

ELUCIDATING THE EVOLUTIONARY CHANGES THAT ENABLE
STAPHYLOCOCCUS AUREUS TO OVERCOME THE SYNERGISTIC IMPACT OF
HOST-IMPOSED MANGANESE STARVATION AND OXIDATIVE STRESS

BY

KATIE FRYE

DISSERTATION

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Microbiology
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Associate Professor Thomas Kehl-Fie, Chair
Professor James Imlay
Professor Andrei Kuzminov
Professor Rachel Whitaker

ABSTRACT

Metals are essential to all forms of life, with 30% of all proteins being reliant on a metal for full functionality. The metal used by metalloproteins can change over time in response to various pressures, presumably including that of metal availability. In line with this, the bioavailability of metals changes both across geological time as well as among macro- and micro-environments. One example of this change in metal availability is seen ~2.5 billion years ago during the Great Oxygenation Event. Among other metals, this oxygenation led to a four orders of magnitude decline in the availability of iron (Fe). This decline placed significant pressure on organisms with metalloproteins relying on this metal to switch to a more available metal. Another example of this change in metal availability occurs during infection by a pathogen where the host sequesters metals in a defense termed nutritional immunity. By limiting these essential metals, the host creates a pressure on pathogens to evolve strategies to circumvent this limitation such as metal-independent enzymes or enzymes relying on alternative metals. Together, these observations lead to the hypothesis that metal availability drove metalloenzyme neofunctionalization. Potentially, this neofunctionalization occurred through iterative rounds of single mutations, each providing a slightly more beneficial enzyme, enabling the shift in metal specificity. However, this hypothesis has not been evaluated as there was not a system that provided evolutionary starting and ending points where biologically relevant stressors, like metal limitation, could be imposed.

Recently, it has been revealed that the superoxide dismutases (SODs) in the bacterial pathogen *Staphylococcus aureus* could be a useful tool in the evaluation of the pressures on metalloenzyme neofunctionalization. *S. aureus* possesses two superoxide dismutases (SODs), one manganese-dependent (SodA) and one capable of functioning with either Mn or Fe (SodM). SODs catalyze the conversion of superoxide to hydrogen peroxide and oxygen, helping to protect the cell from toxic reactive oxygen species. The cambialistic (either of two metals) nature of SodM allows the cell to retain SOD activity even when Mn-starved, promoting resistance to nutritional immunity. SodA is present in all staphylococci, while SodM is only present in *S. aureus* and two closely related species. SodA and SodM are 75% identical at the amino acid level, and they coordinate their metal ions with the same residues. Further analysis suggests that SodM arose following a duplication and subsequent neofunctionalization of *sodA*. These analyses also proposed that the ancestor of SodM was strictly Mn-dependent, not cambialistic. This makes the staphylococcal SODs a unique system where there is a predicted evolutionary starting point and a present-day metalloenzyme, allowing for studies of the evolutionary path from one enzyme to the other.

Previous work identified three residues in the staphylococcal SODs that, when swapped, enable SodA to use Fe as a cofactor and increase the activity of SodM with Mn. In the studies presented herein, I utilized the system provided by the staphylococcal SODs to investigate how metal availability might drive metalloprotein neofunctionalization. By creating a system that mimics the

predicted evolutionary ancestor of *S. aureus* with a duplicated SodA, the second copy of SodA was manipulated with swapping mutations at the residues previously identified to dictate the Fe or Mn-dependent activity of the SODs. These strains then could act as potential evolutionary intermediates, and the effect of physiologically relevant stresses, such as metal limitation and oxidative stress, was evaluated. Preliminary results suggest that swapping one residue in SodA for the one present in SodM leads to an increased selectivity of SodA for Fe over Mn, indicating that non-coordinating residues within the protein might be involved in metal selectivity of the enzyme. These studies also revealed that the same single residue swap in SodA also led to an increased ability for *S. aureus* to resist oxidative stress and Mn-limitation. Additionally, *S. aureus* strains encoding SodA variants with this single residue swap had an increased bacterial burden during infection as compared to a *S. aureus* strain encoding two copies of wild type SodA (*sodA*²). Finally, this work also suggests that there is a disconnect between predictions based on *in vitro* SOD activity and performance of the SodA variant in *S. aureus*. This disconnect might be driven by a reduction of SOD activity of these SodA variants when they are within *S. aureus*. These results demonstrate that biochemically, the path from Mn-dependency to cambialism is simple, but biologically that path is much more constrained.

Note to the reader

When deciding to stay at the University of Illinois Urbana-Champaign after the departure of my first advisor, Dr. Patrick Degnan, it was upon the mutual

agreement of my committee members and the head of the department at the time, Dr. John Cronan, that my contributions to science from both labs would be included in my thesis. Accordingly, Dr. Cronan affectionately refers to my thesis as “Meanderings in Microbiology”. My introduction, Chapter 2, and conclusion will be based on my work performed in Dr. Kehl-Fie’s lab, and Chapter 3 will be the published work from my time in Dr. Degnan’s lab.

TABLE OF CONTENTS

CHAPTER 1: Introduction	1
1.1 Metalloproteins and changes in metal availability.....	1
1.2 Oxidative stress: a universal stress to organisms that encounter oxygen	4
1.3 Superoxide dismutase metal specificity	6
1.4 Eukaryotic superoxide dismutase utilization	10
1.5 Many bacteria require superoxide dismutase to be effective pathogens	12
1.6 Understanding the diversity of superoxide dismutases	16
1.7 Superoxide dismutase diversity and metal availability.....	19
1.8 Metalloprotein metal loading	21
1.9 Staphylococcal SodA and SodM as foundations for studying pressures of metalloenzyme neofunctionalization	23
1.10 <i>Staphylococcus aureus</i> is a serious threat to human health.....	25
1.11 Importance, restriction, and acquisition of transition metals Fe, Mn, and Zn in biology	28
1.12 Aim of this study	37
1.13 Figures.....	39
1.14 References	41

CHAPTER 2: Metal availability can drive metalloenzyme

neofunctionalization, but the path to altered metal dependency is not free rein	68
2.1 Abstract.....	68
2.2 Introduction.....	70
2.3 Results.....	76
2.3.1 Incremental changes in SOD metal activity provide a benefit to <i>S. aureus</i>	76
2.3.2 Absence of a native copy of SodA alters whether SodA protein variants provide a benefit to <i>S. aureus</i>	79
2.3.3 SodA Gly159Leu and Phe19Ile-Gly159Leu swapping mutations lead to increased Fe selectivity.....	82
2.3.4 Leu160Phe mutations likely lead to SodA protein variants that are unable to properly fold or associate with an appropriate metal.....	85
2.3.5 The path to altered metal dependency is restricted in the host.....	86
2.4 Discussion	88
2.5 Methods.....	92
2.5.1 Ethics statement.....	92
2.5.2 Bacterial strains and cloning.....	92
2.5.3 Calprotectin growth assays.....	93
2.5.4 SOD expression analysis.....	94

2.5.5 Evaluation of SOD activity	94
2.5.6 Quantification of SodA peptides containing AA160	95
2.5.7 Animal infections	98
2.6 Figures and Tables	99
2.7 References	123
CHAPTER 3: Mobilization of vitamin B ₁₂ transporters alter competitive	
dynamics in a human gut microbe.....	129
3.1 Summary	129
3.2 Introduction.....	130
3.3 Results.....	133
3.3.1 Distinct corrinoid transporters are encoded on three	
genetically divergent MGE families.....	133
3.3.2 Family 1 corrinoid transport-encoding CTns are	
mobilizable and transport vitamin B ₁₂	135
3.3.3 BoCTn and BtCTn have conserved conjugation and	
integration mechanisms.....	138
3.3.4 CTn-encoded corrinoid transport loci provide distinct	
competitive advantages.....	140
3.3.5 BoCTn and BtCTn mobilize in the mammalian	
Intestinal tract	143
3.4 Discussion	145
3.5 Acknowledgments	151
3.6 Methods.....	151

3.6.1 Resource availability	151
3.6.2 Experimental model and subject details.....	152
3.6.3 Method details	155
3.6.4 Quantification and statistical analysis.....	160
3.7 Figures and Tables	161
3.8 References	269
CHAPTER 4: Conclusions and future directions	276
4.1 Conclusions and future directions	276
4.2 References	281

CHAPTER 1¹:

Introduction

1.1 Metalloproteins and changes in metal availability

Metals are essential for all forms of life, with 30% of all proteins and 50% of all enzymes requiring a metal to function (133, 280, 281). These metalloproteins and enzymes are involved in a variety of important cellular processes including metabolism (e.g., enolase), ribosomes and RNA and DNA polymerases (e.g., structural zinc fingers), and oxidative stress defense (e.g., superoxide dismutase) (66, 88, 192, 207). Interestingly, the identity of the metal utilized by a metalloprotein can change across evolutionary time, but the pressures influencing this shift are not well understood (7, 80, 215, 261).

One pressure that has been proposed to drive these changes in metalloprotein metal utilization is that of metal availability (7, 17, 80, 215, 261). In early life, iron (Fe) was relatively bioavailable and early organisms are thought to have been Fe-philic (4, 7, 80, 160, 245). This assessment is supported by the observation that folds in proteins binding Fe appear to be much older than those binding copper (Cu) and zinc (Zn) (80, 147). This protein fold distinction between Fe binding proteins and Cu and Zn binding proteins likely coincides with the Great Oxygenation Event ~2.5 billion years ago, when concentrations of

¹Portions of this chapter have been published in *Journal of Inorganic Biochemistry*. Modifications have been made to the text to make it appropriate for a thesis introduction chapter and to meet formatting requirements for the dissertation. Copyright © 2022 Elsevier Inc. Citation: Frye KA, Sendra KM, Waldron KJ, Kehl-Fie TE. 2022. Old dogs, new tricks: New insights into the iron/manganese superoxide dismutase family. *Journal of Inorganic Biochemistry*. 230: 111748.

atmospheric dioxygen (O₂) rose (23, 162, 182). During this time, the levels of atmospheric O₂ rose from 0.0001% of the present atmospheric levels (PAL) to 1-10% PAL and then again ~0.5 billion years ago to PAL (4, 160). In line with this rise in O₂, levels of sulfur also rose by ~2 orders of magnitude (160). Likely this is because in the early, low O₂ world, sulfur would have been trapped in the form of sulfide minerals which would have been released when O₂ increased and began oxidizing the sulfur to SO₄²⁻ (4). This also might explain why the levels of Cu and Zn rose by ~14 and ~9 orders of magnitude, respectively, as they would have been precipitated as sulfide minerals before the rise in O₂ (4, 80, 160, 162, 245). On the other hand, Fe underwent a four orders of magnitude decrease in availability (4, 160). These changes in the availability of metals would have placed significant pressure on organisms to switch from metalloproteins relying on metals like Fe to those relying on metals like Cu or Zn (80, 147, 162, 202). In fact, an example of this shift can be seen in a eukaryotic diatom which switched from relying on an Fe-containing cytochrome to a Cu-containing plastocyanin, presumably in response to the dwindling amount of bioavailable Fe in the oceans (215, 261).

While the cytochrome-to-plastocyanin switch is an example of an organism altering the metal utilized in response to changes in metal availability within their environment, it was accomplished through emergence of a new metalloprotein domain. Instead of evolving a new protein domain, the protein domain itself in the Fe/Mn-superoxide dismutase (SOD) protein family evolved to utilize a different metal cofactor. Prior to the Great Oxygenation Event, there

would have been only slight selection pressure for Fe over Mn in this protein family as the bioavailability of these metals would have been similar. These conditions, in addition to phylogenetic data, has led to a prediction that the ancestral Fe/Mn-SOD may have utilized either Fe or Mn, making it cambialistic (162). The subsequent decrease in the bioavailability of Fe would have put significant pressure on organisms relying on this metal to adopt non-Fe utilizing proteins, selecting for organisms with enzymes best adapted to those new conditions (162). Most contemporary aerobes possess a Mn-SOD, while anaerobes still favor Fe-SOD.

Metal levels can differ between habitats, and consequently SOD metal utilization apparently has occasionally shifted even within evolutionary clades. One example of this can be seen in the eukaryotic organism *Tetrahymena pyriformis*. While *T. pyriformis* has an Fe-SOD, the enzyme sequence is more similar to that of eukaryotic Mn-SODs, suggesting that this SOD underwent a substitution in metal cofactor (15, 41). Another study revealed that this cofactor switching might be found throughout the Fe/Mn-SOD family. By utilizing an expanded phylogenetic approach, three distinct subtypes of Fe/Mn-SODs were identified based on the presence or absence of highly conserved GGH or AAQ motifs in the SOD protein sequence (17). The GGH subtype contains mainly Mn-SODs, the AAQ subtype contains mainly Fe-SODs, and the last group is diverse and lacks both motifs (17). However, there was a SOD with a GGH motif that

was Fe-dependent instead of Mn-dependent, and cambialistic SODs have been found in both the GGH and AAQ subtypes. These findings indicate that metal specificities in this protein family interchange evolutionarily (17).

Such changes in metal cofactor utilization, an example of enzyme neofunctionalization, are thought to occur through iterative rounds of mutation and selection, with each cycle resulting in a more beneficial enzyme. A central question remains: how easy is it for metalloenzymes to evolve new metal preferences? Is this something that can occur via a cycle of single mutations, or does it require multiple mutations per cycle? If multiple, simultaneous mutations are required, this would be a much more difficult evolutionary event leaving some organisms unable to attain altered metal cofactor utilization to adapt to changes in their environment or to expand to new habitats with different metal availabilities. While metal availability has been predicted to be a key driving force in this metalloenzyme evolution, this process has not been experimentally reproduced. The work presented in this thesis begins to address this gap by utilizing a pair of SODs that rely on the transition metals Fe and Mn in a pathogenic bacterium, *Staphylococcus aureus*.

1.2 Oxidative stress: a universal stress to organisms that encounter oxygen

As primordial environmental oxygen levels increased and life began to utilize the newly available oxygen, its potential to be converted into toxic byproducts was unveiled. Simply the act of performing aerobic metabolism

exposes organisms to internally generated reactive oxygen species (ROS), including superoxide ($O_2^{\bullet-}$) (189, 195). Superoxide is toxic to the cell, targeting important components of the cell including iron-sulfur cluster containing proteins, mononuclear Fe^{2+} enzymes, and short-chain sugars (28, 137, 138). While intracellular ROS generation is accidental, the immune system actively harnesses the toxic properties of ROS to attack microbial invaders. Immune cells including neutrophils and macrophages generate superoxide as part of the oxidative burst to kill microbes that breach the epithelial barrier (42, 44, 90, 237). Consequently, microbial pathogens need to adapt and evolve mechanisms to bypass the host's oxidative immune response. The importance of the oxidative burst and the threat that ROS pose to potential invaders is revealed by chronic granulomatous disease (CGD). Individuals suffering from CGD have defects in the ability to generate an oxidative burst, preventing immune cells from clearing invaders and thus having enhanced susceptibility to infection (45, 228, 241).

Whether confronted by endogenously or exogenously produced superoxide, both host and pathogen must protect themselves. SODs are the primary means by which most organisms do so (121, 237). These redox-active metalloenzymes are present in almost all forms of life, including humans, plants, pathogenic and non-pathogenic microbes, and even strict anaerobes (37, 180). Given the wide distribution and impact on human health of SODs, since their discovery in 1969 substantial insight has been gained into their function and

distribution (192). In particular, recent work in the context of microbial pathogenesis has brought new insights but also revealed that there is still much to learn (101, 198, 238).

1.3 Superoxide dismutase metal specificity

SODs catalyze the conversion of superoxide ($O_2^{\bullet-}$) into oxygen (O_2) and hydrogen peroxide (H_2O_2). Peroxide can then be converted safely to water (H_2O) by other enzymes (28, 254). The detailed molecular mechanism of SOD catalytic activity, and key questions that remain to be resolved about how they function, have been previously reviewed (254). Given the reaction they catalyze, it is unsurprising that nature has exploited redox active metal cofactors to perform this chemistry. What is surprising though is that three distinct catalytic centers utilizing different metal cofactors have evolved independently via functional convergence. Three structurally distinct SOD protein families exist, each that has activity with a specific set of metals. As a result, SODs are frequently discussed based on the metal that they have activity with, though enzymes active with different metals can belong to the same protein family.

The SOD protein families include, those that are active with nickel (Ni) (Ni-SODs); SODs that are active with Cu or Cu and Zn (Cu/Zn-SODs); and SODs that are active with either Fe or Mn (Fe/Mn-SODs) (Figure 1.1) (254). Ni-SODs were the most recently discovered and are the SODs about which least is known (294). Ni-SOD distribution is more restricted than that of Cu/Zn- or Fe/Mn-SODs, being present primarily in marine cyanobacteria and actinomycetes (81, 262). All

SODs rely on a common chemical process: the sequential oxidation and reduction of superoxide molecules in a 'ping pong' mechanism, whereby the redox-active metal cofactor undergoes sequential reduction and oxidation steps (254). Given the disparate inherent reduction potentials of these catalytic metal ions (E^0 values ranging from 2.26 V for the $\text{Ni}^{2+}/\text{Ni}^{3+}$ redox couple to 0.15 V for the $\text{Cu}^{2+}/\text{Cu}^+$ couple), their different protein architectures must precisely and differentially poise the reduction potential of their respective metal ion to optimize SOD dismutation (optimal $E^0 \sim 0.35$ V) (254). Indeed, this control of the metal's potential by the protein is particularly important for Ni-SODs because the Ni redox couple lies well outside of the range that would enable it to reduce the superoxide substrate (127). The mechanism by which the reduction potential of the catalytic metal is optimized to facilitate dismutation remains enigmatic and is an area of active study.

SODs are usually cofactor specific, catalyzing superoxide dismutation using only their target metal cofactor(s) (Figure 1.1). For example, Ni-SODs and Cu/Zn-SODs are absolutely dependent on their catalytic metal ions, Ni and Cu respectively, and are inactive when loaded with other metal ions *in vitro* (199, 294). Differing from the other SODs, the Fe/Mn-SOD family uniquely evolved the ability to use either metal within a single protein framework (17, 192, 197, 277). Nonetheless, individual SODs from the Fe/Mn family were long regarded as being active exclusively with either Mn or Fe but not with both metals. This assumption was based on analyses of the cofactor specificity of a small number of early model systems, most notably the highly metal-specific Mn-SOD and Fe-

SOD of *Escherichia coli* (211). Similar observations were also made with the mitochondrial Fe/Mn SOD (292). The discovery of Fe/Mn-SODs that are in fact cambialistic, i.e., active with either Mn or Fe, challenged this dogma of singular metal dependency (186). Numerous cambialistic SODs have since been discovered in diverse organisms, and recent studies have shown that cambialism can be physiologically important in bacterial pathogens (17, 101, 181, 186, 205, 206, 293). It is now recognized that the Fe/Mn-SOD family includes Fe-SODs, Mn-SODs, and cam-SODs (2, 86, 101, 111, 186, 193). Furthermore, the SOD enzymes characterized thus far show that even these discrete definitions are too rigid, with metal specificity instead lying on a continuum from Mn-specific to Fe-specific, with varying levels of cambialism in between these extremes (Figure 1.1) (17). The simplistic initial model of narrow metal specificity among Fe/Mn-SODs has likely restrained advances in our understanding of these enzymes, complicating both mechanistic biochemical studies and bioinformatic efforts to predict SODs metal utilization from sequence.

The discovery of cambialistic members of the Fe/Mn-SOD family also raises critical questions about how distinct Mn-specific, Fe-specific, and cambialistic isozymes have evolved within a single metalloprotein family. This diversity of metal specificities is achieved even though all members of the Fe/Mn-SOD family are closely related in sequence, have highly similar structures, and employ nearly identical metal coordination environments of three histidines and one aspartate (His26, His81, Asp167 and His171, using the numbering system from *E. coli* Mn-SOD) (85, 254, 258). A “redox tuning” hypothesis has been

proposed to explain this phenomenon, in which key amino acid residues in the metal's secondary coordination sphere (i.e., residues located near the metal cofactor, but which make no direct contacts with it) adjust each metal's reduction potential to optimize catalysis (114, 278). *In vitro* data from numerous groups support this model. Biochemical and biophysical studies of numerous mutated forms of the *E. coli* SODs demonstrate that these second sphere residues do affect the metal cofactor's reduction potential and, therefore, the catalytic activity (254, 278). Furthermore, it was recently shown that reciprocal switching of just two secondary coordination sphere residues between a pair of SODs from *S. aureus* substantially inverts their metal specificity (17). These two mutations were sufficient to make its Mn-SOD variant highly cambialistic, and to make its cambialistic-SOD highly Mn-specific. Manipulation of redox potential via altering these second sphere residues is thus a relatively simple and plausible mechanism by which evolution can shape SOD metal specificity. Yet, there are few structural clues as to how the chemistry of these non-coordinating residues regulates the metal's reduction potential, due in part to the technical challenges involved in experimentally quantifying SOD reduction potentials with sufficient resolution to correlate redox changes with the structures and activities of diverse mutants.

Notably, the Fe/Mn-SOD family is not the only family of SODs whose metal usage has been informed by investigations studying host-pathogen interactions. The Cu/Zn SOD family had long been thought to absolutely require both Cu for catalysis and Zn for structural integrity and catalytic efficiency.

However, recent work in *Mycobacterium tuberculosis* and *Candida albicans* demonstrated that some SODs within this family do not require Zn (107, 256). Such discoveries urge work to be done to determine how widespread these Cu-only and Fe/Mn cambialistic SODs are across the domains of life (42, 238, 249). While a multitude of studies have been done on SODs, these evolutionarily ancient enzymes still have insights waiting to be discovered (101, 198, 238).

1.4 Eukaryotic superoxide dismutase utilization

While all SODs carry out the same chemical reaction, many organisms possess more than one. Mammals possess three SODs: SOD1, SOD2, and SOD3 (198). SOD1 and SOD3 are both members of the Cu/Zn SOD family, while SOD2 is from the Fe/Mn-SOD family. Many other animals possess a similar set of SODs, but their nomenclature can differ. These SODs localize predominantly to distinct cellular locations, with SOD1 mainly in the cytoplasm and the mitochondrial intermembrane space, SOD2 in the mitochondrial matrix, and SOD3 associated with the extracellular matrix/cell surface. Similar to animals, plants also possess more than one SOD, each compartmentalized in a distinct cellular location (71, 109, 123, 144, 251). As a result of this compartmentalization, these SODs have largely non-overlapping roles in the defense of the cell. Examples of these non-redundant roles are observed across multiple organisms. In *Drosophila melanogaster*, homozygous loss of either *sod1* or *sod2* is lethal (33, 161, 222, 223, 240, 282). Similarly, mice lacking the mitochondrial SOD2 die shortly after birth (171, 175). In humans, genetic

disorders have also revealed the non-overlapping roles of SODs (99, 196). Mutations in SOD1 are associated with familial amyotrophic lateral sclerosis (fALS), abnormalities in SOD2 are associated with mitochondrial stress and cancer, while mutations in SOD3 are associated with vascular dysfunction including hypertension, diabetes, and atherosclerosis (98, 131, 242, 254, 276, 282, 291).

SODs also play important roles during infection by numerous eukaryotic microbes. Four Fe/Mn-SODs were identified in the causative agent of African sleeping sickness, *Trypanosoma brucei*, and overexpression of one of these homologues, SodB, increased survival of the closely related *Trypanosoma cruzi* during macrophage infection (187, 288). Based on phylogenetic evidence, the four SODs of *T. brucei* were likely acquired via at least two independent lateral gene transfer events followed by gene duplications (77). Deletion of one of the SodA alleles from another eukaryotic pathogen, *Leishmania amazonensis*, resulted in its inability to replicate in macrophages and an attenuated ability to generate cutaneous lesions in mice (201).

Similar to protozoan pathogens, pathogenic fungi possess an expanded repertoire of SODs relative to their host organisms. For example, *C. albicans* and other pathogenic fungi possess an additional Mn-dependent SOD in the cytoplasm and an extracellular Cu-only SOD (107, 185, 295). Loss of the Cu-only SOD reduces the ability of *C. albicans* to cause infection. Interestingly, the Mn-dependent SOD is induced in response to host-imposed Cu limitation (26, 174, 185).

1.5 Many bacteria require superoxide dismutase to be effective pathogens

Compared with animals and plants, there is greater diversity in the repertoire of SODs possessed by bacteria. Most pathogenic bacteria possess one or more cytoplasmic SODs belonging to the Fe/Mn-SOD family (42, 180). These SODs are frequently referred to as SodA when Mn-dependent and SodB when Fe-dependent, with this nomenclature being derived from their discovery in *E. coli*. However, there is not always a reliable correlation between name and metal dependency, and such annotations should be interpreted with care. The existence of cambialistic SODs further complicates the issue. In addition to Fe/Mn-SODs, some bacteria possess Cu/Zn SODs, which are generally referred to as SodC (73, 249). Bacterial Cu/Zn SODs are strictly extra-cytoplasmic (249). In Gram-negative bacteria, Cu/Zn family members are found in the periplasm, while in Gram-positive bacteria they are frequently predicted to be attached to the cell surface via lipids (24, 73). For bacteria that encode multiple SODs, the diversity of number and overall repertoire varies significantly from genus to genus and can even differ between species of the same genus (Figure 1.2).

