a PCR assay for identification of staphylococci at genus and species levels. J Clin Microbiol **39:**2541-7.

126. **Matsuhashi, M., M. D. Song, F. Ishino, M. Wachi, M. Doi, M. Inoue, K. Ubukata, N. Yamashita, and M. Konno.** 1986. Molecular cloning of the gene of a penicillin-binding protein supposed to cause high resistance to beta-lactam antibiotics in *Staphylococcus aureus*. J Bacteriol **167:**975-80.

127. **Melles, D. C., R. F. Gorkink, H. A. Boelens, S. V. Snijders, J. K. Peeters, M. J. Moorhouse, P. J. van der Spek, W. B. van Leeuwen, G. Simons, H. A. Verbrugh, and A. van Belkum.** 2004. Natural population dynamics and expansion of pathogenic clones of *Staphylococcus aureus*. J Clin Invest **114:**1732-40.

128. **Midorikawa, K., K. Ouhara, H. Komatsuzawa, T. Kawai, S. Yamada, T. Fujiwara, K. Yamazaki, K. Sayama, M. A. Taubman, H. Kurihara, K. Hashimoto, and M. Sugai.** 2003. *Staphylococcus aureus* susceptibility to innate antimicrobial peptides, beta-defensins and CAP18, expressed by human keratinocytes. Infect Immun **71:**3730-9.

129. **Miller, M., H. A. Cook, E. Y. Furuya, M. Bhat, M. H. Lee, P. Vavagiakis, P. Visintainer, G. Vasquez, E. Larson, and F. D. Lowy.** 2009. *Staphylococcus aureus* in the community: colonization versus infection. PLoS One **4:**e6708.

130. **Moran, G. J., A. Krishnadasan, R. J. Gorwitz, G. E. Fosheim, L. K. McDougal, R. B. Carey, and D. A. Talan.** 2006. Methicillin-resistant *S. aureus* infections among patients in the emergency department. N Engl J Med **355:**666-74.

131. **Morton, T. M., J. L. Johnston, J. Patterson, and G. L. Archer.** 1995. Characterization of a conjugative staphylococcal mupirocin resistance plasmid. Antimicrob Agents Chemother **39:**1272-80.

132. **Munoz, P., J. Hortal, M. Giannella, J. M. Barrio, M. Rodriguez-Creixems, M. J. Perez, C. Rincon, and E. Bouza.** 2008. Nasal carriage of *S. aureus* increases the risk of surgical site infection after major heart surgery. J Hosp Infect **68:**25-31.

133. **Nakama, T., O. Nureki, and S. Yokoyama.** 2001. Structural basis for the recognition of isoleucyl-adenylate and an antibiotic, mupirocin, by isoleucyl-tRNA synthetase. J Biol Chem **276:**47387-93.

134. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3:**418-26.

135. **Nekhotiaeva, N., S. K. Awasthi, P. E. Nielsen, and L. Good.** 2004. Inhibition of *Staphylococcus aureus* gene expression and growth using antisense peptide nucleic acids. Mol Ther **10:**652-9.

136. **Neoh, H. M., L. Cui, H. Yuzawa, F. Takeuchi, M. Matsuo, and K. Hiramatsu.** 2008. Mutated response regulator graR is responsible for phenotypic conversion of Staphylococcus aureus from heterogeneous vancomycin-intermediate resistance to vancomycin-intermediate resistance. Antimicrob Agents Chemother **52:**45-53.

137. **Nguyen, T. X., A. M. Cole, and R. I. Lehrer.** 2003. Evolution of primate theta-defensins: a serpentine path to a sweet tooth. Peptides **24:**1647-54.

138. **Ni Eidhin, D., S. Perkins, P. Francois, P. Vaudaux, M. Hook, and T. J. Foster.** 1998. Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of Staphylococcus aureus. Mol Microbiol **30:**245-57.

139. **Nouwen, J., H. Boelens, A. van Belkum, and H. Verbrugh.** 2004. Human factor in Staphylococcus aureus nasal carriage. Infect Immun **72:**6685-8.

140. **Nouwen, J. L., M. W. Fieren, S. Snijders, H. A. Verbrugh, and A. van Belkum.** 2005. Persistent (not intermittent) nasal carriage of Staphylococcus aureus is the determinant of CPD-related infections. Kidney Int **67:**1084-92.