While many pathogens encode multiple SODs, there are numerous examples of bacteria that possess only a single enzyme. As would be expected, the single SOD is frequently important for virulence-associated phenotypes or infection (87, 206, 221, 227, 293). *Borrelia burgdorferi*, the causative agent of Lyme disease, encodes a single Mn-dependent SOD, SodA (1, 272). Loss of SodA leads to reduction of viability of *B. burgdorferi* after incubation with either

neutrophils or macrophages, as well as a complete loss of detectable bacteria in intradermally infected mice (87). Similarly, *Campylobacter* spp., which are commonly associated with bacterial enteritis in humans and are found in the gastrointestinal tract of 60-80% of poultry broiler flocks worldwide, encode a single SOD, also from the Fe/Mn-SOD family (128). Deletion of the Fe-dependent SodB in *Campylobacter coli* leads to decreased bacterial survival on substrates essential to environmental transmission of this pathogen, as well as decreased chick colonization (227). In *Campylobacter jejuni*, Fe-dependent SodB mutants have decreased survival in human embryonic intestinal cells, indicating a role in intracellular survival (221, 227). *Porphyromonas gingivalis* and *Streptococcus mutans*, both oral pathogens, possess a single cambialistic SOD that contributes to aerotolerance, and thus survival, within the oral cavity (181, 186, 205, 206). Additionally, *Streptococcus pneumoniae*, the most common causative agent of pneumonia, as well as important gut microbes, the Bacteroidetes, each encode a single, cambialistic SOD (86, 111, 216). Loss of SodA makes *S. pneumoniae* less invasive and less pathogenic, as fewer bacteria are recovered from the blood and lungs of infected animals (293). In pathogenic Bacteroidetes, it has been suggested that SOD is a virulence factor, as the ability to withstand temporary tissue aeration would enhance the pathogenicity of these anaerobic microbes (263).

When bacteria encode two SODs, the SODs can either be from different families or both enzymes can belong to the same protein family. *Legionella pneumophila*, the causative agent in Legionnaires' disease, and *Mycobacterium*

tuberculosis, which is responsible for tuberculosis, both encode one SOD from the Fe/Mn family and one from the Cu/Zn family. While the Cu/Zn SOD in *L. pneumophila* is not essential for viability, it is critical for stationary phase survival and therefore perhaps also for extended dormancy between growth in hosts (257). On the other hand, the Fe-dependent SodB in *L. pneumophila* is essential for viability, and presumptively for virulence (244). Similarly, early reports on SodA in *M. tuberculosis* concluded that it must be essential, as attempts to delete it proved unsuccessful (82). Reduced expression of SodA attenuated *M. tuberculosis*, with ~100,000-fold fewer bacteria observed in the lungs and spleens of mice after infection (84). *M. tuberculosis* SodC is important for survival of the oxidative burst during infection, as *sodC* mutants are more susceptible to killing by activated macrophages (224).

As with eukaryotes, it is easy to hypothesize that the unique contribution of multiple SODs in *L. pneumophila*, *M. tuberculosis*, and other bacteria is derived from compartmentalization. However, this idea cannot explain why bacteria have multiple SODs within a single compartment. Varying metal specificities tend to be important in these cases. For example, *S. aureus* encodes two SODs, SodA and SodM, both of which belong to the Fe/Mn-SOD family and are present in the cytoplasm. SodA is strictly Mn-dependent, and SodM is cambialistic, displaying equal activity with Mn or Fe. Loss of SodA, SodM, or both SODs simultaneously results in reduced bacterial loads in a skin abscess model of infection (154, 156). In a systemic model of infection, loss of SodM, but not of

SodA, results in reduced bacterial burdens in liver tissues (101). Thus, despite localizing to the same cellular compartment, the staphylococcal SODs have unique non-overlapping roles during infection.

The diversity of SODs and the unique contribution each SOD provides are further exemplified by the Enterobacteriaceae and Bacilli. The Enterobacteriaceae, including *Escherichia coli* and *Shigella flexneri*, have a common core set of SODs, including one periplasmic Cu/Zn SOD, one cytoplasmic Mn-SOD, and one cytoplasmic Fe-SOD. For avian pathogenic *E. coli* O2 strain E058, SodA is important to virulence, as mutants lacking SodA exhibit significantly increased ingestion and decreased survival in chicken macrophages (100). In addition, mutants lacking SodA exhibit a significant decrease in bacterial burdens in both single-strain infections and competition experiments using chickens (100). In *S. flexneri*, SodB mutants are more susceptible to killing by both mouse macrophages and human polymorphonuclear leukocytes (97). Among Bacilli, both *Bacillus cereus* and *Bacillus anthracis* are predicted to encode four SODs, three that are from the Mn/Fe family (SodA1, SodA2, and SodS/Sod15) and one that is from the Cu/Zn family (SodC), with each of these enzymes differentially contributing to the organisms' success (68, 283, 296). In *B. cereus* 0-9, SodS is Fe-dependent and SodA1 and SodA2 are predicted to be Mn-dependent (296). However, given sequence similarity to *B. anthracis* SODs, SodA2 could in fact be Fe-dependent (273). Each of the four SODs in *B. cereus* 0-9 affect biofilm formation. However, the most significant reduction in the ability to form a biofilm is observed in a single *sodA2* mutant, demonstrating the

importance of SodA2 (296). Interestingly, deletion of Mn-dependent SodA1 leads to a delay in spore formation, while deletion of Fe-dependent SodS leads to a complete loss of swarming abilities (296). In *B. anthracis*, SodA1 and Sod15 are present in the exosporium (68, 177, 259). Assuming similarity, these SODs might also be present in the exosporium of *B. cereus*, perhaps influencing these phenotypes. Finally, individual SodS and SodC mutants have drastically reduced lethality, even lower than that of a SOD-null strain, against the cotton bollworm host, *Helicoverpa armigera*, indicating the importance of these two SODs for pathogenicity in this host (296). The fact that all four of these SODs are present in both *Bacillus* spp. indicates that having a diverse repertoire of SODs is particularly important for these organisms and increases their adaption and ability to handle ROS in diverse environments.

1.6 Understanding the diversity of superoxide dismutases

While the need to protect the cytoplasm and periplasm from ROS damage partially explains diversity associated with bacterial SODs, it does not explain all of it, as there are many examples of bacteria expressing expanded repertoires of cytosolic and periplasmic SODs (42, 97, 275). Again, we can glean insight into the roles of multiple SODs from the study of bacterial pathogens. When examining pathogenic and non-pathogenic bacteria from the same or related species, a common theme emerges, namely that the more pathogenic strains/lineages/serovars in these genres tend to encode an extra SOD. This can be seen in the Bacilli, Staphylococci and Enterobacteriaceae. In Bacilli, the non-

pathogenic *Bacillus subtilis* encodes only one SodA, while both *B. anthracis* and *B. cereus* encode two SodAs, SodA1 and SodA2 (47, 68, 141, 296). Amongst the Staphylococci, all members encode SodA, while only the more pathogenic *S. aureus* lineage encode the cambialistic *sodM* (290). All Salmonellae encode *sodC2*, while only highly virulent strains in serovars such as Typhimurium, Enteritidis, Dublin, Choleraesuis, and Heidelberg possess the phage-encoded *sodC1* (3, 89, 91). This acquisition of a phage encoded SodC homolog is also seen in pathogenic *E. coli*, including the highly virulent O157:H7 serotype (75, 124, 219).

Interestingly, some organisms appear to have potentially redundant protection systems as they have multiple SODs with the same metal dependency. In addition to the three core enterobacterial SODs, some *Salmonella enterica* serovars express four SODs, having gained an additional Cu/Zn SOD. Due to the order of discovery, the common Cu/Zn SOD is referred to as SodC2, while the additional enzyme is SodC1. While many studies have demonstrated the broad importance of SodC in this genus, it was less clear if encoding multiple SodC would provide unique superoxide protection derived from each enzyme. At neutral pH, superoxide cannot cross biological membranes; therefore, any superoxide formed outside the bacteria will stay outside the cell/in the periplasm (in the case of gram-negative organisms). At acidic pH, such as what is encountered within a phagolysosome, superoxide becomes protonated and uncharged ($\text{HO}_2\cdot$), and that form can cross biological membranes (165). Therefore, during infection, superoxide does have the potential to enter the

cytoplasm of the cell. The SodC1 and SodC2 enzymes are both periplasmic SODs from the Cu/Zn family and were originally thought to provide overlapping protection from exogenous superoxide stress; however, now it is appreciated that they have non-overlapping, non-redundant roles. SodC2 mutants in *S. enterica* serovar Choleraesuis have an intracellular survival defect relative to both wild type and SodC1 mutants in macrophages and human colon carcinoma cells (3). Additionally, during *in vivo* competition experiments, double (*sodC1/sodC2*) mutants are not further attenuated from single (*sodC1* or *sodC2*) mutants, indicating that these SODs provide complementary roles (3, 247). While these roles in *S. enterica* serovar Choleraesuis demonstrate that SodC2 is the primary SOD during macrophage and cell survival, many of the findings do not align with what has been seen in *S. enterica* serovar Typhimurium, suggesting that the contributions from each of the SodC may vary from strain to strain and/or serovar to serovar. In *S. enterica* serovar Typhimurium, deletion of SodC1, but not of SodC2, results in a 7- to 10-fold reduction in virulence during intraperitoneal competitions (129, 130, 166). The differences in effects on virulence between the two SodCs in *S. enterica* serovar Typhimurium are likely due to two key differences in the proteins themselves. First, SodC1 forms a dimer like most other Cu/Zn SODs while SodC2 forms a monomer; and second, SodC1 tethers to the periplasm via peptidoglycan binding while SodC2 does not (166, 167, 266). The latter feature may prevent SOD efflux when anti-microbial peptides permeabilize the outer membrane during infection. In *S. enterica* serovar Typhimurium, the periplasmic SodC1 that is essential to virulence protects the

cell specifically from phagocytic superoxide and acts independently from the cytoplasmic SODs, indicating that the target for extracellularly generated superoxide is also extracytoplasmic (65). This is a topic that is still perplexing the field today but the current thought is that carbohydrates or monounsaturated fatty acids in the periplasm/extracellular space may be the target of this extracellularly generated superoxide (137).

1.7 Superoxide dismutase diversity and metal availability

As all known SODs are dependent on metals for function, it stands to reason that some of the diversity of bacterial SODs is driven by the need to retain SOD activity in distinct environments. While it remains to be tested, it seems reasonable to speculate that organisms with only a single, cambialistic SOD such as *P. gingivalis*, *S. mutans*, and *S. pneumoniae* may be well-situated to maintain a defense against oxidative stress despite changing metal availability within the host (181, 205, 206, 293). For organisms with more than one SOD, recent work in *S. aureus* and the fungus *C. albicans* suggests that this enhanced repertoire of SODs enables pathogens to retain SOD activity when the host imposes metal starvation. In *S. aureus*, the Mn-dependent SodA is considered the primary SOD. However, host-imposed Mn starvation enacted in part by the immune effector calprotectin (CP) decreases SodA activity. Concurrent with this, *S. aureus* induces and derives most of its SOD activity from Fe-loaded cambialistic SodM (101). As noted earlier, in the liver of wild type mice, which sequester Mn away

from *S. aureus*, SodM is necessary to achieve wild type bacterial burdens.

However, in CP-deficient mice, which fail to sequester Mn, SodA becomes critical to infect the liver, while SodM is dispensable (101).

C. albicans possesses two cytoplasmic SODs with differing metal specificity, namely a Cu/Zn SOD and a Mn-SOD (42). Similar to *S. aureus*, it uses the differing metal specificity of these enzymes to retain a defense against superoxide in the host. During periods of Cu limitation, such as that which occurs in brain and kidney tissues, *C. albicans* switches its cytosolic SOD from the Cu-requiring Sod1 to the Mn-dependent Sod3 (174).

Similar observations have also been made in laboratory strains of *E. coli*, with the importance of the Fe-dependent SodB and Mn-dependent SodA being dependent on the availability of their cognate metal (9, 29, 226). In accordance with this, *E. coli* cells have a regulatory system to help ensure the switch from SodB to SodA in times of Fe limitation. This system relies on the ferric uptake repressor protein, Fur, and a small RNA, RyhB. When Fe is available, Fur will bind Fe and be active, repressing transcription of Mn-dependent SodA and RyhB, allowing transcription of only Fe-dependent SodB (188). Then, when Fe becomes limiting, Fur no longer represses RyhB, which can then repress SodB, leading to the transcription of only SodA (188). Together, these results in *E. coli* further suggest that possessing multiple SODs with different metal specificities is a common survival strategy used by microbes. While there have been numerous advances in our understanding of the diversity of SODs, these studies also

demonstrate that there is still much to be learned about the forces that drive expansion and diversification of the SOD repertoires possessed by microorganisms.

1.8 Metalloprotein metal loading

Metalloproteins, like SODs, must acquire their cognate metal from the cell in which they are produced. Generally, it is accepted that most metalloproteins acquire their metal cofactor from the buffered/free intracellular metal pools (46, 93). In this model, there are three main pools for each metal (14, 18, 115, 183). The primary pool is fed by metal importers and acts as a metal source for both the secondary pool, consisting of metal regulators, and the tertiary pool, consisting of other metalloproteins (18). The availability of any given metal for a metalloprotein would then be a product of the abundance of that metal within the metal pool, the lability of that metal within the pool, and the thermodynamic stability of the metal complex with the pool ligand as compared to the proteinaceous ligand (18). It is important to note that metal ions within the cell are not considered to be “free”, they will be bound either to a ligand within the metal pool or to a metalloprotein. These ligands thereby “buffer” the availability of the metal within the cell. The identity of ligands within the metal pools has not been fully elucidated but they could include free amino acids, nucleic acids, and proteins, among others (49). According to the Irving-Williams series, all divalent metal ions have intrinsic binding affinities for proteinaceous ligands that place

them in the following order: Mg^{2+} and $\text{Ca}^{2+} < \text{Mn}^{2+} < \text{Fe}^{2+} < \text{Co}^{2+} < \text{Ni}^{2+} < \text{Cu}^{2+} > \text{Zn}^{2+}$ (142). The cell will attempt to fine tune the bioavailability of metals such that metalloproteins can successfully bind cognate metals while not binding noncognate metals (46, 140, 239, 280). For example, the amount of available Zn is kept low, as it has a high affinity for proteins (142, 271, 280). If the amount of available Zn were not limited, it would bind non-zinc metalloproteins, causing them to lose activity (18, 139). To aid in the successful metalation of metalloproteins, the cell attempts to buffer the bioavailability of metals to be inverse that of the Irving-Williams series, e.g. Mn more available than Zn, etc. (46, 140, 280). During non-stressed situations, the cell can achieve proper metal buffering. However, when the cell experiences stress, such as metal scarcity or overabundance, then achieving the proper buffering to ensure the correct metalation of metalloproteins becomes difficult (53). For example, metal loading of Fe/Mn-SODs is influenced by exogenous metal concentrations during both heterologous expression inside *E. coli* cells and, in some cases, in their own native host (17, 29, 186). This influence of exogenous metal concentrations on intracellular metalation is because environmental metal levels influence which metals are available inside the cell and, therefore, the metal pools that help dictate metalloprotein metal loading. Finally, it is not well understood exactly what role, if any, a protein itself plays in selecting or influencing the metal obtained within this buffered system (46, 93).

1.9 Staphylococcal SodA and SodM as foundations for studying pressures on metalloenzyme neofunctionalization

In *S. aureus*, the two SODs (SodA and SodM) are both in the Fe/Mn-SOD family. SodA is a Mn-dependent enzyme important when Mn is available, while SodM is cambialistic and important when the cell is experiencing Mn starvation (61, 101, 154, 275). Despite their difference in metal activity profiles, SodA and SodM have identical protein folds and are 75% identical at the amino acid level (149/199 residues), with most of the differences mapping to the outside of the protein, away from the metal coordination region. Additionally, SodA and SodM coordinate their metal ions with the same 4 residues (His81, His27, His165, and Asp161). Therefore, the basis of their differing metal specificity is not immediately obvious.

Previous work identified secondary sphere residues that influence metal specificity in the staphylococcal SODs (17). These residues were first identified because they differ between SodA and SodM and are within 10 angstroms of the metal coordination region. In SodA they are: F19, G159, and L160, whereas in SodM they are: I19, L159, and F160. For wild-type SodA, the Mn activity is ~ 1836 units mg^{-1} protein and the Fe activity is ~ 4 units mg^{-1} protein leading to an Fe-dependent activity to Mn-dependent activity (or “cambialism ratio”) of 0.002 (17). For wild-type SodM, the Mn activity is ~ 251 units mg^{-1} protein and the Fe activity is ~ 250 units mg^{-1} protein leading to a cambialism ratio of 0.996 (17). When residue 19 is swapped between SodA and SodM, neither SOD variant exhibits a shift in metal specificity. For residue swapping at residue 160, Fe-

dependent activity of SodA L160F is increased to 19 units mg^{-1} protein and Mn-dependent activity of SodM F160L is decreased to 111 units mg^{-1} protein. The metal activity profiles shift slightly further when residue 159 is swapped, with SodA G159L Mn-dependent activity decreasing to 942 units mg^{-1} protein and Fe-dependent activity increasing to 39 units mg^{-1} protein, making it more cambialistic. For residue swapping in SodM, SodM L159G becomes more Mn-dependent with an increase to 808 units mg^{-1} protein while the Fe-dependent activity decreased to 24 units mg^{-1} protein. When residues 159 and 160 are swapped in combination, the SodA and SodM variants make the largest shifts in metal specificities. SodA G159L-L160F Mn-dependent activity decreased to 173 units mg^{-1} protein while Fe-dependent activity increased to 84 units mg^{-1} protein leading to a cambialism ratio of 0.486. SodM L159G-F160L Mn-dependent activity increased to 687 units mg^{-1} protein while Fe-dependent activity decreased to 11 units mg^{-1} protein leading to a cambialism ratio of 0.016. However, when all three residues – 19, 159, and 160 – are swapped, each SOD variant shifts slightly back towards that of the wild-type enzyme with the cambialism ratio of SodA F19I-G159L-L160F decreasing to 0.108 and the cambialism ratio of SodM I19F-L159G-F160L increasing to 0.049. The lesser degree of altered metal activity profiles associated with swapping all three residues indicates there must be other factors dictating metal specificity in these enzymes (17).

Based on the fact that SodM is more similar to SodA than to any other known SOD and the restriction of SodM to only *S. aureus* and highly related species, this previous work hypothesized that *sodM* arose first as a duplication of *sodA* and then via divergence to present-day *sodM* (17, 85, 198, 254, 258). In accordance with this idea, it can be assumed that the ancestral SodM was a strictly Mn-dependent enzyme and that modern SodM has gained the ability to utilize Fe. Together, these observations lead to the hypothesis that the need to maintain a defense against oxidative stress when the organism is Mn-starved selected for the ability of SodM to utilize Fe. The discovery of residues important to dictating metal specificity in SodA and SodM [residues 19, 159, and 160], enabled the creation of potential *S. aureus* evolutionary intermediate strains encoding SodA variants on the path from Mn-dependency to cambialism. This set of *S. aureus* strains, together with the ability to impose physiologically relevant stressors, makes the staphylococcal SODs and *S. aureus* important tools, allowing for studies on metalloenzyme neofunctionalization to be performed within the context of a native host organism. In the following sections, *S. aureus* and metal restriction, including what is mediated by the metal binding host immune protein, CP, will be described to further support their use as tools in this study.

1.10 *Staphylococcus aureus* is a serious threat to human health

Despite over 140 years of investigation, the sophisticated pathogenesis of *S. aureus* remains a threat to human health (208). *S. aureus* is a gram-positive

cocci bacterium that can grow either aerobically or anaerobically. It forms grape-like clusters, and its species name “aureus” is derived from an ancient Roman coin that was gold in color – the typical colony color of this organism. *S. aureus* is widely prevalent in humans, as it asymptomatically colonizes the nasal cavity of around one-third of the human population (163, 170, 210, 246). This nasal colonization is more common in those with compromised immune systems, such as individuals with HIV or diabetes, as well as young children (176, 286). Once it has broken through the epithelial barrier, *S. aureus* can infect every organ within the human body, and it can cause diseases including endocarditis, osteomyelitis, necrotizing fasciitis and pneumonia, urinary tract infections, and surgical site infections, among others (12, 83, 95, 152, 173, 179, 243, 268, 286). Each year in the US alone, infections with Methicillin-resistant *S. aureus* account for over 323,000 hospitalizations and over 10,000 deaths. This high level of morbidity, mortality, and antibiotic-resistant infections has led the CDC and WHO to list *S. aureus* as a “serious” threat to human health and to call for the discovery of novel therapeutics (51).

S. aureus produces an array of virulence factors, and they are promising targets for novel therapeutics. These virulence factors aid in the pathogenesis of *S. aureus* in a multitude of ways, including by disabling the host immune response (e.g., inhibiting the function of neutrophils or T cells, and evading phagocytosis) and by assisting in tissue penetration via production of adhesins, exoenzymes, and toxins (16, 27, 54, 70, 78, 92, 150, 172, 194, 265). Early in infection, it is advantageous for *S. aureus* to produce adhesins to mediate

interactions with the host cell. Among them are the “microbial surface components recognizing adhesive matrix molecules” (MSCRAMMs), which recognize collagen, fibronectin, and fibrinogen (96, 108). Once inside the host organism, *S. aureus* degrades host tissues by releasing a variety of enzymes, including proteases and lipases, leading to spread into deeper tissues (264). Central to the success of *S. aureus* is its arsenal of secreted and produced toxins (264). These toxins account for 10% of the total secretome and have a variety of roles, including lysis of host cells, mediation of cytokine production, and changes in host cell activity and physiology (168, 25, 218, 250, 264). *S. aureus* must carefully regulate the production of these factors in order to ensure successful infection (55). To control this network, *S. aureus* has a collection of global regulators. These include two-component systems (TCSs) and the SarA protein family. The TCSs are thought to sense external signals. *S. aureus* encodes sixteen TCSs, with one being essential and many being important for virulence (116, 190, 229, 287). Meanwhile, the SarA protein family, in conjunction with alternative transcription factors and the response components of TCSs, are thought to act as internal effectors in the response to environmental signals (10, 55, 146). SarA-family members SarA, Rot, and MgrA, in addition to the alternative sigma factor SigB, are considered to be the most important for staphylococcal pathogenesis (58, 146). Through direct and indirect interactions, SarA can control over 120 genes, and *S. aureus* *sarA* mutants have decreased virulence in a variety of infection models (34, 36, 56, 79, 209, 298). While each of

these factors is important for staphylococcal pathogenesis, successful infection by *S. aureus* also requires the ability to combat the various host-imposed strategies for limiting essential nutrients like metals during infection.

1.11 Importance, restriction, and acquisition of transition metals Fe, Mn, and Zn in biology

Unlike the cells of higher organisms, bacteria and other microbes do not fully control the nutrient content of the environment in which they reside. Therefore, all essential nutrients that cannot be synthesized by bacteria must be imported from the environment. For pathogens during infection, this means they must obtain these nutrients from the host. Similar to leveraging the toxicity of ROS, the host restricts the availability of metals to prevent infection (26, 120, 133, 156, 157, 284). This restriction can lead to the inactivation of a myriad of metal-dependent bacterial enzymes, including SODs (26, 120, 133, 156, 157, 284). For bacteria, Fe, Mn, and Zn are the most important metals for metabolism, but are also known to be restricted by the host (53). Despite this metal limitation, pathogens like *S. aureus* are still able to cause disease, indicating that they must possess mechanisms to evade this host-mediated metal restriction.

Importance, restriction, and acquisition of Fe

Iron is the most widely used transition metal in biology. In bacteria, Fe is important for the pentose phosphate pathway, DNA replication, and cofactor synthesis as it is used as a cofactor in metalloproteins either directly or as an Fe-

sulfur cluster or heme (8, 60, 148, 213). Adding to this importance is the versatility in how metalloenzymes can use Fe. This is due to the ability of Fe to shift between its reduced, ferrous state (Fe^{2+}) and oxidized, ferric state (Fe^{3+}).

Fe is extremely limited in its free, bioavailable state, with environmental levels ranging from 10^{-8} – 10^{-9} M and levels within the mammalian host around 10^{-24} M (43, 233, 234). When Fe is encountered by invading pathogens in the host, it is usually bound to proteins such as transferrin, lactoferrin, and ferritin, or is present in heme (126, 153, 248, 284). Extracellular Fe exists as Fe^{3+} at physiological pH and is bound to transferrin to be transported throughout the host (21, 48). Once inside the host cell, Fe is primarily complexed to the Fe storage protein ferritin. The largest pool of Fe within the host is found circulating in erythrocytes associated with hemoglobin. Hemoglobin accounts for up to 80% of total body Fe, while transferrin accounts for less than 1% (67, 255, 285). In an attempt to sequester free hemoglobin away from pathogens, the host deploys haptoglobin, a high-affinity hemoglobin-binding protein (76, 119). The lysis of erythrocytes also releases free heme, which the host sequesters through the use of hemopexin (267). During times of infection, the host can use a variety of mechanisms to limit Fe even further, including altering the level of saturated transferrin from ~25–35% to ~5%, markedly decreasing the amount of circulating Fe (48, 285). Given the fact that Fe is always restricted and that the host has additional methods of restricting Fe even further, it is critical that microbes have mechanisms to counteract this extreme Fe limitation.

One of the reasons *S. aureus* is such a well-adapted pathogen is that it has the ability to obtain Fe from many sources, including directly importing free Fe and heme (19, 118, 119, 191, 255, 260, 270). Further, *S. aureus* also excretes two siderophores, staphyloferrin A and staphyloferrin B, which can extract Fe³⁺ from transferrin and lactoferrin and carry it into the bacterium (119, 252). These siderophores are produced in response to Fe-starvation and are important to staphylococcal virulence (20, 134, 252). Interestingly, staphylococcal abscesses exhibit differences in siderophore production, with some abscesses producing a higher abundance of staphyloferrin A while others generate more staphyloferrin B (220). This difference in siderophore production is likely due to distinct metabolism in each abscess with TCA cycle activity being essential for Fe-regulated synthesis of staphyloferrin A, not staphyloferrin B (253). This suggests that these siderophores are not functionally redundant and instead have niche-specific roles (220).

The preferred Fe source for *S. aureus* is heme (255). When *S. aureus* senses Fe-limitation, it upregulates an alpha toxin, Hla, which promotes hemolysis of erythrocytes (184, 269). This releases the heme-associated hemoglobin, which *S. aureus* can utilize as a source of Fe (270). *S. aureus* can also utilize haptoglobin-hemoglobin complexes, however, it cannot utilize heme sequestered in hemopexin (76, 119, 270). Once *S. aureus* does find heme it can utilize, it relies on the iron-regulated surface-determinant pathway (Isd) to bind, extract, import, and degrade the heme to access the Fe (191, 255). The importance of this ability to import Fe is evidenced by the increased susceptibility

of a *S. aureus* heme uptake mutant to killing by human neutrophils, as well as its decreased bacterial burdens in mouse models of systemic infection (212, 225, 236, 270, 279). Interestingly, the less pathogenic staphylococcal species lack the genes required for production of staphyloferrin B as well as heme utilization (17, 19). These findings further support the importance of Fe acquisition in this pathogen and also suggest that the ability to import Fe may have supported the transition to a pathogenic lifestyle in *S. aureus* (17).

Importance, restriction, and acquisition of Mn

The importance of Mn in cellular processes ranges from metabolism to phosphorylation to oxidative stress defense (9, 101, 156, 214). More specifically, Mn is used as a cofactor for phosphoglycerate mutase, pyruvate kinase, and Mn-SODs as well as in enzymes important for DNA synthesis (151, 213, 230). Similar to Fe, Mn can also shift between its Mn^{2+} and Mn^{3+} oxidation states, allowing its use in redox enzymes. Within the host, Mn levels are <10 nM in the serum and ~10 μ M in the liver (155, 158, 232). Within the mouse, similar Mn concentrations in the liver are observed (151, 158). This metal content also varies in different locations within the host; for example, uninfected mice kidneys contain half the amount of Mn (and Zn) as uninfected livers (158). During infection, bulk Mn levels in the liver are actively restricted and Mn levels are locally restricted at sites of infection in multiple tissues (151, 158). Mn levels in the macrophage phagolysosome are also limited. This is suggested to be

accomplished by a pH-dependent Mn transporter, Nramp1, which pumps Mn out of lysosomes restricting the ability of invading intracellular pathogens to acquire this metal (50, 143, 217).

One mechanism host cells will use to sequester extracellular Mn is that of the deployment of CP. CP is a member of the S100 EF-hand family of calcium binding proteins. Typically, members of this protein family are homodimers; however, CP is a heterodimer of S100A8 and S100A9 leading to its two non-identical metal binding sites (69, 164). The other S100 proteins that bind transition metals have two identical metal-binding sites that present three His and one Asp ligand. In CP, one of the metal-binding sites similarly presents three His and one Asp ligand (His83 and His87 from S100A8 and His20 and Asp30 from S100A9) and it binds Zn at a subpicomolar affinity (40, 156). However, the other site presents six His ligands (His17 and His27 from S100A8 and His91, His95, and two additional histidines from S100A9) and binds Mn with low-nanomolar and Zn with sub-picomolar affinity (40, 69, 125, 203). While capable of binding multiple first-row transition metals, in common culture conditions CP primarily prevents the pathogen *S. aureus* from obtaining Mn, and to some extent Zn, but not Fe or other metals (204, 229, 231). The addition of CP to lab media inhibits the growth of *S. aureus* and other microbes (26, 30, 64, 69, 156, 274, 299). During infection, CP comprises 50% of the neutrophil protein content and can be found at sites of infection in excess of 1 mg/ml (62, 64, 102, 158). This leads to staphylococcal abscesses which are rendered effectively devoid of both Mn and Zn, with the Mn sequestration being dependent upon CP (64, 156). In addition,

CP-deficient mice are more sensitive to infection by a wide variety of fungal and bacterial pathogens, including *S. aureus* (30, 64, 274). The use of CP to impose metal restriction on microbes has greatly enhanced the ability of investigators to evaluate the effects of nutritional immunity as the effects of purified CP in culture often recapitulates what occurs during animal infection studies (64, 72, 74, 101, 112, 113, 132, 158, 229, 230).

In the case of *S. aureus*, similar to Fe, Mn is another important metal to pathogenesis (145). *S. aureus* possesses two distinct Mn uptake systems: an ATP-binding cassette (ABC) system, MntABC, and a natural resistance-associated macrophage protein (NRAMP) homolog, MntH (135, 214). No siderophore-like system has been discovered for Mn transport (151). The ABC system Mn transporters are widely distributed in bacteria and the Nramp homologs are conserved across all domains of life (151, 214). For *S. aureus*, both MntABC and MntH contribute to staphylococcal infection, as loss of both systems leads to decreased bacterial burdens during infection (135, 158). MntH is constitutively expressed, with only a modest increase when Mn is limited, while MntABC is more substantially induced in response to Mn limitation (135, 158, 229). Additionally, the expression of both Mn transport systems is regulated by the metalloregulator, MntR. When MntR is bound to Mn indicating conditions of sufficient Mn, it will repress transcription of *mntABC* (135). When Mn levels are low, MntR will no longer be bound to Mn and will de-repress *mntABC* while leading to maximal expression of *mntH*, together increasing Mn transport in Mn-limited conditions (135). MntABC is the primary transporter responsible for

competing with CP for Mn during infection; accordingly, mutants in the MntABC system exhibit increased killing by human neutrophils, and mice infected with *S. aureus* $\Delta mntC$ mutants exhibit increased survival (63, 64, 122, 231).

Another response to Mn limitation employed by *S. aureus* involves the regulation of the staphylococcal SODs. As noted in section 1.5, *S. aureus* encodes Mn-dependent SodA and cambialistic, Mn- or Fe-dependent, SodM. When Mn is available, the cell relies on SodA and when Mn is depleted, the cell switches to relying on SodM. One way *S. aureus* helps enable this switch is through the alteration of SOD expression. Transcriptional expression of *sodM* is low when Mn is available and increased in response to Mn limitation (101). This expression of SodM is controlled by the global regulator SarA and the alternative sigma factor SigB (13, 154). SarA and SigB are each important for the virulence of *S. aureus*, with SarA regulating the expression of over 100 genes and SigB of over 200 genes, including *sarA* where it acts as an activator (31, 32, 35, 52, 57, 59, 79, 94, 103, 117, 149, 200, 289, 297). SarA interacts directly with the promoter of *sodM* to suppress expression (13). While it is known that SarA and SigB contribute to negatively regulating *sodM* expression, the mechanism of how *S. aureus* responds to Mn availability in order to alter *sodM* expression is unknown. Also contributing to the regulation of SOD expression is post-transcriptional regulation of *sodA* through the presence of a small RNA, RsaC. RsaC is co-transcribed with the MntABC operon therefore making it induced in response to Mn limitation (135, 154, 158, 169). Once transcribed, RsaC binds directly to the SodA ribosome binding site (RBS) leading to repression of SodA

translation (169). This post-translational regulation presumably also helps in the switch from SodA to SodM in response to Mn limitation by potentially enabling SOD activity while also preventing SodA from acting as a sink for metals, such as Mn and Fe, allowing the cell to redirect those metals to other proteins, like SodM. Interestingly, when *E. coli* cells experience iron depletion, the Mn importer *mntH* is strongly induced, suggesting that Mn acts as a substitute and is critical for the cell when Fe is unavailable (110, 136, 139, 159). Accordingly, Mn could substitute for Fe in non-redox metalloenzymes, allowing the low amount of available Fe to be directed to other enzymes, including heme and iron-sulfur cluster containing enzymes (139). This suggests that a metal sparing system might be a common defense used across many organisms.

Importance, restriction, and acquisition of Zn

Among bacteria, 4.9% of the entire proteome is predicted to bind Zn, while in the human host, this rises as high as 10%, showcasing the widespread importance of this metal (5, 6). Unlike other metals, Zn is not redox active, instead providing structure stabilization or catalytic function to proteins (11). Like other transition metals, Zn is limited by the host during infection (178). One example of this is driven by a host Zn transporter, ZIP8 (22). The current model is that, during infection, ZIP8 orients such that it transports Zn from the lysosome into the host cell cytoplasm thereby removing Zn from invading pathogens and redirecting it to the host cell (157). Another method of Zn limitation is through the sequestration of extracellular Zn by members of the S100 protein family. One of

these is CP (S100A8/S100A9), discussed previously as it also sequesters Mn from invading pathogens. While CP is capable of sequestering both Mn and Zn, when *S. aureus* was exposed to a mutant variant of CP incapable of sequestering Mn, the ability of CP to limit bacterial growth was reduced (69). Therefore, the sequestration of Zn alone by CP is not sufficient to limit bacterial growth of *S. aureus*. Additionally, *in vitro* studies have indicated that when subjecting *S. aureus* to CP, there is a reduction in intracellular Mn but not Zn, indicating that *S. aureus* is successfully competing with CP for Zn in culture experiments (229). However, as mentioned above, staphylococcal liver abscesses display a reduction in Zn levels as compared to surrounding tissue indicating the bacteria are experiencing Zn restriction during infection (64). The ability to bind Zn has been reported for many other S100 proteins (S100A1, S100A2, S100A3, S100A5, S100A6, S100A7, S100A12, S100A16 and S100B) although the biological relevance for this Zn binding has yet to be established for the majority of these interactions (105). One S100 protein family member whose Zn-binding biological relevance has been investigated is that of S100A7, or Psoriasin. Psoriasin has two metal binding sites that both bind Zn and it is overexpressed in skin cells of human hosts with psoriasis (38, 39). Additionally Psoriasin has been shown to have antimicrobial activity against *E. coli* (106).

Just like Fe and Mn, the ability to acquire Zn despite its restriction is important for invading pathogens. Many pathogens rely on ABC transporters known as zinc uptake (Znu) systems to acquire this metal during infection (133). The Znu system is required for virulence in pathogens such as *Pseudomonas*

aeruginosa and *Acinetobacter baumannii* (74, 132). In the case of *S. aureus*, it acquires Zn via ABC-type transporters either alone or in combination with a siderophore-like system (113). AdcABC is a member of the most widely distributed bacterial Zn importer family. Its expression is induced when Zn is limited (113). The siderophore-like system utilizes CntKLM and CntE to produce and secrete staphylopin, a broad spectrum metallophore (104, 113). The CntABCDF transporter then imports the metal-bound staphylopin inside the cell (113, 235). Even though staphylopin is capable of binding many first row transition metals, *S. aureus* utilizes it as a zincophore for Zn acquisition (113). Loss of both of these Zn acquisition systems (Adc/Cnt) leads to a significant reduction in *S. aureus* bacterial burdens in hosts, demonstrating the importance of Zn acquisition in this pathogen (112, 113). As noted earlier, *S. aureus* is capable of competing with CP for Zn, and it is the Cnt system, not AdcABC, that is responsible for this (113). Mutants in the Cnt system have decreased bacterial burdens during infection (112, 113).

1.12 Aim of this study

It is clear that metalloprotein metal utilization has changed over evolutionary time (7, 17, 80). However, the factors that drove these changes remain enigmatic due to the lack of knowledge regarding the identity of the ancestral enzyme. The work in this thesis leveraged the unique opportunity provided by the staphylococcal SODs, where SodA represents the ancestral enzyme and SodM represents a neofunctionalized version of SodA, to evaluate if

metal availability can drive metalloprotein evolution (17). Using SOD variants with mutations at key residues, a suite of potential evolutionary intermediates on the path from Mn-dependency to cambialism were created (17). By subjecting these intermediates to metal limitation and increased oxidative stress, the ability of metal limitation to drive metalloenzyme neofunctionalization was evaluated within the context of a native host experiencing physiologically relevant stressors.

1.13 Figures

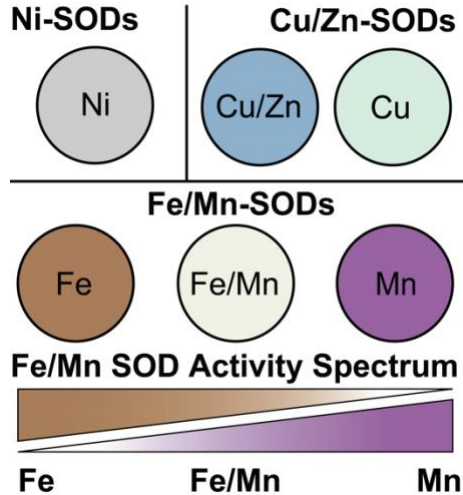


Figure 1.1. Diagram of the three superoxide dismutase (SOD) protein families and their metal-utilization. SODs exist in three distinct protein families that each rely on a distinct set of metals for activity. These families are nickel SODs, copper and zinc SODs, and iron/manganese SODs. Recently, the Cu/Zn-SOD family was discovered to contain enzymes that require only Cu for function. Similarly, the Fe/Mn-SOD family was realized to contain members that are active with either only Mn or only Fe, but also enzymes that can use either metal for catalysis (cambialistic). Moreover, it has become apparent that metal-specificity in this family is best thought of as a spectrum rather than distinct bins. In the diagram spheres represent individual members of each SOD family and the metals they are active with.

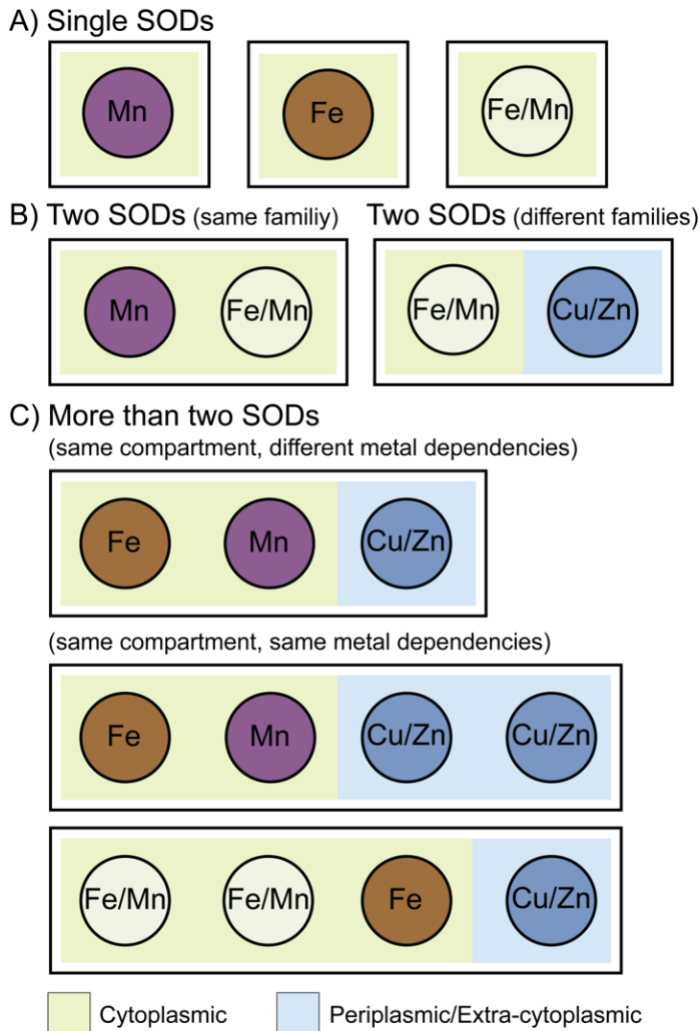


Figure 1.2. Superoxide dismutase (SOD) repertoires possessed by bacteria.

While the threat of superoxide and possession of SOD(s) is nearly universal among bacteria, the repertoire of enzymes they possess is extremely varied. (A) Bacteria can encode only a single SOD that can be either Mn-dependent (purple, e.g., *Borrelia burgdorferi*), Fe-dependent (brown, e.g., *Campylobacter jejuni*), or cambialistic (green, e.g., *Streptococcus mutans*). (B) When bacteria encode two SODs, they can either be from the same SOD family (Fe/Mn family, e.g., *Staphylococcus aureus*), or from different SOD families (Fe/Mn family, light grey, Cu/Zn family, blue e.g., *Mycobacterium tuberculosis*). (C) Finally, there are also many bacteria that encode more than two SODs. This can result in the presence of more than one SOD in a cellular compartment and can result in enzymes from the same family, with either similar or distinct metal specificity, occupying the same compartment (Fe- and/or Mn-dependent in the cytoplasm, tan, Cu/Zn in the periplasm/extra-cytoplasmic space, light blue e.g., *Escherichia coli*, *Salmonella enterica* serovars and Bacilli). Despite carrying out the same chemical reactions, each SOD in these instances appears to uniquely contribute to the lifestyle of the harboring organisms.

1.14 References

1. Aguirre JD, Clark HM, McIlvin M, Vazquez C, Palmere SL, et al. 2013. A manganese-rich environment supports superoxide dismutase activity in a lyme disease pathogen, *Borrelia burgdorferi*. *Journal of Biological Chemistry*. 288(12):8468–78.
2. Amano A, Shizukuishi S, Tamagawa H, Iwakura K, Tsunasawa S, Tsunemitsu A. 1990. Characterization of superoxide dismutases purified from either anaerobically maintained or aerated *Bacteroides gingivalis*. *J Bacteriol*. 172(3):1457–63.
3. Ammendola S, Ajello M, Pasquali P, Kroll JS, Langford PR, et al. 2005. Differential contribution of *sodC1* and *sodC2* to intracellular survival and pathogenicity of *Salmonella enterica* serovar Choleraesuis. *Microbes and Infection*. 7(4):698–707.
4. Anbar AD. 2008. Oceans: Elements and evolution. *Science*. 322(5907):1481–83.
5. Andreini C, Banci L, Bertini I, Rosato A. 2006. Zinc through the three domains of life. *J. Proteome Res*. 5(11):3173–78.
6. Andreini C, Banci L, Bertini I, Rosato A. 2006. Counting the zinc-proteins encoded in the human genome. *J. Proteome Res*. 5(1):196–201.
7. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM. 2008. Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem*. 13(8):1205–18.
8. Andrews SC. 1998. Iron storage in bacteria. In *Advances in Microbial Physiology*. 40:281–351.
9. Anjem A, Varghese S, Imlay JA. 2009. Manganese import is a key element of the OxyR response to hydrogen peroxide in *Escherichia coli*. *Molecular Microbiology*. 72(4):844–58.
10. Arvidson S, Tegmark K. 2001. Regulation of virulence determinants in *Staphylococcus aureus*. *International Journal of Medical Microbiology*. 291(2):159–70.
11. Auld DS. 2009. The ins and outs of biological zinc sites. *Biometals*. 22(1):141–48.

12. Baddour LM, Wilson WR, Bayer AS, Fowler VG, Bolger AF, et al. 2005. Infective endocarditis: Diagnosis, antimicrobial therapy, and management of complications: A statement for healthcare professionals from the committee on rheumatic fever, endocarditis, and kawasaki disease, council on cardiovascular disease in the young, and the councils on clinical cardiology, stroke, and cardiovascular surgery and anesthesia, american heart association: Endorsed by the infectious diseases society of america. *Circulation*. 111(23).
13. Ballal A, Manna AC. 2009. Regulation of superoxide dismutase (*sod*) genes by *sarA* in *Staphylococcus aureus*. *Journal of Bacteriology*.
14. Banci L, Bertini I, Ciofi-Baffoni S, Kozyreva T, Zovo K, Palumaa P. 2010. Affinity gradients drive copper to cellular destinations. *Nature*. 465(7298):645–48.
15. Barra D, Schininà ME, Bossa F, Puget K, Durosay P, et al. 1990. A tetrameric iron superoxide dismutase from the eucaryote *Tetrahymena pyriformis*. *J Biol Chem*. 265(29):17680–87.
16. Bartlett AH, Hulten KG. 2010. *Staphylococcus aureus* pathogenesis: Secretion systems, adhesins, and invasins. *Pediatric Infectious Disease Journal*. 29(9):860–61.
17. Barwinska-Sendra A, Garcia YM, Sendra KM, Baslé A, Mackenzie ES, et al. 2020. An evolutionary path to altered cofactor specificity in a metalloenzyme. *Nat Commun*. 11(1):2738.
18. Barwinska-Sendra A, Waldron KJ. 2017. The Role of intermetal competition and mis-metalation in metal toxicity. In *Advances in Microbial Physiology*. 70:315–79.
19. Beasley FC, Heinrichs DE. 2010. Siderophore-mediated iron acquisition in the staphylococci. *Journal of Inorganic Biochemistry*. 104(3):282–88.
20. Beasley FC, Marolda CL, Cheung J, Buac S, Heinrichs DE. 2011. *Staphylococcus aureus* transporters Hts, Sir, and Sst capture iron liberated from human transferrin by Staphyloferrin A, Staphyloferrin B, and catecholamine stress hormones, respectively, and contribute to virulence. *Infect Immun*. 79(6):2345–55.
21. Becker KW, Skaar EP. 2014. Metal limitation and toxicity at the interface between host and pathogen. *FEMS Microbiology Reviews*. 38(6):1235–49.

22. Begum NA, Kobayashi M, Moriwaki Y, Matsumoto M, Toyoshima K, Seya T. 2002. *Mycobacterium bovis* BCG cell wall and lipopolysaccharide induce a novel gene, BIGM103, encoding a 7-TM protein: identification of a new protein family having Zn-transporter and Zn-metalloprotease signatures. *Genomics*. 80(6):630–45.
23. Bekker A, Holland HD, Wang P-L, Rumble D, Stein HJ, et al. 2004. Dating the rise of atmospheric oxygen. *Nature*. 427(6970):117–20.
24. Benov L, Chang LY, Day B, Fridovich I. 1995. Copper, zinc superoxide dismutase in *Escherichia coli*: Periplasmic localization. *Archives of Biochemistry and Biophysics*. 319(2):508–11.
25. Berube B, Wardenburg J. 2013. *Staphylococcus aureus* α -Toxin: Nearly a century of intrigue. *Toxins*. 5(6):1140–66.
26. Besold AN, Gilston BA, Radin JN, Ramsoomair C, Culbertson EM, et al. 2018. Role of calprotectin in withholding zinc and copper from *Candida albicans*. *Infection and Immunity*. 86(2):1–16.
27. Bestebroer J, Poppelier MJJG, Ulfman LH, Lenting PJ, Denis CV, et al. 2007. *Staphylococcal* superantigen-like 5 binds PSGL-1 and inhibits P-selectin-mediated neutrophil rolling. *Blood*. 109(7):2936–43.
28. Beyer W, Imlay J, Fridovich I. 1991. Superoxide dismutases. *Progress in Nucleic Acid Research and Molecular Biology*. 40(C):221–53.
29. Beyer WF, Fridovich I. 1991. *In vivo* competition between iron and manganese for occupancy of the active site region of the manganese-superoxide dismutase of *Escherichia coli*. *Journal of Biological Chemistry*. 266(1):303–8.
30. Bianchi M, Niemiec MJ, Siler U, Urban CF, Reichenbach J. 2011. Restoration of anti-*Aspergillus* defense by neutrophil extracellular traps in human chronic granulomatous disease after gene therapy is calprotectin-dependent. *Journal of Allergy and Clinical Immunology*. 127(5):1243-1252.e7.
31. Bischoff M, Dunman P, Kormanec J, Macapagal D, Murphy E, et al. 2004. Microarray-based analysis of the *Staphylococcus aureus* sigma B regulon. *Journal of Bacteriology*. 186(13):4085–99.
32. Bischoff M, Entenza JM. 2001. Influence of a functional *sigB* operon on the global regulators *sar* and *agr* in *Staphylococcus aureus*. 183(17):5171–79.