141. **Novakova, D., R. Pantucek, Z. Hubalek, E. Falsen, H. J. Busse, P. Schumann, and I. Sedlacek.** 2010. *Staphylococcus microti* sp. nov., isolated from the common vole (*Microtus arvalis*). Int J Syst Evol Microbiol **60:**566-73.

142. **Nubel, U., J. Dordel, K. Kurt, B. Strommenger, H. Westh, S. K. Shukla, H. Zemlickova, R. Leblois, T. Wirth, T. Jombart, F. Balloux, and W. Witte.** 2010. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. PLoS Pathog **6:**e1000855.

143. **Nylander, J. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey.** 2004. Bayesian phylogenetic analysis of combined data. Syst Biol **53:**47-67.

144. **O'Neill, E., H. Humphreys, and J. P. O'Gara.** 2009. Carriage of both the *fnbA* and *fnbB* genes and growth at 37 degrees C promote FnBP-mediated biofilm development in meticillin-resistant *Staphylococcus aureus* clinical isolates. J Med Microbiol **58:**399-402.

145. **Ogston, A.-.** 1984. Classics in Infectious Diseases. Rev Infect Dis **6:**122-8.

146. **Owen, S. M., D. L. Rudolph, W. Wang, A. M. Cole, A. J. Waring, R. B. Lal, and R. I. Lehrer.** 2004. RC-101, a retrocyclin-1 analogue with enhanced activity against primary HIV type 1 isolates. AIDS Res Hum Retroviruses **20:**1157-65.

147. **Palavecino, E.** 2007. Clinical, Epidemiological, and Laboratory Aspects of Methicillin-Resistant *Staphylococcus aureus* (MRSA) Infections, p. 1-19. *In* Y. Ji (ed.), Methicillin-Resistant *Staphylococcus aureus* (MRSA) Protocols, vol. 391. Humana Press Inc., New Jersey.

148. **Parenti, M. A., S. M. Hatfield, and J. J. Leyden.** 1987. Mupirocin: a topical antibiotic with a unique structure and mechanism of action. Clin Pharm **6:**761-70.

149. **Patti, J. M., B. L. Allen, M. J. McGavin, and M. Hook.** 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. Annu Rev Microbiol **48:**585-617.

150. **Peacock, S. J., I. de Silva, and F. D. Lowy.** 2001. What determines nasal carriage of *Staphylococcus aureus*? Trends Microbiol **9:**605-10.

151. **Posada, D.** 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol **25:**1253-6.

152. **Poyart, C., G. Quesne, C. Boumaila, and P. Trieu-Cuot.** 2001. Rapid and accurate species-level identification of coagulase-negative staphylococci by using the *sodA* gene as a target. J Clin Microbiol **39:**4296-301.

153. **Quinn, G. A., and A. M. Cole.** 2007. Suppression of innate immunity by a nasal carriage strain of *Staphylococcus aureus* increases its colonization on nasal epithelium. Immunology **122:**80-9.

154. **Quinn, G. A., P. M. Tarwater, and A. M. Cole.** 2009. Subversion of interleukin-1-mediated host defence by a nasal carrier strain of *Staphylococcus aureus*. Immunology **128:**e222-9.

155. **Rahman, M., S. Connolly, W. C. Noble, B. Cookson, and I. Phillips.** 1990. Diversity of staphylococci exhibiting high-level resistance to mupirocin. J Med Microbiol **33:**97-100.

156. **Rahman, M., W. C. Noble, and K. G. Dyke.** 1993. Probes for the study of mupirocin resistance in staphylococci. J Med Microbiol **39:**446-9.

157. **Rambaut, A., and A. J. Drummond.** 2007. Tracer v1.5, Available from: <http://beast.bio.ed.ac.uk/Tracer>.

158. **Riesen, A., and V. Perreten.** 2010. *Staphylococcus rostri* sp. nov., a haemolytic bacterium isolated from the noses of healthy pigs. Int J Syst Evol Microbiol **60:**2042-7.

159. **Roben, P. W., A. N. Salem, and G. J. Silverman.** 1995. VH3 family antibodies bind domain D of staphylococcal protein A. J Immunol **154:**6437-45.

160. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572-4.

161. **Said, K. B., J. Ismail, J. Campbell, M. R. Mulvey, A. M. Bourgault, S. Messier, and X. Zhao.** 2010. Regional profiling for determination of genotype diversity of mastitis-specific *Staphylococcus aureus* lineage in Canada by use of clumping factor A, pulsed-field gel electrophoresis, and spa typing. J Clin Microbiol **48:**375-86.