33. Blackney MJ, Cox R, Shepherd D, Parker JD. 2014. Cloning and expression analysis of *Drosophila* extracellular Cu Zn superoxide dismutase. *Bioscience Reports*. 34(6):e00164.
34. Blevins JS, Elasri MO, Allmendinger SD, Beenken KE, Skinner RA, et al. 2003. Role of *sarA* in the pathogenesis of *Staphylococcus aureus* musculoskeletal infection. *Infect Immun*. 71(1):516–23.
35. Blevins JS, Gillaspay AF, Tammy M, Hurlburt BK, Smeltzer MS. 1999. The staphylococcal accessory regulator (*sar*) represses transcription of the *Staphylococcus aureus* collagen adhesin gene (*cna*) in an *agr*-independent manner. 33:317–26.
36. Booth MC, Cheung AL, Hatter KL, Jett BD, Callegan MC, Gilmore MS. 1997. *Staphylococcal* accessory regulator (*sar*) in conjunction with *agr* contributes to *Staphylococcus aureus* virulence in endophthalmitis. *Infect Immun*. 65(4):1550–56.
37. Brioukhanov AL, Netrusov AI. 2004. Catalase and superoxide dismutase: distribution, properties, and physiological role in cells of strict anaerobes. *Biochemistry (Moscow)*. 69(9):949–62.
38. Brodersen DE, Nyborg J, Kjeldgaard M. 1999. Zinc-binding site of an S100 protein revealed. Two crystal structures of Ca²⁺-bound human psoriasin (S100A7) in the Zn²⁺-loaded and Zn²⁺-free states. *Biochemistry*. 38(6):1695–1704.
39. Broome A-M, Ryan D, Eckert RL. 2003. S100 protein subcellular localization during epidermal differentiation and Psoriasis. *J Histochem Cytochem*. 51(5):675–85.
40. Brophy MB, Hayden JA, Nolan EM. 2012. Calcium ion gradients modulate the zinc affinity and antibacterial activity of human calprotectin. *Journal of the American Chemical Society*. 134(43):18089–100.
41. Brown JR, Doolittle WF. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev*. 61(4):456–502.
42. Broxton CN, Culotta VC. 2016. SOD enzymes and microbial pathogens: surviving the oxidative storm of infection. *PLoS Pathog*. 12(1):e1005295.
43. Bullen JJ. 1981. The significance of iron in infection. *Clinical Infectious Diseases*. 3(6):1127–38.
44. Burton NA, Schürmann N, Casse O, Steeb AK, Claudi B, et al. 2014. Disparate impact of oxidative host defenses determines the fate of *Salmonella* during systemic infection in mice. *Cell Host & Microbe*. 15(1):72–83.

45. Buvelot H, Posfay-Barbe KM, Linder P, Schrenzel J, Krause K-H. 2016. *Staphylococcus aureus*, phagocyte NADPH oxidase and chronic granulomatous disease. FEMS Microbiology Reviews, p. fuw042.
46. Capdevila DA, Edmonds KA, Giedroc DP. 2017. Metallochaperones and metalloregulation in bacteria. Essays In Biochemistry. 61(2):177–200.
47. Casillas-Martinez L, Setlow P. 1997. Alkyl hydroperoxide reductase, catalase, MrgA, and superoxide dismutase are not involved in resistance of *Bacillus subtilis* spores to heat or oxidizing agents. J Bacteriol. 179(23):7420–25.
48. Cassat JE, Skaar EP. 2012. Metal ion acquisition in *Staphylococcus aureus*: Overcoming nutritional Immunity. Seminars in Immunopathology. 34(2):215–35.
49. Cavet JS, Perry RD, Brunke S, Darwin KH, Fierke CA, et al. 2015. Trace metals in host–microbe interactions: The microbe perspective. In *trace metals and infectious diseases*, ed JO Nriagu, EP Skaar. Cambridge (MA): MIT Press.
50. Cellier MF, Courville P, Champion C. 2007. Nramp1 phagocyte intracellular metal withdrawal defense. Microbes and Infection. 9(14–15):1662–70.
51. Centers for Disease Control and Prevention (U.S.). 2019. Antibiotic resistance threats in the United States, 2019. Centers for Disease Control and Prevention (U.S.).
52. Chan PANF, Foster SJ, Ingham E, Clements MO. 1998. The *Staphylococcus aureus* alternative sigma factor σ^B controls the environmental stress response but not starvation survival or pathogenicity in a mouse abscess model. 180(23):6082–89.
53. Chandrangsu P, Rensing C, Helmann JD. 2017. Metal homeostasis and resistance in bacteria. Nature Reviews Microbiology. 15(6):338–50.
54. Chavakis T, Hussain M, Kanse SM, Peters G, Bretzel RG, et al. 2002. *Staphylococcus aureus* extracellular adherence protein serves as anti-inflammatory factor by inhibiting the recruitment of host leukocytes. Nat Med. 8(7):687–93.
55. Cheung AL, Bayer AS, Zhang G, Gresham H, Xiong YQ. 2004. Regulation of virulence determinants *in vitro* and *in vivo* in *Staphylococcus aureus*. FEMS Immunology and Medical Microbiology. 40(1):1–9.
56. Cheung AL, Eberhardt KJ, Chung E, Yeaman MR, Sullam PM, et al. 1994. Diminished virulence of a *sar-/agr-* mutant of *Staphylococcus aureus* in the rabbit model of endocarditis. J. Clin. Invest. 94(5):1815–22.

57. Cheung AL, Koomey JM, Butler CA, Projan SJ, Fischetti VA. 1992. Regulation of exoprotein expression in *Staphylococcus aureus* by a locus (*sar*) distinct from *agr*. *Proceedings of the National Academy of Sciences of the United States of America*. 89(14):6462–66.
58. Cheung AL, Nishina KA, Trottonda MP, Tamber S. 2008. The SarA protein family of *Staphylococcus aureus*. *International Journal of Biochemistry and Cell Biology*. 40(3):355–61.
59. Chien Y, Manna AC, Projan SJ, Cheung AL. 1999. SarA, a global regulator of virulence determinants in *Staphylococcus aureus*, binds to a conserved motif essential for *sar*-dependent gene regulation. 274(52):37169–76.
60. Choby JE, Skaar EP. 2016. Heme synthesis and acquisition in bacterial pathogens. *Journal of Molecular Biology*. 428(17).
61. Clements MO, Watson SP, Foster SJ. 1999. Characterization of the major superoxide dismutase of *Staphylococcus aureus* and its role in starvation survival, stress resistance, and pathogenicity. *Journal of Bacteriology*. 181(13):3898–3903.
62. Clohessy PA, and Golden BE. 1995. Calprotectin-mediated zinc chelation as a biostatic mechanism in host defence. *Scandinavian Journal of Immunology*. 42(5):551–56.
63. Coady A, Xu M, Phung Q, Cheung TK, Bakalarski C, et al. 2015. The *Staphylococcus aureus* ABC-Type manganese transporter MntABC Is critical for reinitiation of bacterial replication following exposure to phagocytic oxidative burst. *PLOS ONE*. 10(9):e0138350.
64. Corbin, B. D., Seeley, E. H., Raab, A., Feldmann, J., Miller, M. R., Torres, V. J., Anderson, K. L., Dattilo, B. M., Anderson, K. L., Dunman, P. M., Gerads, R., Caprioli, R., Nacken, W., Chazin WJ, Skaar EP. 2008. Metal chelation and inhibition of bacterial growth in tissue abscesses. *Science*. 319(February):962–66.
65. Craig M, Slauch JM. 2009. Phagocytic superoxide specifically damages an extracytoplasmic target to inhibit or kill *Salmonella*. *PLoS ONE*. 4(3):e4975.
66. Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, et al. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*. 288(5466):640–49.
67. Crichton RR. 2002. Inorganic biochemistry of iron metabolism: from molecular mechanisms to clinical consequences. Hoboken: Wiley [Imprint] John Wiley & Sons, Incorporated.

68. Cybulski RJ, Sanz P, Alem F, Stibitz S, Bull RL, O'Brien AD. 2009. Four superoxide dismutases contribute to *Bacillus anthracis* virulence and provide spores with redundant protection from oxidative stress. *Infect Immun.* 77(1):274–85.
69. Damo SM, Kehl-Fie TE, Sugitani N, Holt ME, Rath S, et al. 2013. Molecular basis for manganese sequestration by calprotectin and roles in the innate immune response to invading bacterial pathogens. *Proceedings of the National Academy of Sciences.* 110(10):3841–46.
70. de Haas CJC, Veldkamp KE, Peschel A, Weerkamp F, Van Wamel WJB, et al. 2004. Chemotaxis inhibitory protein of *Staphylococcus aureus*, a bacterial antiinflammatory agent. *Journal of Experimental Medicine.* 199(5):687–95.
71. del Río LA, Lyon DS, Olah I, Glick B, Salin ML. 1983. Immunocytochemical evidence for a peroxisomal localization of manganese superoxide dismutase in leaf protoplasts from a higher plant. *Planta.* 158(3):216–24.
72. Diaz-Ochoa VE, Lam D, Lee CS, Klaus S, Behnsen J, et al. 2016. *Salmonella* mitigates oxidative stress and thrives in the inflamed gut by evading calprotectin-mediated manganese sequestration. *Cell Host and Microbe.* 19(6):814–25.
73. D'Orazio M, Folcarelli S, Mariani F, Colizzi V, Rotilio G, Battistoni A. 2001. Lipid modification of the Cu, Zn superoxide dismutase from *Mycobacterium tuberculosis*. *Biochemical Journal.* 359(1):17–22.
74. D'Orazio M, Mastropasqua MC, Cerasi M, Pacello F, Consalvo A, et al. 2015. The capability of *Pseudomonas aeruginosa* to recruit zinc under conditions of limited metal availability is affected by inactivation of the ZnuABC transporter. *Metallomics.* 7(6):1023–35.
75. D'Orazio M, Scotti R, Nicolini L, Cervoni L, Rotilio G, et al. 2008. Regulatory and structural properties differentiating the chromosomal and the bacteriophage-associated *Escherichia coli* O157:H7 Cu, Zn superoxide dismutases. *BMC Microbiol.* 8(1):166.
76. Dryla A, Gelbmann D, Von Gabain A, Nagy E. 2003. Identification of a novel iron regulated staphylococcal surface protein with haptoglobin-haemoglobin binding activity: Staphylococcal haptoglobin receptor. *Molecular Microbiology.* 49(1):37–53.

77. Dufernez F, Yernaux C, Gerbod D, Noël C, Chauvenet M, et al. 2006. The presence of four iron-containing superoxide dismutase isozymes in Trypanosomatidae: Characterization, subcellular localization, and phylogenetic origin in *Trypanosoma brucei*. Free Radical Biology and Medicine. 40(2):210–25.
78. DuMont AL, Nygaard TK, Watkins RL, Smith A, Kozhaya L, et al. 2011. Characterization of a new cytotoxin that contributes to *Staphylococcus aureus* pathogenesis: The cytotoxin LukAB kills human phagocytes. Molecular Microbiology. 79(3):814–25.
79. Dunman PM, Murphy E, Haney S, Palacios D, Tucker-Kellogg G, et al. 2001. Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci. J Bacteriol. 183(24):7341–53.
80. Dupont CL, Butcher A, Valas RE, Bourne PE, Caetano-Anolles G. 2010. History of biological metal utilization inferred through phylogenomic analysis of protein structures. Proceedings of the National Academy of Sciences. 107(23):10567–72.
81. Dupont CL, Neupane K, Shearer J, Palenik B. 2008. Diversity, function and evolution of genes coding for putative Ni-containing superoxide dismutases. Environ Microbiol. 10(7):1831–43.
82. Dussurget O, Stewart G, Neyrolles O, Pescher P, Young D, Marchal G. 2001. Role of *Mycobacterium tuberculosis* copper-zinc superoxide dismutase. Infect Immun. 69(1):529–33.
83. Edelsberg J, Taneja C, Zervos M, Haque N, Moore C, et al. 2009. Trends in US hospital admissions for skin and soft tissue infections. Emerg. Infect. Dis. 15(9):1516–18.
84. Edwards KM, Cynamon MH, Voladri RKR, Hager CC, DeStefano MS, et al. 2001. Iron-cofactored superoxide dismutase inhibits host responses to *Mycobacterium tuberculosis*. Am J Respir Crit Care Med. 164(12):2213–19.
85. Edwards RA, Baker HM, Whittaker MM, Whittaker JW, Jameson GB, Baker EN. 1998. Crystal structure of *Escherichia coli* manganese superoxide dismutase at 2.1-Å resolution. J Biol Inorg Chem. 3(2):161–71.
86. Eijkelkamp BA, Morey JR, Ween MP, Ong CLY, McEwan AG, et al. 2014. Extracellular zinc competitively inhibits manganese uptake and compromises oxidative stress management in *Streptococcus pneumoniae*. PLoS ONE. 9(2).

87. Esteve-Gassent MD, Elliott NL, Seshu J. 2009. *sodA* is essential for virulence of *Borrelia burgdorferi* in the murine model of Lyme disease. *Molecular Microbiology*. 71(3):594–612.
88. Faller LD, Johnson AM. 1974. Direct measurement of proton release by yeast enolase upon binding magnesium ions. *FEBS Letters*. 44(3):298–301.
89. Fang FC, DeGroote MA, Foster JW, Baumler AJ, Ochsner U, et al. 1999. Virulent *Salmonella typhimurium* has two periplasmic Cu, Zn-superoxide dismutases. *Proceedings of the National Academy of Sciences*. 96(13):7502–7.
90. Fenlon LA, Slauch JM. 2014. Phagocyte roulette in *Salmonella* killing. *Cell Host & Microbe*. 15(1):7–8.
91. Figueroa-Bossi N, Bossi L. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. *Mol Microbiol*. 33(1):167–76.
92. Forsgren A, Nordström K. 1974. Protein A from *Staphylococcus aureus*: The biological significance of its reaction with IgG. *Ann NY Acad Sci*. 236(1 Recent Advanc):252–66.
93. Foster AW, Osman D, Robinson NJ. 2014. Metal preferences and metallation. *Journal of Biological Chemistry*. 289(41):28095–103.
94. Foster JAL SJ. 1999. Interactive regulatory pathways control virulence determinant production and stability in response to environmental conditions in *Staphylococcus aureus*. 1:323–31.
95. Foster T. 1996. *Staphylococcus*. In *Medical Microbiology*, ed S Baron. Galveston (TX): University of Texas Medical Branch at Galveston. 4th ed.
96. Foster T. 1998. Surface protein adhesins of *Staphylococcus aureus*. *Trends in Microbiology*. 6(12):484–88.
97. Franzon VL, Arondel J, Sansonetti PJ. 1990. Contribution of superoxide dismutase and catalase activities to *Shigella flexneri* pathogenesis. *Infect Immun*. 58(2):529–35.
98. Fukai T. 2002. Extracellular superoxide dismutase and cardiovascular disease. *Cardiovascular Research*. 55(2):239–49.
99. Fukai T, Ushio-Fukai M. 2011. Superoxide dismutases: Role in redox signaling, vascular function, and diseases. *Antioxidants & Redox Signaling*. 15(6):1583–1606.

100. Gao Q, Xia L, Wang X, Ye Z, Liu J, Gao S. 2019. SodA contributes to the virulence of avian pathogenic *Escherichia coli* O2 Strain E058 in experimentally infected chickens. *J Bacteriol.* 201(6).
101. Garcia YM, Barwinska-Sendra A, Tarrant E, Skaar EP, Waldron KJ, Kehl-Fie TE. 2017. A Superoxide dismutase capable of functioning with iron or manganese promotes the resistance of *Staphylococcus aureus* to calprotectin and nutritional immunity. *PLOS Pathogens.* 13(1):e1006125.
102. Gebhardt C, Németh J, Angel P, Hess J. 2006. S100A8 and S100A9 in inflammation and cancer. *Biochemical Pharmacology.* 72(11):1622–31.
103. Gertz S, Engelmann S, Schmid R, Ziebandt AK, Tischer K, et al. 2000. Characterization of the σ B regulon in *Staphylococcus aureus*. *Journal of Bacteriology.* 182(24):6983–91.
104. Ghssein G, Brutesco C, Ouerdane L, Fojcik C, Izaute A, et al. 2016. Biosynthesis of a broad-spectrum nicotianamine-like metallophore in *Staphylococcus aureus*. *Science.* 352(6289):1105–9.
105. Gilston BA, Skaar EP, Chazin WJ. 2016. Binding of transition metals to S100 proteins. *Sci. China Life Sci.* 59(8):792–801.
106. Gläser R, Harder J, Lange H, Bartels J, Christophers E, Schröder J-M. 2005. Antimicrobial psoriasin (S100A7) protects human skin from *Escherichia coli* infection. *Nat Immunol.* 6(1):57–64.
107. Gleason JE, Galaleldeen A, Peterson RL, Taylor AB, Holloway SP, et al. 2014. *Candida albicans* SOD5 represents the prototype of an unprecedented class of Cu-only superoxide dismutases required for pathogen defense. *Proceedings of the National Academy of Sciences.* 111(16):5866–71.
108. Gordon RJ, Lowy FD. 2008. Pathogenesis of Methicillin-Resistant *Staphylococcus aureus* Infection. *CLIN INFECT DIS.* 46(S5):S350–59.
109. Grace SC. 1990. Phylogenetic distribution of superoxide dismutase supports an endosymbiotic origin for chloroplasts and mitochondria. *Life Sciences.* 47(21):1875–86.
110. Grass G, Franke S, Taudte N, Nies DH, Kucharski LM, et al. 2005. The metal permease ZupT from *Escherichia coli* is a transporter with a broad substrate spectrum. *J Bacteriol.* 187(5):1604–11.
111. Gregory EM. 1985. Characterization of the O₂-induced manganese-containing superoxide dismutase from *Bacteroides fragilis*. *Archives of Biochemistry and Biophysics.* 238(1):83–89.

112. Grim KP, Radin JN, Solórzano PKP, Morey JR, Frye KA, et al. 2020. Intracellular accumulation of staphylopine can sensitize *Staphylococcus aureus* to host-imposed zinc starvation by chelation-independent toxicity. *J Bacteriol.* 202(9):e00014-20.
113. Grim KP, San Francisco B, Radin JN, Brazel EB, Kelliher JL, et al. 2017. The metallophore staphylopine enables *Staphylococcus aureus* to compete with the host for zinc and overcome nutritional immunity. *mBio.* 8(5).
114. Grove LE, Xie J, Yikilmaz E, Miller A-F, Brunold TC. 2008. Spectroscopic and computational investigation of second-sphere contributions to redox tuning in *Escherichia coli* iron superoxide dismutase. *Inorg. Chem.* 47(10):3978–92.
115. Gu M, Imlay JA. 2013. Superoxide poisons mononuclear iron enzymes by causing mismetallation. *Molecular Microbiology.* 89(1):123–34.
116. Haag AF, Bagnoli F. 2015. The role of two-component signal transduction systems in *Staphylococcus aureus* virulence regulation. In *Staphylococcus aureus*, ed F Bagnoli, R Rappuoli, G Grandi. 409:145–98. Cham: Springer International Publishing.
117. Hacker RG, Rachid S, Ohlsen K, Wallner U, Hecker M, Ziebuhr W. 2000. Alternative transcription factor σ B is involved in regulation of biofilm expression in a *Staphylococcus aureus* mucosal isolate. 182(23):6824–26.
118. Haley KP, Skaar EP. 2012. A battle for iron: host sequestration and *Staphylococcus aureus* acquisition. *Microbes and Infection.* 14(3):217–27.
119. Hammer ND, Skaar EP. 2011. Molecular mechanisms of *Staphylococcus aureus* iron acquisition. *Annu. Rev. Microbiol.* 65(1):129–47.
120. Hammer ND, Skaar EP. 2012. The impact of metal sequestration on *Staphylococcus aureus* metabolism. *Current Opinion in Microbiology.* 15(1):10–14.
121. Hampton MB, Kettle AJ, Winterbourn CC. 1996. Involvement of superoxide and myeloperoxidase in oxygen-dependent killing of *Staphylococcus aureus* by neutrophils. *Infect Immun.* 64(9):3512–17.
122. Handke LD, Gribenko AV, Timofeyeva Y, Scully IL, Anderson AS. 2018. MntC-dependent manganese transport is essential for *Staphylococcus aureus* oxidative stress resistance and virulence. *msphere.* 3(4):1–16.
123. Hayakawa T, Kanematsu S, Asada K. 1985. Purification and characterization of thylakoid-bound Mn-superoxide dismutase in spinach chloroplasts. *Planta.* 166(1):111–16.

124. Hayashi T. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Research*. 8(1):11–22.
125. Hayden JA, Brophy MB, Cunden LS, Nolan EM. 2013. High-affinity manganese coordination by human calprotectin is calcium-dependent and requires the histidine-rich site formed at the dimer interface. *Journal of the American Chemical Society*. 135(2):775–87.
126. Hentze MW, Muckenthaler MU, Andrews NC. 2004. Balancing acts: molecular control of mammalian iron metabolism. *Cell*. 117(3):285–97.
127. Herbst RW, Guce A, Bryngelson PA, Higgins KA, Ryan KC, et al. 2009. Role of conserved tyrosine residues in NiSOD catalysis: a case of convergent evolution. *Biochemistry*. 48(15):3354–69.
128. Hermans D, Pasmans F, Messens W, Martel A, Van Immerseel F, et al. 2012. Poultry as a host for the zoonotic pathogen *Campylobacter jejuni*. *Vector-Borne and Zoonotic Diseases*. 12(2):89–98.
129. Ho TD, Figueroa-Bossi N, Wang M, Uzzau S, Bossi L, Slauch JM. 2002. Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of *Salmonella enterica* serovar *Typhimurium*. *J Bacteriol*. 184(19):5234–39.
130. Ho TD, Slauch JM. 2001. Characterization of grvA, an antivirulence gene on the Gifsy-2 phage in *Salmonella enterica* serovar *Typhimurium*. *J Bacteriol*. 183(2):611–20.
131. Holley AK, Dhar SK, St. Clair DK. 2013. Curbing cancer's sweet tooth: Is there a role for MnSOD in regulation of the Warburg effect? *Mitochondrion*. 13(3):170–88.
132. Hood MI, Mortensen BL, Moore JL, Zhang Y, Kehl-Fie TE, et al. 2012. Identification of an *Acinetobacter baumannii* zinc acquisition system that facilitates resistance to calprotectin-mediated zinc sequestration. *PLoS Pathog*. 8(12):e1003068.
133. Hood MI, Skaar EP. 2012. Nutritional immunity: Transition metals at the pathogen-host interface. *Nature Reviews Microbiology*. 10(8):525–37.
134. Horsburgh MJ, Ingham E, Foster SJ. 2001. In *Staphylococcus aureus*, *fur* is an interactive regulator with PerR, contributes to virulence, and is necessary for oxidative stress resistance through positive regulation of catalase and iron homeostasis. *J Bacteriol*. 183(2):468–75.

135. Horsburgh MJ, Wharton SJ, Cox AG, Ingham E, Peacock S, Foster SJ. 2002. MntR modulates expression of the PerR regulon and superoxide resistance in *Staphylococcus aureus* through control of manganese uptake: Manganese uptake and oxidative stress in *S. aureus*. *Molecular Microbiology*. 44(5):1269–86.
136. Ikeda JS, Janakiraman A, Kehres DG, Maguire ME, Slauch JM. 2005. Transcriptional regulation of *sitABCD* of *Salmonella enterica* Serovar Typhimurium by MntR and Fur. *J Bacteriol*. 187(3):912–22.
137. Imlay JA. 2003. Pathways of oxidative damage. *Annual Review of Microbiology*. 57(1):395–418.
138. Imlay JA. 2008. Cellular defenses against superoxide and hydrogen peroxide. *Annual Review of Biochemistry*. 77(1):755–76.
139. Imlay JA. 2014. The mismetallation of enzymes during oxidative stress. *Journal of Biological Chemistry*. 289(41):28121–28.
140. Imlay JA. 2015. Common mechanisms of bacterial metal homeostasis. In *Trace Metals and Infectious Diseases*, ed JO Nriagu, EP Skaar. Cambridge (MA): MIT Press.
141. Inaoka T, Matsumura Y, Tsuchido T. 1998. Molecular cloning and nucleotide sequence of the superoxide dismutase gene and characterization of its product from *Bacillus subtilis*. *J Bacteriol*. 180(14):3697–3703.
142. Irving H, Williams RJP. 1948. Order of stability of metal complexes. *Nature*. 162(4123):746–47.
143. Jabado N, Jankowski A, Dougaparsad S, Picard V, Grinstein S, Gros P. 2000. Natural resistance to intracellular infections. *Journal of Experimental Medicine*. 192(9):1237–48.
144. Jackson C, Dench J, Moore AL, Halliwell B, Foyer CH, Hall DO. 1978. Subcellular localisation and identification of superoxide dismutase in the leaves of higher plants. *Eur J Biochem*. 91(2):339–44.
145. Jakubovics NS, Jenkinson HF. 2001. Out of the iron age: new insights into the critical role of manganese homeostasis in bacteria. *Microbiology*. 147(7):1709–18.
146. Jenul C, Horswill AR. 2019. Regulation of *Staphylococcus aureus* virulence. *Microbiology Spectrum*. 6(1).

147. Ji H-F, Chen L, Jiang Y-Y, Zhang H-Y. 2009. Evolutionary formation of new protein folds is linked to metallic cofactor recruitment. *BioEssays*. 31(9):975–80.
148. Johnson DC, Dean DR, Smith AD, Johnson MK. 2005. Structure, function, and formation of biological iron-sulfur clusters. *Annu. Rev. Biochem.* 74(1):247–81.
149. Jonsson I, Arvidson S, Foster S, Tarkowski A. 2004. Sigma factor B and RsbU are required for virulence in *Staphylococcus aureus*-induced arthritis and sepsis. 72(10):6106–11.
150. Jönsson K, McDevitt D, McGavin MH, Patti JM, Höök M. 1995. *Staphylococcus aureus* expresses a major histocompatibility complex class II analog. *Journal of Biological Chemistry*. 270(37):21457–60.
151. Juttukonda LJ, Skaar EP. 2015. Manganese homeostasis and utilization in pathogenic bacteria. *Molecular Microbiology*. 97(2):216–28.
152. Kale P, Dhawan B. 2016. The changing face of community-acquired methicillin-resistant *Staphylococcus aureus*. *Indian Journal of Medical Microbiology*. 34(3):275–85.
153. Kaplan J. 2002. Mechanisms of cellular iron acquisition: another iron in the fire. *Cell*. 111(5):603–6.
154. Karavolos MH, Horsburgh M, Ingham E, Foster S. 2003. Role and regulation of the superoxide dismutases of *Staphylococcus aureus*. *Microbiology*. 149(10):2749–58.
155. Keen CL, Ensunsa JL, Clegg MS. 2000. Manganese metabolism in animals and humans including the toxicity of manganese. *Met Ions Biol Syst*. 37:89–121.
156. Kehl-Fie TE, Chitayat S, Hood MI, Damo S, Restrepo N, et al. 2011. Nutrient metal sequestration by calprotectin inhibits bacterial superoxide defense, enhancing neutrophil killing of *Staphylococcus aureus*. *Cell Host and Microbe*. 10(2):158–64.
157. Kehl-Fie TE, Skaar EP. 2010. Nutritional immunity beyond iron: a role for manganese and zinc. *Current Opinion in Chemical Biology*. 14(2):218–24.
158. Kehl-Fie TE, Zhang Y, Moore JL, Farrand AJ, Hood MI, et al. 2013. MntABC and MntH contribute to systemic *Staphylococcus aureus* infection by competing with calprotectin for nutrient manganese. *Infection and Immunity*. 81(9):3395–3405.

159. Kehres DG, Janakiraman A, Slauch JM, Maguire ME. 2002. Regulation of *Salmonella enterica* Serovar Typhimurium *mntH* Transcription by H₂O₂, Fe²⁺, and Mn²⁺. *J Bacteriol.* 184(12):3151–58.
160. Khademian M, Imlay JA. 2021. How microbes evolved to tolerate oxygen. *Trends in Microbiology.* 29(5):428–40.
161. Kirby K, Hu J, Hilliker AJ, Phillips JP. 2002. RNA interference-mediated silencing of Sod2 in *Drosophila* leads to early adult-onset mortality and elevated endogenous oxidative stress. *Proceedings of the National Academy of Sciences.* 99(25):16162–67.
162. Kirschvink JL, Gaidos EJ, Bertani LE, Beukes NJ, Gutzmer J, et al. 2000. Paleoproterozoic snowball Earth: Extreme climatic and geochemical global change and its biological consequences. *Proceedings of the National Academy of Sciences.* 97(4):1400–1405.
163. Kluytmans JAJW, Wertheim HFL. 2005. Nasal carriage of *Staphylococcus aureus* and prevention of nosocomial infections. *Infection.* 33(1):3–8.
164. Korndörfer IP, Brueckner F, Skerra A. 2007. The Crystal structure of the human (S100A8/S100A9) heterotetramer, calprotectin, illustrates how conformational changes of interacting α -helices can determine specific association of two EF-hand proteins. *Journal of Molecular Biology.* 370(5):887–98.
165. Korshunov SS, Imlay JA. 2002. A potential role for periplasmic superoxide dismutase in blocking the penetration of external superoxide into the cytosol of Gram-negative bacteria. *Mol Microbiol.* 43(1):95–106.
166. Krishnakumar R, Craig M, Imlay JA, Slauch JM. 2004. Differences in enzymatic properties allow SodCI but not SodCII to contribute to virulence in *Salmonella enterica* serovar *Typhimurium* strain 14028. *J Bacteriol.* 186(16):5230–38.
167. Krishnakumar R, Kim B, Mollo EA, Imlay JA, Slauch JM. 2007. Structural properties of periplasmic SodCI that correlate with virulence in *Salmonella enterica* serovar *Typhimurium*. *J Bacteriol.* 189(12):4343–52.
168. Kusch H, Engelmann S. 2014. Secrets of the secretome in *Staphylococcus aureus*. *International Journal of Medical Microbiology.* 304(2):133–41.
169. Lalaouna D, Baude J, Wu Z, Tomasini A, Chicher J, et al. 2019. RsaC sRNA modulates the oxidative stress response of *Staphylococcus aureus* during manganese starvation. *Nucleic Acids Research.* 47(18):9871–87.

170. Laux C, Peschel A, Krismer B. 2019. *Staphylococcus aureus* colonization of the human nose and interaction with other microbiome members. In Gram-Positive Pathogens, ed VA Fischetti, RP Novick, JJ Ferretti, DA Portnoy, M Braunstein, JI Rood, pp. 723–30. Washington, DC, USA: ASM Press.
171. Lebovitz RM, Zhang H, Vogel H, Cartwright J, Dionne L, et al. 1996. Neurodegeneration, myocardial injury, and perinatal death in mitochondrial superoxide dismutase-deficient mice. *Proceedings of the National Academy of Sciences*. 93(18):9782–87.
172. Lee LY, Miyamoto YJ, McIntyre BW, Höök M, McCrea KW, et al. 2002. The *Staphylococcus aureus* Map protein is an immunomodulator that interferes with T cell-mediated responses. *J. Clin. Invest.* 110(10):1461–71.
173. Lew DP, Waldvogel FA. 2004. Osteomyelitis. *The Lancet*. 364(9431):369–79.
174. Li CX, Gleason JE, Zhang SX, Bruno VM, Cormack BP, Culotta VC. 2015. *Candida albicans* adapts to host copper during infection by swapping metal cofactors for superoxide dismutase. *Proc Natl Acad Sci USA*. 112(38):E5336–42.
175. Li Y, Huang T-T, Carlson EJ, Melov S, Ursell PC, et al. 1995. Dilated cardiomyopathy and neonatal lethality in mutant mice lacking manganese superoxide dismutase. *Nat Genet*. 11(4):376–81.
176. Liu GY. 2009. Molecular pathogenesis of *Staphylococcus aureus* infection. *Pediatr Res*. 65(5 Part 2):71R-77R.
177. Liu H, Bergman NH, Thomason B, Shallom S, Hazen A, et al. 2004. Formation and composition of the *Bacillus anthracis* endospore. *J Bacteriol*. 186(1):164–78.
178. Lonergan ZR, Skaar EP. 2019. Nutrient zinc at the host–pathogen interface. *Trends in Biochemical Sciences*. 44(12):1041–56.
179. Lowy F. 1998. *Staphylococcus aureus* infections. *New England Journal of Medicine*. 339:520–32.
180. Lynch M, Kuramitsu H. 2000. Expression and role of superoxide dismutases (SOD) in pathogenic bacteria. *Microbes and infection / Institut Pasteur*. 2(10):1245–55.
181. Lynch MC, Kuramitsu HK. 1999. Role of superoxide dismutase activity in the physiology of *Porphyromonas gingivalis*. *Infect Immun*. 67(7):3367–75.

182. Lyons TW, Reinhard CT, Planavsky NJ. 2014. The rise of oxygen in Earth's early ocean and atmosphere. *Nature*. 506(7488):307–15.
183. Macomber L, Imlay JA. 2009. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc. Natl. Acad. Sci. U.S.A.* 106(20):8344–49.
184. Malachowa N, Whitney AR, Kobayashi SD, Sturdevant DE, Kennedy AD, et al. 2011. Global changes in *Staphylococcus aureus* gene expression in human blood. *PLoS One*. 6(4):e18617.
185. Martchenko M, Alarco A-M, Harcus D, Whiteway M. 2004. Superoxide dismutases in *Candida albicans*: transcriptional regulation and functional characterization of the hyphal-induced SOD5 gene. *Mol Biol Cell*. 15(2):456–67.
186. Martin ME, Byers BR, Olson MOJ, Salin ML, Arceneaux JE, Tolbert C. 1986. A *Streptococcus mutans* superoxide dismutase that is active with either manganese or iron as a cofactor. *Journal of Biological Chemistry*. 261(20):9361–67.
187. Martínez A, Prolo C, Estrada D, Rios N, Alvarez MN, et al. 2019. Cytosolic Fe-superoxide dismutase safeguards *Trypanosoma cruzi* from macrophage-derived superoxide radical. *Proc Natl Acad Sci USA*. 116(18):8879–88.
188. Masse E, Gottesman S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences*. 99(7):4620–25.
189. Massey V, Strickland S, Mayhew SG, Howell LG, Engel PC, et al. 1969. The production of superoxide anion radicals in the reaction of reduced flavins and flavoproteins with molecular oxygen. *Biochemical and Biophysical Research Communications*. 36(6):891–97.
190. Matsuo M, Kato F, Oogai Y, Kawai T, Sugai M, Komatsuzawa H. 2010. Distinct two-component systems in methicillin-resistant *Staphylococcus aureus* can change the susceptibility to antimicrobial agents. *J Antimicrob Chemother*. 65(7):1536–37.
191. Mazmanian SK, Skaar EP, Gaspar AH, Humayun M, Gornicki P, et al. 2003. Passage of heme-iron across the envelope of *Staphylococcus aureus*. *Science*. 299(5608):906–9.
192. McCord JM, Fridovich I. 1969. Superoxide dismutase: An enzymic function for erythrocyte hemocuprein (hemocuprein). *Journal of Biological Chemistry*. 244(22):6049–55.

193. Meier B, Barra D, Bossa F, Calabrese L, Rotilio G. 1982. Synthesis of either Fe- or Mn-superoxide dismutase with an apparently identical protein moiety by an anaerobic bacterium dependent on the metal supplied. *Journal of Biological Chemistry*. 257(23):13977–80.
194. Menestrina G, Dalla Serra M, Prévost G. 2001. Mode of action of β -barrel pore-forming toxins of the staphylococcal α -hemolysin family. *Toxicon*. 39(11):1661–72.
195. Messner KR, Imlay JA. 1999. The identification of primary sites of superoxide and hydrogen peroxide formation in the aerobic respiratory chain and sulfite reductase complex of *Escherichia coli*. *Journal of Biological Chemistry*. 274(15):10119–28.
196. Miao L, St. Clair DK. 2009. Regulation of superoxide dismutase genes: Implications in disease. *Free Radical Biology and Medicine*. 47(4):344–56.
197. Miller A-F. 2004. Superoxide dismutases: active sites that save, but a protein that kills. *Current Opinion in Chemical Biology*. 8(2):162–68.
198. Miller AF. 2012. Superoxide dismutases: Ancient enzymes and new insights. *FEBS Letters*. 586(5):585–95.
199. Ming LJ, Valentine JS. 1990. NMR studies of nickel(II)-substituted derivatives of bovine copper-zinc superoxide dismutase with nickel(II) bound in the copper site. *J. Am. Chem. Soc.* 112(17):6374–83.
200. Misra TK. 2000. SarA represses *agr* operon expression in a purified *in vitro* *Staphylococcus aureus* transcription system. 182(20):5893–97.
201. Mittra B, Laranjeira-Silva MF, Miguel DC, Perrone Bezerra de Menezes J, Andrews NW. 2017. The iron-dependent mitochondrial superoxide dismutase SODA promotes *Leishmania* virulence. *Journal of Biological Chemistry*. 292(29):12324–38.
202. Moore EK, Jelen BI, Giovannelli D, Raanan H, Falkowski PG. 2017. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nature Geosci.* 10(9):629–36.
203. Nakashige TG, Stephan JR, Cunden LS, Brophy MB, Wommack AJ, et al. 2016. The Hexahistidine motif of host-defense protein human calprotectin contributes to zinc withholding and its functional versatility. *J. Am. Chem. Soc.* 138(37):12243–51.
204. Nakashige TG, Zhang B, Krebs C, Nolan EM. 2015. Human calprotectin is an iron-sequestering host-defense protein. *Nature Chemical Biology*. 11(10):765–71.

205. Nakayama K. 1992. Nucleotide sequence of *Streptococcus mutans* superoxide dismutase gene and isolation of insertion mutants. *J Bacteriol.* 174(15):4928–34.
206. Nakayama K. 1994. Rapid viability loss on exposure to air in a superoxide dismutase-deficient mutant of *Porphyromonas gingivalis*. *J Bacteriol.* 176(7):1939–43.
207. Netz DJA, Stith CM, Stümpfig M, Köpf G, Vogel D, et al. 2012. Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. *Nat Chem Biol.* 8(1):125–32.
208. Newsom SWB. 2008. Ogston's coccus. *Journal of Hospital Infection.* 70(4):369–72.
209. Nilsson IM, Bremell T, Rydén C, Cheung AL, Tarkowski A. 1996. Role of the staphylococcal accessory gene regulator (*sar*) in septic arthritis. *Infect Immun.* 64(11):4438–43.
210. Noble WC, Valkenburg HA, Wolters CHL. 1967. Carriage of *Staphylococcus aureus* in random samples of a normal population. *Journal of Hygiene.* 65(4):567–73.
211. Ose DE, Fridovich I. 1976. Superoxide dismutase. Reversible removal of manganese and its substitution by cobalt, nickel or zinc. *Journal of Biological Chemistry.* 251(4):1217–18.
212. Palazzolo-Ballance AM, Reniere ML, Braughton KR, Sturdevant DE, Otto M, et al. 2008. Neutrophil microbicides induce a pathogen survival response in community-associated methicillin-resistant *Staphylococcus aureus*. *J Immunol.* 180(1):500–509.
213. Palmer LD, Skaar EP. 2016. Transition metals and virulence in bacteria. *Annual Review of Genetics.* 50(1):67–91.
214. Papp-Wallace KM, Maguire ME. 2006. Manganese transport and the role of manganese in virulence. *Annu Rev Microbiol.* 60:187–209.
215. Peers G, Price NM. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature.* 441(7091):341–44.
216. Pennington CD, Gregory EM. 1986. Isolation and reconstitution of iron- and manganese-containing superoxide dismutases from *Bacteroides thetaiotaomicron*. *Journal of Bacteriology.* 166(2):528–32.
217. Peracino B, Wagner C, Balest A, Balbo A, Pergolizzi B, et al. 2006. Function and mechanism of action of dictyostelium Nrap1 (Slc11a1) in bacterial infection: Nrap1 in dictyostelium. *Traffic.* 7(1):22–38.

218. Peraro MD, van der Goot FG. 2016. Pore-forming toxins: ancient, but never really out of fashion. *Nat Rev Microbiol.* 14(2):77–92.
219. Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, et al. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature.* 409(6819):529–33.
220. Perry WJ, Spraggins JM, Sheldon JR, Grunenwald CM, Heinrichs DE, et al. 2019. *Staphylococcus aureus* exhibits heterogeneous siderophore production within the vertebrate host. *Proc Natl Acad Sci U S A.* 116(44):21980–82.
221. Pesci EC, Cottle DL, Pickett CL. 1994. Genetic, enzymatic, and pathogenic studies of the iron superoxide dismutase of *Campylobacter jejuni*. *Infect Immun.* 62(7):2687–94.
222. Phillips JP, Campbell SD, Michaud D, Charbonneau M, Hilliker AJ. 1989. Null mutation of copper/zinc superoxide dismutase in *Drosophila* confers hypersensitivity to paraquat and reduced longevity. *Proceedings of the National Academy of Sciences.* 86(8):2761–65.
223. Phillips JP, Tainer JA, Getzoff ED, Boulianne GL, Kirby K, Hilliker AJ. 1995. Subunit-destabilizing mutations in *Drosophila* copper/zinc superoxide dismutase: neuropathology and a model of dimer dysequilibrium. *Proceedings of the National Academy of Sciences.* 92(19):8574–78.
224. Piddington DL, Fang FC, Laessig T, Cooper AM, Orme IM, Buchmeier NA. 2001. Cu, Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. *Infect Immun.* 69(8):4980–87.
225. Pishchany G, Sheldon JR, Dickson CF, Alam MT, Read TD, et al. 2014. IsdB-dependent hemoglobin binding is required for acquisition of heme by *Staphylococcus aureus*. *The Journal of Infectious Diseases.* 209(11):1764–72.
226. Pugh SY, DiGuseppi JL, Fridovich I. 1984. Induction of superoxide dismutases in *Escherichia coli* by manganese and iron. *J Bacteriol.* 160(1):137–42.
227. Purdy D, Cawthraw S, Dickinson JH, Newell DG, Park SF. 1999. Generation of a superoxide dismutase (SOD)-deficient mutant of *Campylobacter coli*: Evidence for the significance of SOD in *Campylobacter* survival and colonization. *Appl Environ Microbiol.* 65(6):2540–46.

228. Quie PG, Kaplan EL, Page AR, Gruskay FL, Malawista SE. 1968. Defective polymorphonuclear-leukocyte function and chronic granulomatous disease in two female children. *N Engl J Med.* 278(18):976–80.
229. Radin JN, Kelliher JL, Párraga Solórzano PK, Kehl-Fie TE. 2016. The two-component system ArlRS and alterations in metabolism enable *Staphylococcus aureus* to resist calprotectin-induced manganese starvation. *PLoS Pathogens.* 12(11):1–23.
230. Radin JN, Kelliher JL, Solórzano PKP, Grim KP, Ramezanifard R, et al. 2019. Metal-independent variants of phosphoglycerate mutase promote resistance to nutritional immunity and retention of glycolysis during infection. *PLoS Pathog.* 15(7):e1007971
231. Radin JN, Zhu J, Brazel EB, McDevitt CA, Kehl-Fie TE. 2019. Synergy between nutritional immunity and independent host defenses contributes to the importance of the MntABC manganese transporter during *Staphylococcus aureus* infection. *Infection and Immunity.* 87(1):1–14.
232. Rahil-Khazen R, Bolann BJ, Myking A, Ulvik RJ. 2002. Multi-element analysis of trace element levels in human autopsy tissues by using inductively coupled atomic emission spectrometry technique (ICP-AES). *J Trace Elem Med Biol.* 16(1):15–25.
233. Ratledge C. 2007. Iron metabolism and infection. *Food Nutr Bull.* 28:S515–23.
234. Ratledge C, Dover LG. 2000. Iron metabolism in pathogenic bacteria. *Annu. Rev. Microbiol.* 54(1):881–941.
235. Remy L, Carrière M, Derré-Bobillot A, Martini C, Sanguinetti M, Borezée-Durant E. 2013. The *Staphylococcus aureus* Opp1 ABC transporter imports nickel and cobalt in zinc-depleted conditions and contributes to virulence: Nickel and cobalt uptake in *Staphylococcus aureus*. *Molecular Microbiology.* 87(4):730–43.
236. Reniere ML, Skaar EP. 2008. *Staphylococcus aureus* haem oxygenases are differentially regulated by iron and haem. *Molecular Microbiology.* 69(5):1304–15.
237. Rigby KM, DeLeo FR. 2012. Neutrophils in innate host defense against *Staphylococcus aureus* infections. *Seminars in Immunopathology.* 34(2):237–59.
238. Robinett NG, Peterson RL, Culotta VC. 2018. Eukaryotic copper-only superoxide dismutases (SODs): A new class of SOD enzymes and SOD-like protein domains. *Journal of Biological Chemistry.* 293(13):4636–43.

239. Robinson NJ. 2007. A more discerning zinc exporter. *Nat Chem Biol.* 3(11):692–93.
240. Rogina B, Helfand SL. 2000. Cu, Zn superoxide dismutase deficiency accelerates the time course of an age-related marker in *Drosophila melanogaster*. *Biogerontology.* 1(2):163–69.
241. Roos D. 2016. Chronic granulomatous disease. *Br Med Bull.* 118(1):50–63.
242. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, et al. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature.* 362(6415):59–62.
243. Sader HS, Mendes RE, Jones RN, Flamm RK. 2016. Antimicrobial susceptibility patterns of community- and hospital-acquired methicillin-resistant *Staphylococcus aureus* from United States hospitals: results from the AWARE surveillance program (2012–2014). *Diagnostic Microbiology and Infectious Disease.* 86(1):76–79.
244. Sadosky AB, Wilson JW, Steinman HM, Shuman HA. 1994. The iron superoxide dismutase of *Legionella pneumophila* is essential for viability. *J Bacteriol.* 176(12):3790–99.
245. Saito MA, Sigman DM, Morel FMM. 2003. The bioinorganic chemistry of the ancient ocean: the co-evolution of cyanobacterial metal requirements and biogeochemical cycles at the Archean–Proterozoic boundary? *Inorganica Chimica Acta.* 356:308–18.
246. Sakr A, Brégeon F, Mège J-L, Rolain J-M, Blin O. 2018. *Staphylococcus aureus* nasal colonization: An update on mechanisms, epidemiology, risk factors, and subsequent infections. *Front. Microbiol.* 9:2419.
247. Sansone A, Watson PR, Wallis TS, Langford PR, Kroll JS. 2002. The role of two periplasmic copper- and zinc-cofactored superoxide dismutases in the virulence of *Salmonella choleraesuis*. *Microbiology.* 148(3):719–26.
248. Schaible UE, Kaufmann SHE. 2004. Iron and microbial infection. *Nature Reviews Microbiology.* 2(12):946–53.
249. Schatzman SS, Culotta VC. 2018. Chemical warfare at the microorganismal level: A closer look at the superoxide dismutase enzymes of pathogens. *ACS Infectious Diseases.* 4(6):893–903.
250. Seilie ES, Bubeck Wardenburg J. 2017. *Staphylococcus aureus* pore-forming toxins: The interface of pathogen and host complexity. *Seminars in Cell and Developmental Biology.* 72:101–16.

251. Sevilla F, López-Gorgé J, del Río LA. 1982. Characterization of a manganese superoxide dismutase from the higher plant *Pisum sativum*. *Plant Physiol.* 70(5):1321–26.
252. Sheldon JR, Heinrichs DE. 2018. Recent developments in understanding the iron acquisition strategies of gram positive pathogens. , pp. 592–630.
253. Sheldon JR, Marolda CL, Heinrichs DE. 2014. TCA cycle activity in *Staphylococcus aureus* is essential for iron-regulated synthesis of staphyloferrin A, but not staphyloferrin B: the benefit of a second citrate synthase. *Molecular Microbiology.* 92(4):824–39.
254. Sheng Y, Abreu IA, Cabelli DE, Maroney MJ, Miller AF, et al. 2014. Superoxide dismutases and superoxide reductases. *Chemical Reviews.* 114(7):3854–3918.
255. Skaar EP, Humayun M, Bae T, DeBord KL, Schneewind O. 2004. Iron-source preference of *Staphylococcus aureus* infections. *Science.* 305(5690):1626–28.
256. Spagnolo L, Törö I, D’Orazio M, O’Neill P, Pedersen JZ, et al. 2004. Unique features of the *sodC*-encoded superoxide dismutase from *Mycobacterium tuberculosis*, a fully functional copper-containing enzyme lacking zinc in the active site. *Journal of Biological Chemistry.* 279(32):33447–55.
257. St John G, Steinman HM. 1996. Periplasmic copper-zinc superoxide dismutase of *Legionella pneumophila*: role in stationary-phase survival. *J Bacteriol.* 178(6):1578–84.
258. Stallings WC, Powers TB, Patridge KA, Fee JA, Ludwig ML. 1983. Iron superoxide dismutase from *Escherichia coli* at 3.1-Å resolution: a structure unlike that of copper/zinc protein at both monomer and dimer levels. *Proceedings of the National Academy of Sciences.* 80(13):3884–88.
259. Steichen C, Chen P, Kearney JF, Turnbough, CL. 2003. Identification of the immunodominant protein and other proteins of the *Bacillus anthracis* exosporium. *J Bacteriol.* 185(6):1903–10.
260. Stojiljkovic I, Perkins-Balding D. 2002. Processing of heme and heme-containing proteins by bacteria. *DNA and Cell Biology.* 21(4):281–95.
261. Strzepek RF, Harrison PJ. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature.* 431(7009):689–92.
262. Sutherland KM, Ward LM, Colombero C -R., Johnston DT. 2021. Inter-domain horizontal gene transfer of nickel-binding superoxide dismutase. *Geobiology.* 19(5):450–59.