162. **Said, K. B., K. Ramotar, G. Zhu, and X. Zhao.** 2009. Repeat-based subtyping and grouping of *Staphylococcus aureus* from human infections and bovine mastitis using the R-domain of the clumping factor A gene. Diagn Microbiol Infect Dis **63:**24-37.

163. **Said, K. B., G. Zhu, and X. Zhao.** 2010. Organ- and host-specific clonal groups of *Staphylococcus aureus* from human infections and bovine mastitis revealed by the clumping factor A gene. Foodborne Pathog Dis **7:**111-9.

164. **Sakoulas, G., and R. C. Moellering, Jr.** 2008. Increasing antibiotic resistance among methicillin-resistant *Staphylococcus aureus* strains. Clin Infect Dis **46 Suppl 5:**S360-7.

165. **Sakwinska, O., M. Giddey, M. Moreillon, D. Morisset, A. Waldvogel, and P. Moreillon.** 2011. Host range and human-bovine host shift in *Staphylococcus aureus*. Appl Environ Microbiol.

166. **Sakwinska, O., G. Kuhn, C. Balmelli, P. Francioli, M. Giddey, V. Perreten, A. Riesen, F. Zysset, D. S. Blanc, and P. Moreillon.** 2009. Genetic diversity and ecological

success of *Staphylococcus aureus* strains colonizing humans. Appl Environ Microbiol **75:**175-83.

167. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:**5463-7.

168. **Schaffer, A. C., R. M. Solinga, J. Cocchiaro, M. Portoles, K. B. Kiser, A. Risley, S. M. Randall, V. Valtulina, P. Speziale, E. Walsh, T. Foster, and J. C. Lee.** 2006. Immunization with *Staphylococcus aureus* clumping factor B, a major determinant in nasal carriage, reduces nasal colonization in a murine model. Infect Immun **74:**2145-53.

169. **Shah, M. M., H. Iihara, M. Noda, S. X. Song, P. H. Nhung, K. Ohkusu, Y. Kawamura, and T. Ezaki.** 2007. *dnaJ* gene sequence-based assay for species identification and phylogenetic grouping in the genus *Staphylococcus*. Int J Syst Evol Microbiol **57:**25-30.

170. **Shopsin, B., M. Gomez, S. O. Montgomery, D. H. Smith, M. Waddington, D. E. Dodge, D. A. Bost, M. Riehman, S. Naidich, and B. N. Kreiswirth.** 1999. Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J Clin Microbiol **37:**3556-63.

171. **Shopsin, B., M. Gomez, M. Waddington, M. Riehman, and B. N. Kreiswirth.** 2000. Use of coagulase gene (*coa*) repeat region nucleotide sequences for typing of methicillin-resistant *Staphylococcus aureus* strains. J Clin Microbiol **38:**3453-6.

172. **Shukla, S. K., M. E. Karow, J. M. Brady, M. E. Stemper, J. Kislow, N. Moore, K. Wroblewski, P. H. Chyou, D. M. Warshauer, K. D. Reed, R. Lynfield, and W. R. Schwan.** 2010. Virulence genes and genotypic associations in nasal carriage, community-associated methicillin-susceptible and methicillin-resistant USA400 *Staphylococcus aureus* isolates. J Clin Microbiol **48:**3582-92.

173. **Sieradzki, K., and A. Tomasz.** 2003. Alterations of cell wall structure and metabolism accompany reduced susceptibility to vancomycin in an isogenic series of clinical isolates of *Staphylococcus aureus*. J Bacteriol **185:**7103-10.

174. **Silverman, G. J., C. S. Goodyear, and D. L. Siegel.** 2005. On the mechanism of staphylococcal protein A immunomodulation. Transfusion **45:**274-80.

175. **Sinha, B., P. P. Francois, O. Nusse, M. Foti, O. M. Hartford, P. Vaudaux, T. J. Foster, D. P. Lew, M. Herrmann, and K. H. Krause.** 1999. Fibronectin-binding protein acts as *Staphylococcus aureus* invasin via fibronectin bridging to integrin alpha5beta1. Cell Microbiol **1:**101-17.