263. Tally FP, Goldin BR, Jacobus NV, Gorbach SL. 1977. Superoxide dismutase in anaerobic bacteria of clinical significance. *Infect Immun.* 16(1):20–25.
264. Tam K, Torres VJ. 2019. *Staphylococcus aureus* secreted toxins and extracellular enzymes. *Microbiol Spectr.* 7(2).
265. Thakker M, Park J-S, Carey V, Lee JC. 1998. *Staphylococcus aureus* serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. *Infect Immun.* 66(11):5183–89.
266. Tidhar A, Rushing MD, Kim B, Slauch JM. 2015. Periplasmic superoxide dismutase SodCI of *Salmonella* binds peptidoglycan to remain tethered within the periplasm: SodCI binds peptidoglycan. *Molecular Microbiology.* 97(5):832–43.
267. Tolosano E, Fagoonee S, Morello N, Vinchi F, Fiorito V. 2010. Heme scavenging and the other facets of hemopexin. *Antioxidants & Redox Signaling.* 12(2):305–20.
268. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. 2015. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin Microbiol Rev.* 28(3):603–61.
269. Torres VJ, Attia AS, Mason WJ, Hood MI, Corbin BD, et al. 2010. *Staphylococcus aureus* *fur* regulates the expression of virulence factors that contribute to the pathogenesis of Pneumonia. *Infect Immun.* 78(4):1618–28.
270. Torres VJ, Pishchany G, Humayun M, Schneewind O, Skaar EP. 2006. *Staphylococcus aureus* IsdB is a hemoglobin receptor required for heme iron utilization. *J Bacteriol.* 188(24):8421–29.
271. Tottey S, Waldron KJ, Firbank SJ, Reale B, Bessant C, et al. 2008. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature.* 455(7216):1138–42.
272. Troxell B, Xu H, Yang XF. 2012. *Borrelia burgdorferi*, a pathogen that lacks iron, encodes manganese-dependent superoxide dismutase essential for resistance to Streptonigrin. *Journal of Biological Chemistry.* 287(23):19284–93.
273. Tu WY, Pohl S, Gray J, Robinson NJ, Harwood CR, Waldron KJ. 2012. Cellular iron distribution in *Bacillus anthracis*. *Journal of Bacteriology.* 194(5):932–40.

274. Urban CF, Ermert D, Schmid M, Abu-Abed U, Goosmann C, et al. 2009. Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against *Candida albicans*. PLoS Pathogens. 5(10).
275. Valderas MW, Hart ME. 2001. Identification and characterization of a second superoxide dismutase gene (*sodM*) from *Staphylococcus aureus*. Journal of Bacteriology. 183(11):3399–3407.
276. Van Remmen H, Williams MD, Guo Z, Estlack L, Yang H, et al. 2001. Knockout mice heterozygous for Sod2 show alterations in cardiac mitochondrial function and apoptosis. American Journal of Physiology-Heart and Circulatory Physiology. 281(3):H1422–32.
277. Vance CK, Miller A-F. 1998. A simple proposal that can explain the inactivity of metal-substituted superoxide dismutases. J. Am. Chem. Soc. 120(3):461–67.
278. Vance CK, Miller A-F. 2001. Novel insights into the basis for *Escherichia coli* superoxide dismutase's metal ion specificity from Mn-substituted FeSOD and its very high E_m. Biochemistry. 40(43):13079–87.
279. Visai L, Yanagisawa N, Josefsson E, Tarkowski A, Pezzali I, et al. 2009. Immune evasion by *Staphylococcus aureus* conferred by iron-regulated surface determinant protein IsdH. Microbiology. 155(3):667–79.
280. Waldron KJ, Robinson NJ. 2009. How do bacterial cells ensure that metalloproteins get the correct metal? Nature Reviews Microbiology. 7(1):25–35.
281. Waldron KJ, Rutherford JC, Ford D, Robinson NJ. 2009. Metalloproteins and metal sensing. Nature. 460(7257):823–30.
282. Wang Y, Branicky R, Noë A, Hekimi S. 2018. Superoxide dismutases: Dual roles in controlling ROS damage and regulating ROS signaling. The Journal of Cell Biology. jcb.201708007.
283. Wang Y, Mo X, Zhang L, Wang Q. 2011. Four superoxide dismutase (isozymes) genes of *Bacillus cereus*. Ann Microbiol. 61(2):355–60.
284. Weinberg ED. 1975. Nutritional immunity: Host's attempt to withhold iron from microbial invaders. The Journal of the American Medical Association. 231(1):39–41.
285. Weinberg ED. 2009. Iron availability and infection. Biochimica et Biophysica Acta - General Subjects. 1790(7):600–605.

286. Wertheim HF, Melles DC, Vos MC, van Leeuwen W, van Belkum A, et al. 2005. The role of nasal carriage in *Staphylococcus aureus* infections. The Lancet Infectious Diseases. 5(12):751–62.
287. White MJ, Boyd JM, Horswill AR, Nauseef WM. 2014. Phosphatidylinositol-specific phospholipase C contributes to survival of *Staphylococcus aureus* USA300 in human blood and neutrophils. Infect. Immun. 82(4):1559–71.
288. Wilkinson SR, Prathalingam SR, Taylor MC, Ahmed A, Horn D, Kelly JM. 2006. Functional characterisation of the iron superoxide dismutase gene repertoire in *Trypanosoma brucei*. Free Radical Biology and Medicine. 40(2):198–209.
289. Wolz C, Po P. 2000. Agr-independent regulation of fibronectin-binding protein(s) by the regulatory locus *sar* in *Staphylococcus aureus*. 36:230–43.
290. Wright Valderas M, Gatson JW, Wreyford N, Hart ME. 2002. The superoxide dismutase gene *sodM* is unique to *Staphylococcus aureus*: Absence of *sodM* in coagulase-negative staphylococci. Journal of Bacteriology. 184(9):2465–72.
291. Xu Y, Miriyala S, Fang F, Bakthavatchalu V, Noel T, et al. 2015. Manganese superoxide dismutase deficiency triggers mitochondrial uncoupling and the Warburg effect. Oncogene. 34(32):4229–37.
292. Yang M, Cobine PA, Molik S, Naranuntarat A, Lill R, et al. 2006. The effects of mitochondrial iron homeostasis on cofactor specificity of superoxide dismutase 2. EMBO J. 25(8):1775–83.
293. Yesilkaya H, Kadioglu A, Gingles N, Alexander JE, Mitchell TJ, Andrew PW. 2000. Role of manganese-containing superoxide dismutase in oxidative stress and virulence of *Streptococcus pneumoniae*. Infect Immun. 68(5):2819–26.
294. Youn H-D, Kim E-J, Roe J-H, Hah YC, Kang S-O. 1996. A novel nickel-containing superoxide dismutase from *Streptomyces* spp. Biochemical Journal. 318(3):889–96.
295. Youseff BH, Holbrook ED, Smolnycki KA, Rappleye CA. 2012. Extracellular superoxide dismutase protects *Histoplasma* yeast cells from host-derived oxidative stress. PLoS Pathog. 8(5):e1002713.
296. Zhang J, Wang H, Huang Q, Zhang Y, Zhao L, et al. 2020. Four superoxide dismutases of *Bacillus cereus* 0–9 are non-redundant and perform different functions in diverse living conditions. World J Microbiol Biotechnol. 36(1):12.

297. Ziebandt A-K, Weber H, Rudolph J, Schmid R, Höper D, et al. 2001. Extracellular proteins of *Staphylococcus aureus* and the role of SarA and B. *Proteomics*. 1(4):480–93.
298. Zielinska AK, Beenken KE, Mrak LN, Spencer HJ, Post GR, et al. 2012. *sarA*-mediated repression of protease production plays a key role in the pathogenesis of *Staphylococcus aureus* USA300 isolates. *Mol Microbiol*. 86(5):1183–96.
299. Zygiel EM, Nolan EM. 2018. Transition metal sequestration by the host-defense protein calprotectin. *Annual Review of Biochemistry*. 87(1):621–43.

CHAPTER 2¹:

Metal availability can drive metalloenzyme neofunctionalization, but the path to altered metal dependency is not free rein

2.1 Abstract

Metalloproteins are important to all forms of life with up to half of all enzymes requiring a metal for full function. Metal utilization by metalloproteins can change over time, but the factors that govern these changes are not well understood. A protein family where these metal cofactor switches have been observed is the Fe/Mn superoxide dismutases (SODs). This protein family is evolutionarily ancient and has members in every domain of life, emphasizing its biological importance. As organisms rely on import of essential metals from the environment, changes in environmental metal availability are predicted to be a driver of metalloenzyme evolution, but this process has not been studied in systems with physiologically relevant stresses. Additionally the evolutionary path from one metal dependency to another is predicted to have occurred through single incremental changes that provided an advantage to the organism while encountering metal limitation, yet this has not been specifically demonstrated in biologically relevant systems. One example of a situation in which an organism will encounter metal limitation is during infection of a host. When pathogens infect a host, the host will limit the availability of essential nutrients, including transition metals such as manganese (Mn). The staphylococcal abscess is a

¹Caroline Vermilya contributed to Figures 2.5, 2.6, 2.13, 2.14, Jana N. Radin and Saika Hossain contributed to Figure 2.19.

prototypical example of nutritional immunity and is rendered virtually devoid of Mn. In addition to metal limitation, invading pathogens must also cope with superoxide. For this superoxide defense, pathogens use SODs; however, as SODs are metalloenzymes, they can be inactivated by nutritional immunity. The pathogen *Staphylococcus aureus* encodes two highly related SODs: SodA, which is Mn-dependent, and SodM, which can utilize either Mn or iron (Fe) (i.e. cambialistic). This cambialistic activity of SodM is necessary for *S. aureus* to cause infection as it enables the bacterium to retain SOD activity even when host-imposed Mn starvation inactivates SodA. Previous analyses indicated that modern SodM arose following duplication and subsequent divergence of *sodA*, suggesting that ancestral SodM was strictly Mn-dependent and that host-imposed Mn starvation drove evolution of the ancestral Mn-dependent enzyme to the modern cambialistic one. Due to the evolutionary relatedness and protein sequence similarity between SodA and SodM, the staphylococcal SODs enable studies of metalloenzyme evolution within a native organism experiencing physiologically relevant stresses. Previously, three residues were identified that differ between SodA and SodM, and when reciprocally mutated interconvert the metal specificity of these two enzymes *in vitro*. Here, this panel of enzymes that lie in a spectrum of metal utilization between SodA and SodM was integrated into the *S. aureus sodM* locus creating a collection of strains representing potential evolutionary intermediates between the ancestral and present-day *S. aureus* strains. We found that residue swapping at a single residue in SodA (Gly159Leu) led to an increased ability of *S. aureus* strains encoding that variant to resist Mn

limitation and oxidative stress. *S. aureus* strains encoding this single residue SodA variant also had an increased bacterial burden during mouse infections as compared to the parental strain encoding two copies of SodA. When we examined *S. aureus* strains encoding SodA variants with two residue swapping mutations (Gly159Leu-Leu160Phe) that had increased Fe-dependent activity *in vitro* as compared to the variant with one residue swapping mutation, surprisingly we found that they had a growth benefit only when a copy of native SodA was present and did not lead to increased bacterial burdens during infection. The cumulative data reveal that metal availability can act as a driver of evolution of metalloenzymes; yet while it is biochemically easy to evolve from Mn-dependency to cambialism, the biology reveals that there are complex factors affecting that path.

2.2 Introduction

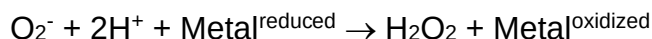
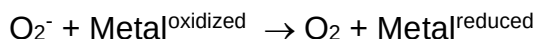
Metals are essential for all forms of life. In fact, 30% of all proteins and 50% of all enzymes require a metal to function (27, 62, 63). These metalloproteins and enzymes are involved in a variety of important cellular processes including metabolism, replication, and oxidative stress defense. Changes in metalloprotein metal utilization can occur, and it is thought that metal availability can drive these changes (2, 5, 18, 51, 58). Early organisms are thought to have been iron-philic, supported by the observation that folds in proteins binding iron (Fe) are predicted to be much older than those binding copper (Cu) and zinc (Zn) (2, 18, 31). Additionally, the anoxic environments early

organisms experienced would have meant that Cu and Zn would both have been precipitated as their sulfide minerals, leading to low concentrations of these metals (1, 35). When concentrations of atmospheric dioxygen (O_2) rose from $\sim 10^{-5}$ of present atmospheric level (PAL) to $\sim 1-10\%$ PAL ~ 2.5 billion years ago with the Great Oxygenation Event and then again to the PAL ~ 0.5 billion years ago, the levels of dissolved metals in the environment also changed (1, 7, 35, 36, 40). With the increasing O_2 , oxidized H_2S meant that the precipitated sulfide Cu and Zn minerals were released over time, leading to 14 orders of magnitude increase for Cu and 9 orders of magnitude increase for Zn (1, 35). This increase in O_2 had the opposite effect on Fe concentrations, with Fe levels dropping by four orders of magnitude due to its oxidation to Fe^{3+} which can react with OH^- to form an insoluble compound (1, 35). This decrease in the bioavailability of Fe created significant pressure for these iron-philic organisms to switch to non-Fe cofactors (18, 31, 36, 49).

Enzyme neofunctionalization, including changes in metal cofactor utilization, is thought to occur through iterative rounds of mutation and selection, with each cycle resulting in a more beneficial enzyme. While metal availability has been predicted to be a key driving force in this metalloenzyme evolution, it has not been tested for metalloenzymes in their native context due to the absence of a system that recapitulates the stresses experienced by the organism. Understanding the general principles surrounding metalloenzyme metal utilization and how metalloenzymes can alter their metal utilization is of extreme importance, not only to organismal biology, but also to biotechnology.

Metalloproteins are vital to many aspects of biotechnology and the ability to tune the metal selectivity of these enzymes could help significantly advance the field (38, 62).

Changes in metal utilization have been observed within a metalloprotein family of superoxide dismutases (SODs). These SOD enzymes are utilized as the primary mode of defense against superoxide ($O_2^{\cdot -}$) in most organisms (26, 28, 47, 55, 56). Superoxide is universally formed in organisms that experience oxygen through normal aerobic metabolism (43, 45). Superoxide is also generated during infection through the oxidative burst of immune cells (12, 13, 21, 55). The importance of these metalloenzymes is emphasized by the fact that they are present in almost all forms of life, including strict anaerobes as even they will briefly encounter some level of oxygen (10, 39). SODs can be classified into three structurally distinct protein families, each with a specific metal activity profile: some are active with nickel (Ni), some are active with Cu or Cu and Zn, and others are active with Fe, Mn, or either metal (i.e. cambialistic) (56). Each of these SOD protein families catalyze the dismutation of superoxide through the sequential oxidation and reduction of substrate molecules coupled with the sequential oxidation and reduction of the redox-active metal cofactor (56):



In order to optimize these reactions, the protein architecture of a SOD must precisely regulate the reactivity of the metal ion (56). How the protein accomplishes this becomes especially interesting for SODs in the Fe/Mn-SOD

family which optimize their different catalytic metal ions within a single protein framework (5, 44, 46, 60). This means the protein must be able to adjust both the redox couple reduction potential (E°) of free Fe ($\sim 0.77\text{V}$) and also that of Mn ($\sim 1.5\text{V}$) to that of SOD dismutation ($\sim 0.35\text{ V}$) (56). Fe/Mn-SODs coordinate their metal ions with universally conserved residues including three histidines and one aspartate sidechain (His26, His81, Asp167, and His171, using the numbering system from *Escherichia coli* Mn-SOD). Thus, the mechanism for how they optimize dismutation with either Fe or Mn within the same protein framework is not explained by their coordinating ligands (19, 56, 57). Instead, a “redox tuning” hypothesis has been proposed where residues near the metal cofactor that make no direct contacts with it (i.e., the secondary coordination sphere) tune each metal’s reduction potential to optimize catalysis (25, 61). Therefore, the overall protein architecture, and not just the central metal coordination region, is critical to ensure successful superoxide dismutation by SODs.

Changes in the environment, such as the rise in O_2 and corresponding changes in metal levels, would have selected for organisms with enzymes best adapted to those altered conditions (36). This can be seen in the Fe/Mn-superoxide dismutase (SOD) protein family, which is widely distributed across the tree of life with members found in the genomes of Eukarya, Archaea, and Bacteria (5). The eukaryotic organism, *Tetrahymena pyriformis*, encodes a SOD that is more similar to eukaryotic Mn-SODs despite its reliance on Fe (4, 11). This suggests that this SOD has undergone a substitution in metal cofactor (4, 11). Additionally, a recent study expanding the phylogenetic analyses of the

Fe/Mn-SODs found three distinct subtypes based on the presence or absence of highly conserved GGH or AAQ motifs in the SOD protein sequence (5). While the GGH subtype contains mainly Mn-SODs and the AAQ subtype contains mainly Fe-SODs, cambialistic SODs have been found in both subtypes and an Fe-SOD was found with a GGH motif (5). These analyses demonstrate that within these SOD protein clades, there are outliers that use a different metal (5).

As other metalloproteins, SODs must acquire their cognate metal from the environment. According to the Irving-Williams series, all divalent metal ions have intrinsic binding affinities for proteinaceous ligands that place them in the following order: Mg^{2+} and $\text{Ca}^{2+} < \text{Mn}^{2+} < \text{Fe}^{2+} < \text{Co}^{2+} < \text{Ni}^{2+} < \text{Cu}^{2+} > \text{Zn}^{2+}$) (30). To promote proper metalation, the bioavailability of metals is fine-tuned within the cell at concentrations inverse to that of the Irving-Williams series, i.e., Mn is more available than Fe, which is more available than Zn, etc. (14, 29, 62). Therefore, metal loading is at least partially governed by the relative bioavailability within the cell (5, 9, 42, 62). The metal loading of *E. coli* Mn-SOD has been proposed to be dictated by the availability of Mn within the cell, as Mn-SOD has extremely low metal binding selectivity between Mn or Fe (9, 16, 48, 50, 54, 64). In this way, metal acquisition is governed by both the concentration of the metal and competition against other ligands for that metal (14, 22). However, the exact role/features of a metalloprotein that contribute to metal selectivity remain unknown.

Organisms obtain essential nutrients, including metals, from their environment. During infection, pathogens must import metals within the host. The host takes advantage of this essentiality by sequestering metals in a defense termed nutritional immunity. The most well-studied example of nutritional immunity is that of Fe sequestration; however, Mn, Zn, and Cu are also withheld during infection (8, 27, 33). During infection by the pathogen *Staphylococcus aureus*, metal sequestration by CP contributes to rendering the staphylococcal abscess effectively devoid of Mn and Zn, leading to a loss of function of key metalloenzymes in the cell (15, 17, 32, 34, 52). By utilizing CP as a tool in culture, physiologically relevant metal limitation that mimics what is experienced during infection can enable studies of the effects of metal limitation.

S. aureus encodes two SODs, SodA and SodM, both in the Fe/Mn-SOD family, that differentially contribute to oxidative stress defense (24, 59). SodA is Mn-dependent while SodM is cambialistic (24). When Mn-limited, *S. aureus* switches from relying predominantly on SodA to relying on SodM, enabling the maintenance of SOD activity even when metal starved, such as during infection (24). *S. aureus* SodA and SodM are 75% identical at the amino acid level, and cambialistic SodM is more similar to Mn-dependent SodA than to any other SOD (5, 59). Additionally, all staphylococci encode SodA, but only the pathogenic *S. aureus* and two closely related species encode SodM (66). These observations suggest that *sodM* arose as first a duplication of *sodA* and then via a subsequent modification of the duplicated Mn-dependent SodA to present-day, cambialistic,

SodM (5). Thus, the staphylococcal SODs provide both an evolutionary start and endpoint enabling studies on whether small changes in metal-dependent activity can provide a biological advantage in the context of naturally occurring stressors.

Previously, secondary sphere residues in the *S. aureus* SODs that help dictate the metal activity profiles of these enzymes were identified (5). However, whether or not these small changes in metal activity profiles are sufficient to provide a benefit to the organism remained unknown. In order to begin to address this question, a suite of *S. aureus* strains encoding SodA protein variants with incremental changes in their Fe/Mn metal activity profiles was created (5). Each of these strains acts as a potential evolutionary intermediate, enabling studies into whether metal limitation can drive metalloenzyme neofunctionalization. These studies revealed that, while biochemical conversion between a Mn-dependent and cambialistic enzyme can occur, biologically, the path is constrained and increased Fe-dependency is not the only predictor of success. Additionally, these results suggest that some protein variants have altered metal selectivity, indicating that the protein itself may have some control over metal loading.

2.3 Results

2.3.1 Incremental changes in SOD metal activity provide a benefit to *S. aureus*. Previously, it was demonstrated that swapping three key amino acids between *S. aureus* SodA and SodM *in vitro* leads to significant changes in the metal activity profiles of these enzymes (5). In order to address whether changes

in SodA metal dependency lead to benefits in resisting Mn-limitation and oxidative stress, growth of a *S. aureus* strain with a duplicated copy of *sodA* in the *sodM* locus (*sodA*²), which mimics the predicted ancestral *S. aureus* strain, was assessed (Figure 2.1). Physiologically relevant Mn limitation was imposed using increasing concentrations of CP, and elevated oxidative stress was achieved by the addition of the oxidative stress generator, paraquat (PQ). In metal replete medium (0 µg/ml CP), *S. aureus sodA*² grew as well as wild type (Figures 2.2, 2.3). When Mn was restricted by the addition of CP, *S. aureus sodA*² displayed a significant growth defect as compared to wild type (Figures 2.2, 2.3, 2.4). These trends were also observed when PQ was present (Figures 2.2, 2.3, 2.4). In fact, the *sodA*² strain grew about as well as a *S. aureus* strain lacking the cambialistic SodM (Δ *sodM*) in all conditions tested (Figures 2.2, 2.3, 2.4). Together, these results demonstrate that in conditions that mimic infection, the predicted ancestral strain, *sodA*², exhibits significant growth defects as compared to present-day *S. aureus*.

By incorporating amino acid swapping mutations in the second copy of SodA at residues previously identified to dictate metal utilization, SodA metal dependency can be shifted to be more like that of SodM. In this way, *S. aureus* strains carrying these SodA protein variants act as presumptive evolutionary intermediates with incremental changes in metal activity profiles between the ancestral *S. aureus* strain with two copies of Mn-dependent SodA and the present-day *S. aureus* strain with Mn-dependent SodA and cambialistic SodM (Figure 2.1). Now that the baseline differences between present-day *S. aureus*

and the predicted ancestral *S. aureus* (*sodA*²) had been established, each of the SodA protein variants could be evaluated for whether they provide a benefit to *S. aureus* in resisting physiologically relevant stressors. When metals were available (0 µg/ml CP), all *sodA*² SodA-point mutant (PM) strains grew as well as *sodA*² and wild-type *S. aureus* (Figures 2.2, 2.3). When metals were restricted, *sodA*² SodA-PM strains encoding SodA Phe19Ile, Leu160Phe, Phe19Ile-Gly159Leu, and Phe19Ile-Leu160Phe protein variants had a significant growth defect as compared to wild-type *S. aureus* (Figures 2.2, 2.3, 2.4). *S. aureus* *sodA*² SodA-PM strains encoding SodA Gly159Leu, Gly159Leu-Leu160Phe, and Phe19Ile-Gly159Leu-Leu160Phe protein variants grew significantly better than *S. aureus* *sodA*² (Figures 2.2, 2.3, 2.4). Additionally, the growth benefits observed for these strains were also observed in media with PQ and CP (Figure 2.2). *In vitro*, these SodA protein variants (Gly159Leu, Gly159Leu-Leu160Phe, and Phe19Ile-Gly159Leu-Leu160Phe) also have increased Fe-dependent activity (Table 2.1) (5). This, combined with their improved ability to resist oxidative stress when Mn starved, is consistent with the idea that metal limitation could select for changes in metalloprotein metal utilization. To ensure that the differences in growth observed for the *S. aureus* *sodA*² and *sodA*² SodA-PM strains was not due to altered expression of the SODs, SodA and SodM expression was assessed. Utilizing transcriptional YFP reporter fusions, the expression of *sodA* and *sodM* were similar to wild type in *S. aureus* *sodA*² and most of the *sodA*² SodA-PM strains, even in strains that exhibited growth defects in response to Mn limitation (Figures 2.5, 2.6). In addition, even though

expression from the *sodM* locus in a *sodA*² SodA-PM Gly159Leu strain was slightly lower, the strains still grew significantly better than *sodA*² (Figure 2.2, 2.6). Finally, ectopic expression of *sodM* completely rescued growth of *sodA*² and each of the *sodA*² SodA-PM strains indicating that the growth defects are driven by the loss of SodM (Figure 2.7). Collectively, these results indicate that SodA variants with increased capacity to utilize Fe are beneficial to *S. aureus* under conditions of combined metal limitation and oxidative stress.