176. **Smith, E. J., L. Visai, S. W. Kerrigan, P. Speziale, and T. J. Foster.** 2011. The Sbi protein: a multifunctional immune evasion factor of *Staphylococcus aureus*. Infect Immun.

177. **Stocsits, R. R., H. Letsch, J. Hertel, B. Misof, and P. F. Stadler.** 2009. Accurate and efficient reconstruction of deep phylogenies from structured RNAs. Nucleic Acids Res **37:**6184-93.

178. **Sukumaran, J., and M. T. Holder.** 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics **26:**1569-71.

179. **Supre, K., S. De Vliegher, I. Cleenwerck, K. Engelbeen, S. Van Trappen, S. Piepers, O. C. Sampimon, R. N. Zadoks, P. De Vos, and F. Haesebrouck.** 2010. *Staphylococcus devriesei* sp. nov., isolated from teat apices and milk of dairy cows. Int J Syst Evol Microbiol **60:**2739-44.

180. **Sutherland, R., R. J. Boon, K. E. Griffin, P. J. Masters, B. Slocombe, and A. R. White.** 1985. Antibacterial activity of mupirocin (pseudomonic acid), a new antibiotic for topical use. Antimicrob Agents Chemother **27:**495-8.

181. **Svec, P., M. Vancanneyt, I. Sedlacek, K. Engelbeen, V. Stetina, J. Swings, and P. Petras.** 2004. Reclassification of *Staphylococcus pulvereri* Zakrzewska-Czerwinska *et al*. 1995 as a later synonym of *Staphylococcus vitulinus* Webster *et al.* 1994. Int J Syst Evol Microbiol **54:**2213-5.

182. **Takahashi, T., I. Satoh, and N. Kikuchi.** 1999. Phylogenetic relationships of 38 taxa of the genus *Staphylococcus* based on 16S rRNA gene sequence analysis. Int J Syst Bacteriol **49 Pt 2:**725-8.

183. **Takeuchi, O., K. Hoshino, T. Kawai, H. Sanjo, H. Takada, T. Ogawa, K. Takeda, and S. Akira.** 1999. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity **11:**443-51.

184. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24:**1596-9.

185. **Tang, Y. Q., J. Yuan, G. Osapay, K. Osapay, D. Tran, C. J. Miller, A. J. Ouellette, and M. E. Selsted.** 1999. A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. Science **286:**498-502.

186. **Taponen, S., H. Simojoki, M. Haveri, H. D. Larsen, and S. Pyorala.** 2006. Clinical characteristics and persistence of bovine mastitis caused by different species of coagulase-negative staphylococci identified with API or AFLP. Vet Microbiol **115:**199-207.

187. **Taponen, S., K. Supre, V. Piessens, E. Van Coillie, S. De Vliegher, and J. M. Koort.** 2011. *Staphylococcus agnetis* sp. nov., a coagulase-variable species from bovine subclinical and mild clinical mastitis. Int J Syst Evol Microbiol.

188. **Thomas, C. M., J. Hothersall, C. L. Willis, and T. J. Simpson.** 2010. Resistance to and synthesis of the antibiotic mupirocin. Nat Rev Microbiol **8:**281-9.

189. **Thomas, R., and T. Brooks.** 2006. Attachment of *Yersinia pestis* to human respiratory cell lines is inhibited by certain oligosaccharides. J Med Microbiol **55:**309-15.

190. **Thomas, R. J.** 2010. Receptor mimicry as novel therapeutic treatment for biothreat agents. Bioeng Bugs **1:**17-30.

191. **Trabi, M., H. J. Schirra, and D. J. Craik.** 2001. Three-dimensional structure of RTD-1, a cyclic antimicrobial defensin from Rhesus macaque leukocytes. Biochemistry **40:**4211-21.

192. **Tran, D., P. A. Tran, Y. Q. Tang, J. Yuan, T. Cole, and M. E. Selsted.** 2002. Homodimeric theta-defensins from rhesus macaque leukocytes: isolation, synthesis, antimicrobial activities, and bacterial binding properties of the cyclic peptides. J Biol Chem **277:**3079-84.

193. **Trulzsch, K., B. Grabein, P. Schumann, A. Mellmann, U. Antonenka, J. Heesemann, and K. Becker.** 2007. *Staphylococcus pettenkoferi* sp. nov., a novel coagulase-negative staphylococcal species isolated from human clinical specimens. Int J Syst Evol Microbiol **57:**1543-8.