2.3.2 Absence of a native copy of SodA alters whether SodA protein variants provide a benefit to *S. aureus*. To further investigate the metal activity profiles of each of the SodA protein variants, measurement of SOD activity as well as the proportion of Fe-dependent SOD activity of these SodA protein variants was attempted within the native organism, *S. aureus*, while experiencing physiologically relevant stressors. Unfortunately, in conditions used previously to assess the growth of these strains, many of the *sodA*² SodA-PM strains exhibited poor growth, requiring the cells to be harvested at a low O.D.₆₀₀ value. This led to an appreciable level of activity associated with wild-type SodA in wild-type bacteria, therefore limiting the ability to evaluate the activity associated with the SodA protein variants (Figure 2.8). Given that small changes were expected between the protein variants, the native copy of SodA was eliminated to reduce background and facilitate quantifying changes in activity derived from the SodA protein variant encoded in the *sodM* locus. When wild-type SodA is absent (Δ *sodA* *sodA*²), the only SOD these strains have is the one in the *sodM* locus (Figure 2.1). In all conditions tested, *S. aureus* Δ *sodA* *sodA*² grew significantly

better than a SOD null mutant ($\Delta sodA \Delta sodM$) (Figures 2.9, 2.10, 2.11). When the cells were in a metal-replete environment (0 $\mu\text{g/ml}$ CP), *S. aureus* $\Delta sodA$ $sodA^2$ grew as well as a *S. aureus* strain encoding only SodM ($\Delta sodA$) (Figures 2.9, 2.10). When PQ was added, the $\Delta sodA$ $sodA^2$ strain grew slightly better than $\Delta sodA$ (Figure 2.9). When Mn was restricted, the $\Delta sodA$ $sodA^2$ strain exhibited a significant growth defect as compared to $\Delta sodA$ whether or not PQ was also present, further supporting the importance of SodM for *S. aureus* (Figures 2.9, 2.10, 2.11) (24). Ectopic complementation with *sodM* led to a complete rescue of $\Delta sodA$ $sodA^2$ growth defects in each condition tested (Figure 2.12). Therefore, we assessed both growth and SOD activity of cells expressing SodA variants in the $\Delta sodA$ background.

When assessing growth of the $\Delta sodA$ $sodA^2$ SodA-PM strains, $\Delta sodA$ $sodA^2$ SodA-PM Phe19Ile and $\Delta sodA$ $sodA^2$ SodA-PM Gly159Leu strains maintained their same growth phenotypes whether or not a wild-type copy of SodA was present in all conditions tested, i.e., $\Delta sodA$ $sodA^2$ SodA-PM Phe19Ile grew equivalent to $\Delta sodA$ $sodA^2$ whereas $\Delta sodA$ $sodA^2$ SodA-PM Gly159Leu had a growth benefit (Figures 2.2, 2.4, 2.9, 2.11). Unexpectedly, strains encoding a SodA Phe19Ile-Gly159Leu protein variant had significantly improved growth as compared to $\Delta sodA$ $sodA^2$ when the cells were treated with CP or CP and PQ. However, when wild-type SodA was present ($sodA^2$ SodA-PM Phe19Ile-Gly159Leu), no growth benefit was observed (Figures 2.2, 2.4, 2.9, 2.11). Also surprising were the growth differences for some of the $\Delta sodA$ $sodA^2$ SodA-PM strains with SodA protein variants containing Leu160Phe swapping mutations

(Figures 2.2, 2.4, 2.9, 2.11). When cells were treated with PQ, $\Delta sodA$ $sodA^2$ SodA-PM Leu160Phe and $\Delta sodA$ $sodA^2$ SodA-PM Phe19Ile-Leu160Phe grew significantly worse than the parental strain, $\Delta sodA$ $sodA^2$ (Figure 2.9 C). However, when wild-type SodA was present ($sodA^2$ SodA-PM Leu160Phe and $sodA^2$ SodA-PM Phe19Ile-Leu160Phe), these strains had equivalent growth to the parental strain, $sodA^2$ (Figure 2.2 C). *S. aureus* $\Delta sodA$ $sodA^2$ SodA-PM Phe19Ile-Leu160Phe also had a significant growth defect compared to a $\Delta sodA$ $sodA^2$ strain when cells were treated with 480 μ g/ml CP and PQ (Figure 2.9). Interestingly, *S. aureus* $\Delta sodA$ $sodA^2$ SodA-PM Gly159Leu-Leu160Phe and $\Delta sodA$ $sodA^2$ SodA-PM Phe19Ile-Gly159Leu-Leu160Phe strains grew equivalently to $\Delta sodA$ $sodA^2$ in all conditions tested (Figures 2.9, 2.11). However, when wild-type SodA was present ($sodA^2$ SodA-PM Gly159Leu-Leu160Phe and $sodA^2$ SodA-PM Phe19Ile-Gly159Leu-Leu160Phe), these strains had growth benefits when CP or CP and PQ were present (Figures 2.2, 2.4). Together, these results suggest that the Leu160Phe mutation in SodA may be detrimental to the ability of *S. aureus* to resist metal limitation and increased oxidative stress. As with the $sodA^2$ strains, ectopic complementation with *sodM* was performed and SOD transcriptional expression was assessed in $\Delta sodA$ $sodA^2$ and each of the $\Delta sodA$ $sodA^2$ SodA-PM strains. Complementation with *sodM* led to a complete rescue of growth defects for each *S. aureus* strain tested (Figure 2.12). Expression of *sodA* was slightly lower than that of wildtype in $\Delta sodA$ $sodA^2$ and the majority of the $\Delta sodA$ $sodA^2$ SodA-PM strains, while it was largely equivalent to wildtype in $\Delta sodA$ $sodA^2$ SodA-PM Gly159Leu-Leu160Phe and Phe19Ile-

Gly159Leu-Leu160Phe strains with the addition of CP (Figure 2.13). Between $\Delta sodA$ $sodA^2$ and the majority of the $\Delta sodA$ $sodA^2$ SodA-PM strains, *sodA* expression was equivalent (Figure 2.13). As there is no longer a functional SOD in the *sodA* locus, these minor differences in expression at this locus are not likely to lead to changes in growth of these strains. For *sodM* expression, $\Delta sodA$ $sodA^2$ and all of the $\Delta sodA$ $sodA^2$ SodA-PM strains had equivalent or increased *sodM* expression relative to wild type *S. aureus* (Figure 2.14). Cumulatively, as the SodA Gly159Leu-Leu160Phe and SodA Phe19Ile-Gly159Leu-Leu160Phe variants no longer provided a benefit in the absence of a native copy of SodA (Figure 2.9 B, D), this indicates that factors other than or in addition to the ability to utilize Fe influence whether a mutation in SodA is going to be beneficial to *S. aureus* when experiencing the combined effect of Mn limitation and increased oxidative stress.

2.3.3 SodA Gly159Leu and Phe19Ile-Gly159Leu swapping mutations lead to increased Fe selectivity. Now that the growth differences had been established for *S. aureus* strains encoding SodA protein variants that lacked a native copy of SodA, the metal activity profile of each of the SodA protein variants was further investigated. This was accomplished by assessing SOD activity as well as the proportion of Fe-dependent SOD activity of these SodA protein variants within the native organism, *S. aureus*, while experiencing physiologically relevant stressors. As growth differences were observed among strains carrying the SodA protein variants and these growth differences changed across conditions (+/- CP, +/- PQ), we wondered what metal these SodA protein

variants would load when *S. aureus* experiences the physiologically relevant stresses of metal limitation and increased oxidative stress. As for cells maintaining wild-type SodA (*sodA*²), due to poor growth of the strains, SOD activity was assessed following overnight growth in metal-restricted media with or without the addition of CP. In these conditions, Δ *sodA sodA*² and each of the Δ *sodA sodA*² SodA-PM strains grew mostly similar to Δ *sodA sodA*² with or without the addition of CP but, importantly, they grew to a high enough density to enable the evaluation of SOD activity in these strains (Figure 2.15). Additionally, wild-type *S. aureus* demonstrated some Mn starvation at 0 CP, as activity from both SodA and SodM was observed (Figure 2.8). When 240 µg/ml of CP was present, the cells exhibited a slight growth defect as compared to no CP, and wild-type *S. aureus* shifted to relying more on SodM for SOD activity, although some SodA activity was still observed (Figures 2.8, 2.15). The addition of increasing amounts of CP (480 µg/ml and 960 µg/ml) did not significantly alter growth of the strains, and the shift to relying on SodM for SOD activity while retaining some SodA activity was still observed (Figure 2.8, 2.15). Therefore, we focused on evaluating SOD activity in Δ *sodA sodA*² and Δ *sodA sodA*² SodA-PM strains in 0 µg/ml and 240 µg/ml of CP.

At 0 µg/ml CP, *S. aureus* Δ *sodA sodA*² and Δ *sodA sodA*² SodA-PM strains encoding SodA protein variants Phe19Ile, Gly159Leu, or Phe19Ile-Gly159Leu had similar SOD activity, while Δ *sodA sodA*² SodA-PM strains encoding SodA protein variants including Leu160Phe swapping mutations (Leu160Phe, Phe19Ile-Leu160Phe, Gly159Leu-Leu160Phe, and Phe19Ile-Gly159Leu-

Leu160Phe) had dramatically less SOD activity (Figure 2.16). For SodA protein variants including the Leu160Phe swapping mutation, 8-10x more total protein was required to obtain SOD activity values within the linear range used for quantification, further emphasizing their low levels of SOD activity. While no growth defects were observed for any of the *S. aureus* $\Delta sodA$ $sodA^2$ SodA-PM strains at 0 $\mu\text{g/ml}$ CP, the SOD activity data for SodA variants containing the Leu160Phe mutation suggest that the growth defects observed for strains encoding these variants might be explained by the substantial decrease in SOD activity observed even in the absence of CP.

At 240 $\mu\text{g/ml}$ CP, the reduction in SOD activity of $\Delta sodA$ $sodA^2$ SodA-PM strains encoding SodA protein variants including Leu160Phe swapping mutations (Leu160Phe, Phe19Ile-Leu160Phe, Gly159Leu-Leu160Phe, and Phe19Ile-Gly159Leu-Leu160Phe) was more pronounced. Even when maximum amounts of total protein were loaded (7x more than samples without Leu160Phe swapping mutations), SOD activity within the linear range was not detected (Figure 2.16). These results further suggest that the growth defects observed for $\Delta sodA$ $sodA^2$ SodA-PM strains encoding SodA protein variants including Leu160Phe swapping mutations might be explained by a substantial decrease in SOD activity observed with or without CP. For a *S. aureus* $\Delta sodA$ $sodA^2$ strain and $\Delta sodA$ $sodA^2$ SodA-PM strains encoding SodA protein variants Phe19Ile or Phe19Ile-Gly159Leu, roughly equal SOD activity was observed (Figure 2.16). Additionally, the data

suggested that the SodA Gly159Leu protein variant conferred slightly increased SOD activity; however, the difference has not yet been shown to be statistically significant (Figure 2.16).

In order to assess metal loading of the SOD protein variants, Fe-dependent SOD activity was inactivated by H₂O₂. Both in the presence and absence of CP, SodA protein variant Gly159Leu had significantly less SOD activity as compared to wild-type SodA when the samples were treated with H₂O₂ (Figure 2.16). Additionally, SodA protein variant Phe19Ile-Gly159Leu had significantly less SOD activity as compared to wild-type SodA when the samples were treated with H₂O₂ in the absence of CP (Figure 2.16). This indicates that these SodA protein variants are obtaining Fe in these conditions. Interestingly, this is occurring in an environment when wild-type SodA and SodA protein variant Phe19Ile are obtaining Mn. This suggests that there is a difference in the metal selectivity in SodA protein variants Gly159Leu and Phe19Ile-Gly159Leu and that the protein itself may be contributing to metal selection.

2.3.4 Leu160Phe mutations likely lead to SodA protein variants that are unable to properly fold or associate with an appropriate metal. The low level of SOD activity observed in $\Delta sodA$ *sodA*² SodA-PM strains with SodA protein variants containing the Leu160Phe mutation is likely due to one of two scenarios: 1) the Leu160Phe mutation is leading to protein instability, and thus degradation and absence of the protein, or 2) the protein is present, but the Leu160Phe mutation is leading to an inability to acquire the appropriate metal/properly fold. In order to begin to address this, cells were harvested as

performed for SOD activity analysis and protein abundance was quantified using LC-MS (Figure 2.17). The amount of SOD peptide pieces containing residue 160 was then compared among *S. aureus* *sodA*², *sodA*² SodA-PM Leu160Phe, Δ *sodA* *sodA*², and Δ *sodA* *sodA*² SodA-PM Leu160Phe strains (Figure 2.18). While there is an apparent difference in total SOD levels between *S. aureus* *sodA*² and *sodA*² SodA-PM Leu160Phe, no statistical significance is observed. Analysis of the individual data points revealed that, while the three replicates for *sodA*² do trend higher than *sodA*² SodA-PM Leu160Phe, there is one *sodA*² data point that is much higher than the other two. Regardless of whether wild-type SodA was present, the quantity of peptide pieces containing the Leu160Phe mutation was equivalent (Figure 2.18). This demonstrates that the difference in SOD activity observed in *S. aureus* Δ *sodA* *sodA*² SodA-PM strains encoding SodA protein variants including the Leu160Phe swapping mutation is not due to protein instability. Therefore, this reduction in SOD activity is most likely explained by an inability to acquire the appropriate metal/properly fold that is caused by the Leu160Phe mutation.

2.3.5 The path to altered metal dependency is restricted in the host.

S. aureus strains encoding SodA protein variant Gly159Leu led to a growth benefit in the presence of Mn limitation and oxidative stress with or without a wild-type copy of SodA (Figures 2.2, 2.9). However, *S. aureus* strains encoding SodA protein variant Gly159Leu-Leu160Phe only led to a growth benefit in the presence of a wild-type copy of SodA (Figures 2.2, 2.9). As *S. aureus* *sodA*² grew just about as well as Δ *sodM* during in culture growth in the presence of Mn

limitation and oxidative stress, we predicted that the *sodA*² strain would also have bacterial burdens similar to that of Δ *sodM* during infection. If that was the case, we also wondered if *S. aureus* strains encoding SodA protein variants Gly159Leu or Gly159Leu-Leu160Phe could rescue any decrease in bacterial burdens observed for the *S. aureus* *sodA*² strain. Wild-type C57BL/6 mice were retro-orbitally infected with *S. aureus* wild type, Δ *sodM*, *sodA*², *sodA*² SodA-PM Gly159Leu, and *sodA*² SodA-PM Gly159Leu-Leu160Phe and bacterial burdens were enumerated four days post-infection. As observed previously, infection with Δ *sodM* led to a decreased bacterial burden during infection as compared to wild type (Figure 2.19) (24). Infections with *S. aureus* *sodA*² also led to decreased bacterial burdens, similar to Δ *sodM* (Figure 2.19). Infections with the *S. aureus* *sodA*² SodA-PM Gly159Leu strain led to a significant increase in bacterial burdens as compared to Δ *sodM* and *sodA*² and no significant differences in bacterial burdens as compared to wild-type *S. aureus* (Figure 2.19). This demonstrates that the SodA Gly159Leu protein variant is advantageous during infection. For the *S. aureus* strain encoding a SodA Gly159Leu-Leu160Phe protein variant, bacterial burdens less than wild type and similar to that of Δ *sodM* were observed (Figure 2.19). Even though the bacterial burdens for *S. aureus* *sodA*² SodA-PM Gly159Leu and *S. aureus* *sodA*² SodA-PM Gly159Leu-Leu160Phe are not statistically different from one another, the bacterial burdens for the *S. aureus* *sodA*² SodA-PM Gly159Leu-Leu160Phe strain appear to be intermediate between that of *sodA*² and *S. aureus* *sodA*² SodA-PM Gly159Leu

(Figure 2.19). Cumulatively, these results demonstrate that the path to altered metal dependency in the host is restricted, with not all SodA protein variants that have a growth benefit in culture leading to a benefit during infection.

2.4 Discussion

The Fe/Mn-SOD family are ancient enzymes. They have had to adapt to changes in the bioavailability of metals, such as what was driven by extreme shifts in atmosphere, such as the Great Oxygenation Event, or by host-restriction, such as nutritional immunity (8, 27, 33, 36, 40). While it follows that this availability of metals was likely a driving force in the evolution of metalloproteins, this has not been tested in a system that imposes true physiological stresses. The unique opportunity presented by the staphylococcal SODs was capitalized upon to study metalloenzyme neofunctionalization (5, 59). This represents the first time a study has been able to ask these questions using a known evolutionary start and end point in a native system subjected to physiologically relevant stresses. These studies revealed that regardless of whether a wild-type copy of SodA is present, single Gly159Leu swapping mutations are beneficial for *S. aureus*. Strains encoding this mutation had increased Fe-selectivity (Figure 2.16), an increased ability to withstand Mn limitation and oxidative stress (Figures 2.2, 2.9), and led to greater bacterial burdens during infection (Figure 2.19). These studies show that just a single mutation can allow metalloenzymes to gain a new metal specificity, allowing an

evolutionary pathway from one dependency to another. This further supports that metal limitation within the host could have acted as a driving force of SOD metalloenzyme evolution (23).

SodA Gly159Leu-Leu160Phe and Phe191Ile-Gly159Leu-Leu160Phe variants led to a benefit to growth of *S. aureus* when resisting Mn limitation and oxidative stress only in the presence of a wild-type copy of SodA (Figures 2.2, 2.9). This suggests that wild-type SodA might be helping to stabilize/mask the defect of these mutant SodA variants. Perhaps this need for a “helper” protein helps explain why non-*S. aureus* staphylococcal species do not encode a cambialistic SOD (5, 66). In these species, only a single SOD is present and the biochemically easy shift from Mn-dependency to cambialism might therefore be limited by the absence of a wild-type enzyme that can stabilize/mask defects from iterative rounds of mutant enzyme. In line with this, a native copy of SodA would have been present through the neofunctionalization of SodM in *S. aureus* (5). This could have stabilized any detrimental mutations structurally in a heterodimer or through providing SOD activity for the cell. Another likely constraint on metal switching of metalloenzymes is the ability of the organism to obtain the metal. In fact, metal availability in the cytosol varies widely among organisms (63). If iterative rounds of mutations lead to an enzyme that requires a metal that the organism cannot successfully acquire, then that enzyme will not be beneficial even if *in vitro* it has higher activity. Interestingly, it was previously shown that the presence of SodM in *S. aureus* coincides with the acquisition of Fe-uptake systems and transition to an opportunistic pathogen (5). This supports

that the neofunctionalization of SodA to SodM was supported by an increase in the ability of the organism to acquire Fe and also provides another rationale for why other, non-*S. aureus* staphylococci did not evolve an Fe-utilizing SOD. Broadly, it also supports that metal limitation and the ability to acquire metals can act as a drivers of metalloprotein neofunctionalization.

Most metalloproteins, including the Fe/Mn-SOD family, acquire their metals from carefully balanced cytoplasmic metal pools (6, 63). These pools ensure that the bioavailability of metals is inverse to that of the Irving Williams series, with metals with the tightest binding having the lowest concentrations and metals with the weakest binding having the highest, i.e.: Mg^{2+} and Ca^{2+} , Mn^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , then Cu^{2+} and Zn^{2+} (6, 20, 30, 62, 63). One study with Mn-SOD in *E. coli* observed that a mutation in the conserved glutamic acid residue at the dimer interface to an alanine (E170A) led to changes in metal selectivity. When mutated, the *E. coli* Mn-SOD was found to only bind Fe even when grown in conditions with excess Mn, indicating a large shift in metal selectivity of the enzyme (65). However, this change in metal selectivity was also accompanied by a complete loss of SOD activity even when selectively loaded with Mn, indicating this mutation likely had more drastic effects on the function of the SOD enzyme as a whole (65). Additionally, this mutation is not naturally occurring as it was performed in the double glutamate bridge that forms across the dimer interface and is conserved in both Mn- and Fe-SODs (65). Surprisingly, analysis of metal loading in this study demonstrated that even when wild-type SodA and the SodA Phe19Ile protein variant were acquiring Mn, the SodA Gly159Leu and SodA

Phe19Ile-Gly159Leu protein variants were acquiring Fe (Figure 2.16). While metal loading is thought to be driven largely by metal availability, the differential metal loading, and thus selectivity, observed for SodA Gly159Leu and Phe19Ile-Gly159Leu protein variants suggests that the protein itself might play a role in metal selection. These results reveal mutations in secondary sphere residues that retained SOD activity and also led to changes in metal selectivity of a metalloenzyme. This potential ability of mutations in metalloproteins to at least partially influence the metal it obtains has the possibility to influence the evolution of that metalloprotein. Not only do these results represent a shift in the paradigm for metal loading of metalloproteins, but also represent promise for synthetic biology leading to the ability to selectively tune metal loading in metalloproteins.

Cumulatively, the combined work with the suite of potential evolutionary intermediate SodA variants with incremental changes in their Fe/Mn metal activity profiles indicated that aspects of the organism's biology need to be considered when studying the pressures on metalloenzyme neofunctionalization. This work also suggests that metal loading by metalloenzymes might be driven in part by protein structure rather than solely by metal availability, urging future studies to investigate the role that secondary sphere amino acids may play in metal selectivity of SODs. Finally, while biochemically simple, the biological path from one metal dependency to another is not as straightforward.

2.5 Methods

2.5.1 Ethics statement. All animal experiments were approved by the University of Illinois Urbana-Champaign Institutional Animal Care and Use Committee (IACUC license number 18038) and performed according to NIH guidelines, the Animal Welfare Act, and U.S. federal law.

2.5.2 Bacterial strains and cloning. *E. coli* were grown at 37°C in Luria broth (LB) with shaking at 180 rpm or on Luria broth agar supplemented with 100 µg/ml of ampicillin or 10 µg/ml of trimethoprim as needed for plasmid maintenance and selection. *S. aureus* strain Newman was used and grown at 37°C in tryptic soy broth (TSB) on a roller drum or on tryptic soy agar (TSA) supplemented with 10 µg/ml of chloramphenicol or 10 µg/ml of trimethoprim as needed for plasmid maintenance and selection. All bacterial strains were stored at -80°C in growth medium containing 30% glycerol. *E. coli* DH5α were used for propagation of pAH5 (41), pKORI (3), and their derivatives, and DH5αλpir were used for propagation of pKK30 (37) and derivatives.

The Δ*sodA*, Δ*sodM*, and Δ*sodA*Δ*sodM* *S. aureus* strains and the *sodA* and *sodM* reporter constructs were previously generated (24, 41). Plasmids containing the reporter constructs in yellow fluorescent protein (YFP)-containing vector, pAH5, were electroporated into RN4220 before being transduced into their final recipient strains via Φ85 phage. For the generation of *sodA*², 1kb fragments upstream and downstream of *sodM*, as well as *sodA* from the 5' to the 3' end of the gene were amplified using the indicated primers from wild-type *S. aureus*. These fragments were then combined using NEBuilder® HiFi DNA

Assembly Master Mix into pKORI (Table 2.2, New England Biolabs) (3). The amino acid swapping mutations were introduced through site-directed mutagenesis using the indicated primers (Table 2.3). All plasmid constructs were verified by sequencing of the region inserted into the cloning site of the plasmid. *S. aureus* strains were constructed via allelic exchange in $\Delta sodM$ and $\Delta sodA\Delta sodM$ (3). The hemolytic activity of all staphylococcal strains was confirmed by plating on blood agar plates. All $sodA^2$, $sodA^2$ SodA-PM, and derivative strains were verified by sequencing of *sodA* inserted into the *sodM* locus.

For complementation experiments, *sodM* and its 5' intergenic region including the native promoter of *sodM* were amplified and cloned into the pKK30 vector using the indicated primers (Table 2.3) (37). The pKK30 plasmids were electroporated into RN4220 before being transduced into their final recipient strains via $\Phi 85$ phage. See Tables 2.2 and 2.3 for the full list of strains, plasmids, and primers used in this study.