194. **Turkheimer, F. E., R. Hinz, and V. J. Cunningham.** 2003. On the undecidability among kinetic models: from model selection to model averaging. J Cereb Blood Flow Metab **23:**490-8.

195. **Uhlen, M., B. Guss, B. Nilsson, S. Gatenbeck, L. Philipson, and M. Lindberg.** 1984. Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. J Biol Chem **259:**1695-702.

196. **Van Belkum, A., N. H. Riewarts Eriksen, M. Sijmons, W. Van Leeuwen, M. Van den Bergh, J. Kluytmans, F. Espersen, and H. Verbrugh.** 1997. Coagulase and protein A polymorphisms do not contribute to persistence of nasal colonisation by *Staphylococcus aureus*. J Med Microbiol **46:**222-32.

197. **van Belkum, A., N. J. Verkaik, C. P. de Vogel, H. A. Boelens, J. Verveer, J. L. Nouwen, H. A. Verbrugh, and H. F. Wertheim.** 2009. Reclassification of *Staphylococcus aureus* Nasal Carriage Types. J Infect Dis **199:**1820-6.

198. **Vandenesch, F., T. Naimi, M. C. Enright, G. Lina, G. R. Nimmo, H. Heffernan, N. Liassine, M. Bes, T. Greenland, M. E. Reverdy, and J. Etienne.** 2003. Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerg Infect Dis **9:**978-84.

199. **Venkataraman, N., A. L. Cole, P. Svoboda, J. Pohl, and A. M. Cole.** 2005. Cationic polypeptides are required for anti-HIV-1 activity of human vaginal fluid. J Immunol **175:**7560-7.

200. **von Eiff, C., K. Becker, K. Machka, H. Stammer, and G. Peters.** 2001. Nasal carriage as a source of *Staphylococcus aureus* bacteremia. Study Group. N Engl J Med **344:**11-6.

201. **Walsh, T. R., and R. A. Howe.** 2002. The prevalence and mechanisms of vancomycin resistance in *Staphylococcus aureus*. Annu Rev Microbiol **56:**657-75.

202. **Wenzel, R. P., and T. M. Perl.** 1995. The significance of nasal carriage of *Staphylococcus aureus* and the incidence of postoperative wound infection. J Hosp Infect **31:**13-24.

203. **Wertheim, H. F., D. C. Melles, M. C. Vos, W. van Leeuwen, A. van Belkum, H. A. Verbrugh, and J. L. Nouwen.** 2005. The role of nasal carriage in *Staphylococcus aureus* infections. Lancet Infect Dis **5:**751-62.

204. **Wertheim, H. F., E. Walsh, R. Choudhurry, D. C. Melles, H. A. Boelens, H. Miajlovic, H. A. Verbrugh, T. Foster, and A. van Belkum.** 2008. Key role for clumping factor B in *Staphylococcus aureus* nasal colonization of humans. PLoS Med **5:**e17.

205. **Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis.** 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol Phylogenet Evol **25:**361-71.

206. **Williams, R. J., B. Henderson, and S. P. Nair.** 2002. *Staphylococcus aureus* fibronectin binding proteins A and B possess a second fibronectin binding region that may have biological relevance to bone tissues. Calcif Tissue Int **70:**416-21.

207. **Woodford, N., A. P. Watson, S. Patel, M. Jevon, D. J. Waghorn, and B. D. Cookson.** 1998. Heterogeneous location of the *mupA* high-level mupirocin resistance gene in *Staphylococcus aureus*. J Med Microbiol **47:**829-35.

208. **Yasin, B., W. Wang, M. Pang, N. Cheshenko, T. Hong, A. J. Waring, B. C. Herold, E. A. Wagar, and R. I. Lehrer.** 2004. Theta defensins protect cells from infection by herpes simplex virus by inhibiting viral adhesion and entry. J Virol **78:**5147-56.

209. **Zhang, L., K. Jacobsson, J. Vasi, M. Lindberg, and L. Frykberg.** 1998. A second IgG-binding protein in *Staphylococcus aureus*. Microbiology **144 (Pt 4):**985-91.

210. **Zharkikh, A.** 1994. Estimation of evolutionary distances between nucleotide sequences. J Mol Evol **39:**315-29.

211. **Zwickl, D. J.** 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation. The University of Texas at Austin, Austin.