2.5.3 Calprotectin growth assays. Calprotectin (CP) growth assays and *sodM* complementation assays were performed largely as previously described (5, 17, 24). Briefly, *S. aureus* cultures were grown for 16 hours overnight in 5 ml Chelex-treated RPMI (nRPMI, RPMI-1640 tissue culture medium with L-glutamine, supplemented with 1% casamino acids, treated with Chelex-100 resin) supplemented with 1 mM $MgCl_2$, 100 μM $CaCl_2$, and 1 μM $FeSO_4$. Overnight cultures were then diluted 1:100 in 96-well plates with 100 μl of a growth medium consisting of 38% TSB and 62% calprotectin buffer (20 mM Tris

pH 7.5, 100 mM NaCl, 3 mM CaCl₂) + 1uM ZnSO₄, FeSO₄, and MnCl₂ +/- 0.01mM paraquat (PQ) containing varying concentrations of CP as indicated. All cultures were incubated at 37°C with overnight cultures grown in a roller drum while 96-well plates grown with shaking at 180 rpm. Growth was assessed by measuring optical density at 600 nm (OD₆₀₀).

2.5.4 SOD expression analysis. Transcriptional expression from *sodA* or *sodM* was assessed similar to prior work and CP growth assays as described above, with the following changes (24). Overnight cultures were grown in 5 ml TSB before dilution into the assay medium as above. Expression of *sodA* and *sodM* were determined by measuring yellow fluorescence (excitation and emission wavelengths of 505nm and 535nm, respectively) and normalizing to OD₆₀₀ and fluorescence from a promoter-less control vector.

2.5.5 Evaluation of SOD activity. SOD activity was assessed using a gel-based nitro blue tetrazolium assay as described previously (5, 24). For evaluating SOD activity in wild-type *S. aureus* cells (Figure 2.8), cells were grown identically to as described above for calprotectin growth assays. When cells reached an OD₆₀₀ value of 0.3, cells were then harvested, resuspended in 0.5mM KPO₄ buffer at pH 7.8 with 0.1mM EDTA, and lysed via mechanical disruption before being centrifuged to remove insoluble material.

For all other evaluations of SOD activity, *S. aureus* strains were grown in 5 ml of TSB for 3 hours and then the cells were pelleted and washed twice in 1 ml of nRPMI. The washed cultures were then inoculated 1:100 into 5 ml of 38% 3x

nRPMI supplemented with 3% CAA, 1mM MgCl₂, 100uM CaCl₂, and 1uM FeSO₄ and 62% calprotectin buffer with or without CP as indicated. These cultures were grown for 16 hours overnight before growth was assessed using optical density at 600 nm (OD₆₀₀). The cells were then harvested, resuspended in 0.5mM KPO₄ buffer at pH 7.8 with 0.1mM EDTA, and lysed via mechanical disruption before being centrifuged to remove insoluble material.

For both mid-log harvested and overnight harvested cell lysates, protein concentration in each cell lysate was determined using a BCA Assay Kit (Pierce). Varying amounts of protein from each cell lysate were resolved on a 10% native polyacrylamide gel such that the SOD activity from the lysate would be within the linear range of SOD activity of purified SodA loaded with Mn. The quantity of purified SodA-Mn protein was: 0.002 µg, 0.004 µg, 0.006 µg, 0.008 µg, 0.010 µg, and 0.012 µg. After resolution, gels were incubated in buffer containing 0.05M potassium phosphate buffer, pH 7.8, 0.1mM EDTA, 0.25mM nitro blue tetrazolium chloride, and 0.025mM riboflavin. Then, gels were exposed to light for development. To evaluate the proportion of SOD activity derived from iron-loaded enzyme, prior to incubation in buffer, gels were incubated in water with or without 20mM H₂O₂ for 20 minutes. Gels were imaged using a Bio-Rad imager Universal Hood II and analyzed using Image Lab (Bio-Rad). Units of SOD activity per mg of protein were calculated by generation of a standard curve of SOD activity using SodA-Mn loaded controls.

2.5.6 Quantification of SodA peptides containing AA160. *S. aureus* strains were grown, cell lysates were harvested, and protein concentration was

assessed in a manner identical to what is described above for SOD activity analysis from overnight cultures. Once harvested, samples were flash frozen and kept at -80°C until processing to ensure protein stability.

An aliquot of 50 µg total protein was taken from each protein extract and prepared for liquid chromatography mass spectrometry (LC-MS) analysis. Proteins were first denatured with 6 M guanidine hydrochloride in 100 mM triethylammonium bicarbonate (TEAB). Protein disulfide bonds were then reduced with 10 mM Tris(2-carboxyethyl)phosphine (TCEP), and cysteine residues were alkylated with 40 mM 2-chloroacetamide (CAA). These reactions were facilitated by incubating the samples at 95°C for 15 minutes in the dark. After the samples cooled, they were diluted fourfold with 100 mM TEAB, and the proteins were digested with LysC (Wako) at a ratio of 1:100 (w/w) for 3 hours at 30°C. Trypsin (Pierce) was added at a ratio of 1:50 (w/w) for a second digestion overnight at 37°C. On the following day, the samples were acidified with 10% trifluoroacetic acid, desalted with StageTips (53), and dried in a vacuum centrifuge.

Dried peptides were suspended in 5% acetonitrile (ACN) with 0.1% formic acid (FA), and 500 ng from each sample was injected into an UltiMate 3000 RSLC nanoflow system coupled online to a Q Exactive HF-X mass spectrometer (Thermo Scientific). Peptides were separated at a flow rate of 300 nL/min using a 25 cm Acclaim PepMap 100 C18 column (2 µm particle size, 75 µm ID) and mobile phases of 0.1% FA (A) and 80% ACN in 0.1% FA (B). The reversed phase gradient ran from 5% B to 35% B over the course of 80 minutes, followed

by an increase to 50% B over 10 minutes; the column was then washed at 90% B and equilibrated again at 5% B. The mass spectrometer was run in the positive mode with full MS scans from 350 to 1500 m/z being collected at 120k resolution (50 ms max IT, 3×10^6 AGC target). The top 15 ions from each MS scan were then subjected to HCD (30 NCE), and the resulting fragment ions were analyzed. The MS2 scans were collected at 15k resolution with an AGC target of 5×10^4 and a maximum IT of 30 ms. The isolation window was set to 1.2 m/z , and the dynamic exclusion time was set to 60 s.

The raw LC-MS data was further processed with Mascot Distiller v2.8.1.1 and the in-house Mascot server v2.8.0 (Matrix Science) to provide general protein identifications. Relative quantitation of the diagnostic peptides TPILGLDVWEHAYYLLK (wild type) and TPILGFDVWEHAYYLLK (Leu160Phe SodA protein variant) was accomplished with Byonic v4.2.1 and Byologic v4.2 (Protein Metrics Inc). Settings for the Mascot searches included a peptide mass tolerance of 10 ppm and a fragment mass tolerance of 0.02 Da. A tryptic digest with a maximum of 2 missed cleavages was specified along with a fixed modification for cysteine carbamidomethylation and variable modifications of protein N-terminal acetylation, methionine oxidation, and asparagine/glutamine deamidation. Searches were done against the Uniprot *Staphylococcus aureus* proteome (2,889 entries), and the false discovery rate (FDR) was calculated using a reverse decoy database strategy.

The settings for the Byonic search were similar, except the peptide mass tolerance was 15 ppm and the fragment mass tolerance was 25 ppm; these

settings were suggested by an initial search of the data using Byonic Preview. In addition, the analysis was done using a database of only potential SOD sequences, not the whole *S. aureus* proteome. Quantitation was done with Byologic using the initial peptide identifications imported from Byonic. Extracted ion chromatogram (XIC) peak areas were calculated for each peptide using a m/z window of 20 ppm, an apex search window of 2 min, an XIC area window of 2 min, and a minimum elution time prediction score of 200. Reported XIC values were based on the +3 charge state of the peptides TPILGLDVWEHAYYLK and TPILGFDVWEHAYYLK because this charge state was reliably identified in every sample and had a higher overall intensity than its +2 counterpart. Relative quantity of SodA peptide pieces containing the amino acid 160 are reported here as XIC +3 values and labeled as “Abundance of Peptide Pieces with AA160”.

2.5.7 Animal infections. Animal infections were performed largely as described previously (15, 24, 32). Ten-week-old female C57BL/6 mice were infected retro-orbitally with 1×10^7 CFU of *S. aureus* suspended in phosphate-buffered saline (PBS). Four days post-infection, mice were sacrificed and their kidneys were homogenized before determining bacterial burden by plating for serial dilutions for CFUs.

2.6 Figures and Tables

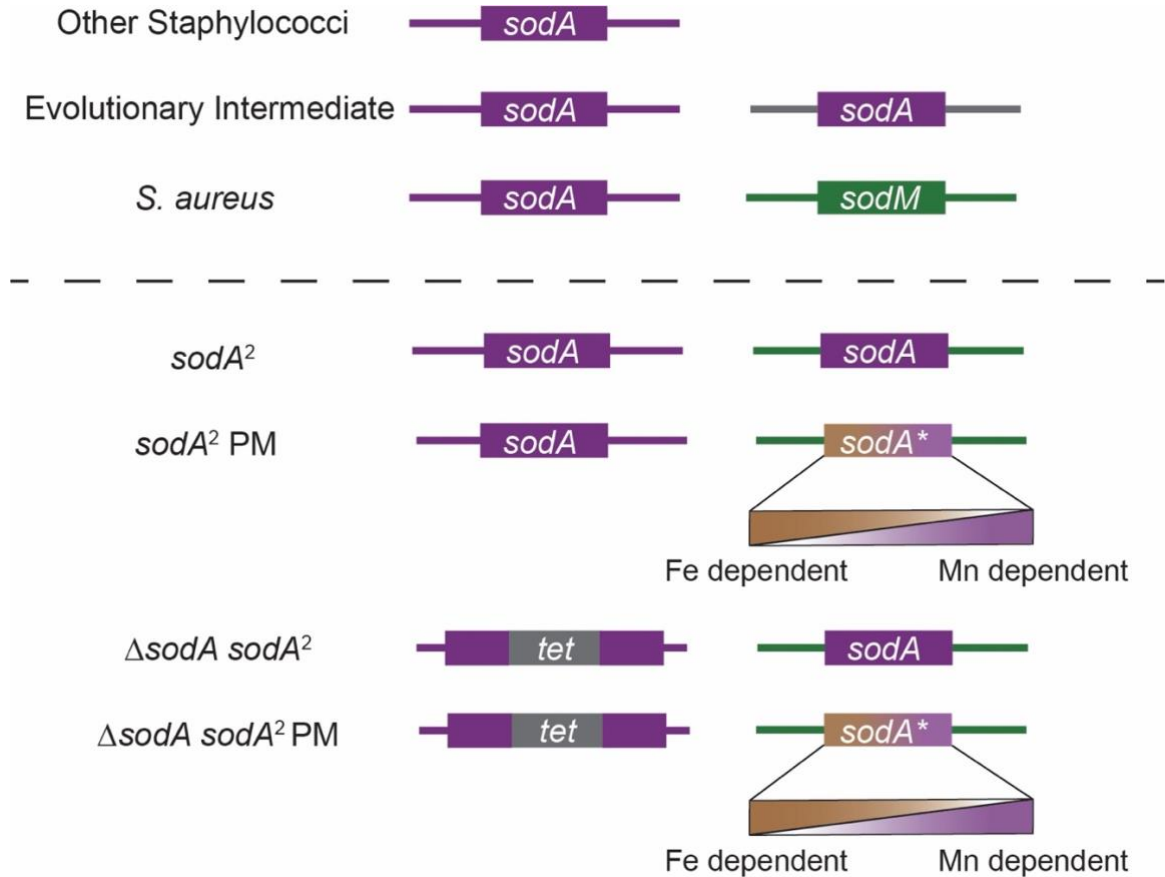


Figure 2.1. Schematic of *S. aureus* SOD evolution and strains used in this study. All staphylococci encode *sodA* while only *S. aureus* and two other closely related species encode *sodM*. Present-day *sodM* is predicted to have evolved first as a duplication of *sodA* and subsequent neofunctionalization to *sodM*. In an attempt to recreate this ancestral *S. aureus* strain, a *sodA*² strain containing the native copy of SodA and a second copy of SodA in the *sodM* locus was made. Incremental mutations that make SodA more like SodM were constructed in the second copy of *sodA* (*sodA*^{*}) making various *sodA*² point mutant (PM) strains in an attempt to recreate potential evolutionary intermediates on the path from Mn-dependency to cambialism. The *sodA*² and *sodA*² PM strains were also created with a disruption in SodA leading to strains Δ *sodA* *sodA*² and Δ *sodA* *sodA*² PM with a single copy of *sodA* (or a *sodA* PM variant) encoded in the *sodM* locus.

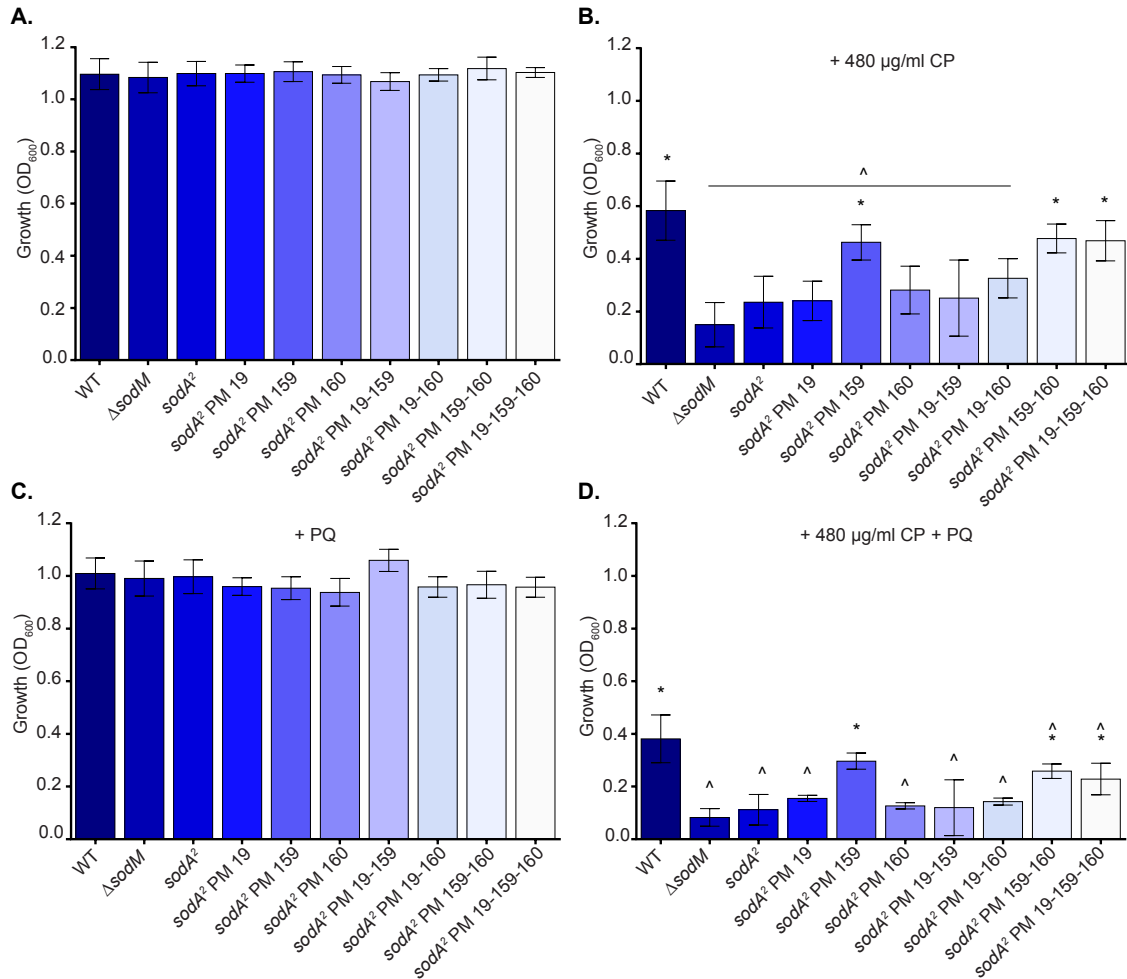


Figure 2.2. SodA variants with increased Fe-dependent activity provide a benefit to *S. aureus* in the presence of Mn starvation and oxidative stress. *S. aureus* wild type, Δ sodM, sodA², and sodA² point mutant (PM) strains were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP) in the absence (A, B) or presence (C, D) of 0.01mM paraquat (PQ). Growth was assessed by measuring optical density (OD₆₀₀) at 10 hrs. Error bars indicate SEM, p<0.05 relative to wild type (^) or sodA² (*) via two-way ANOVA with Dunnett's post-test, n = 6 – 21 (A, B) n = 6 – 11 (C, D).

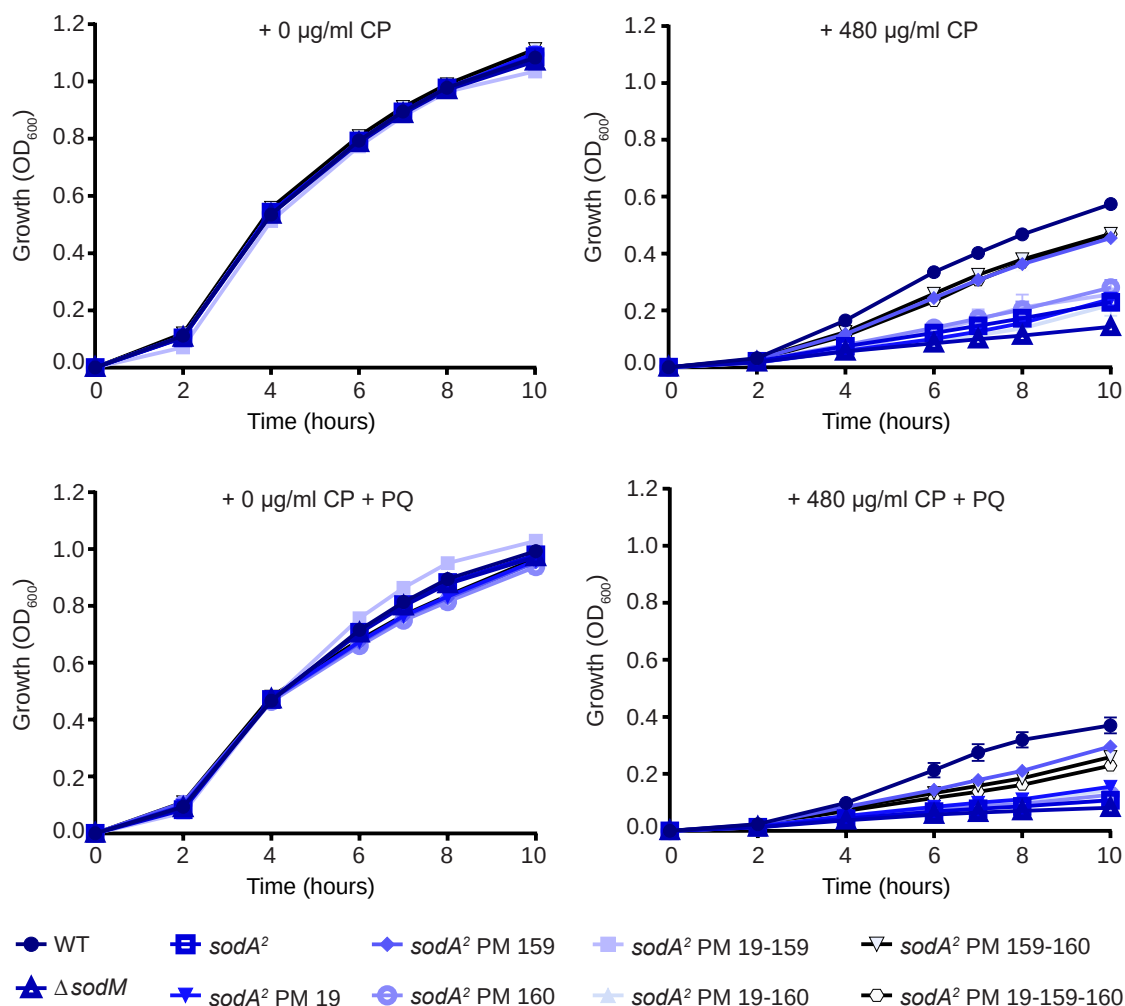


Figure 2.3. SodA variants with increased Fe-dependent activity provide a benefit to *S. aureus* in the presence of Mn starvation and oxidative stress. *S. aureus* wild type, Δ *sodM*, *sodA*², and *sodA*² point mutant (PM) strains were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP) in the absence or presence of 0.01mM paraquat (PQ). Growth was assessed by measuring optical density (OD₆₀₀) over time for 10 hrs. Error bars indicate SEM, n = 6 – 21 (no PQ) and n = 6 – 11 (+ PQ).

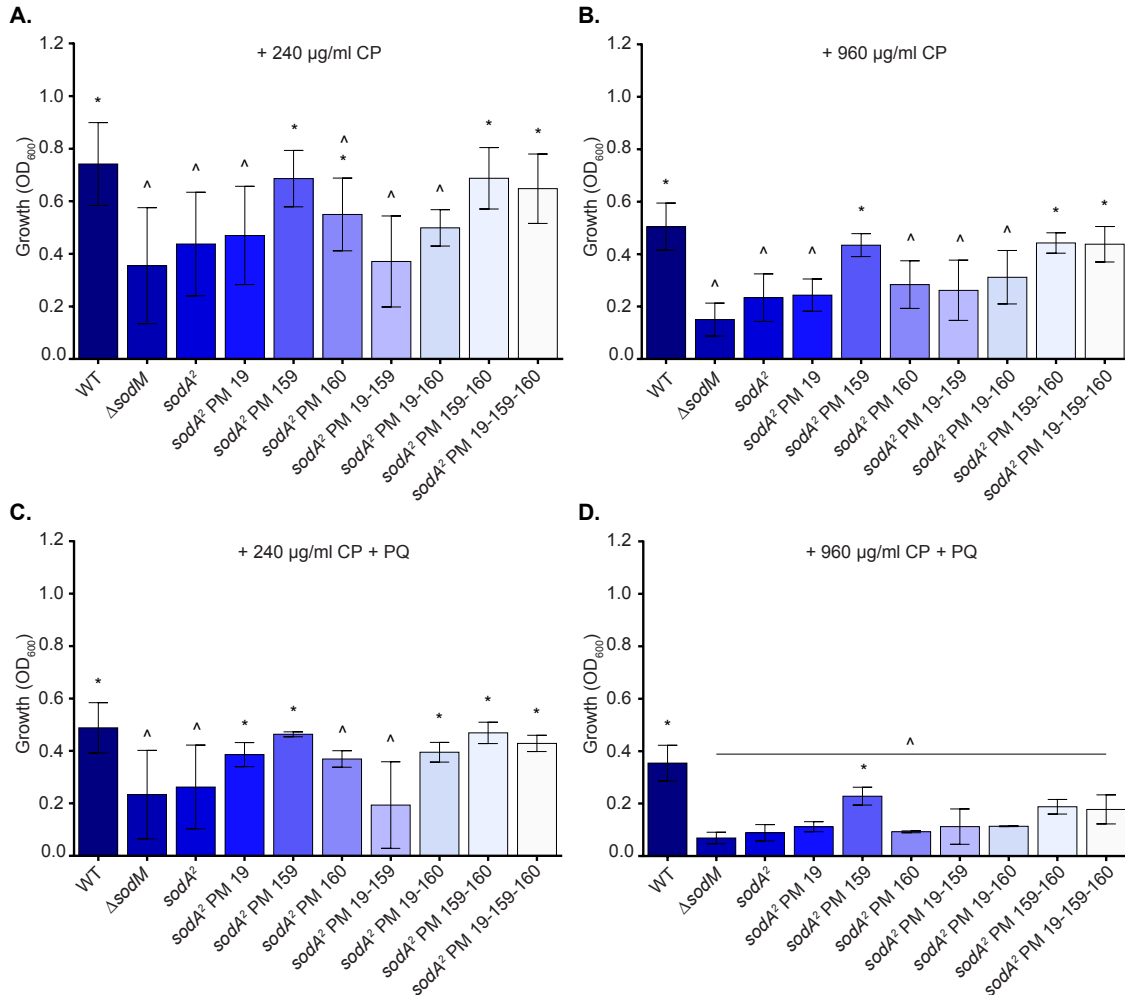


Figure 2.4. SodA variants with increased Fe-dependent activity provide a benefit to *S. aureus* in the presence of Mn starvation and oxidative stress. *S. aureus* wild type, Δ sodM, sodA², and sodA² point mutant (PM) strains were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of 240 µg/ml CP (A, C) or 960 µg/ml CP (B, D) in the absence (A, B) or presence (C, D) of 0.01mM paraquat (PQ). Growth was assessed by measuring optical density (OD₆₀₀) at 10 hrs. Error bars indicate SEM, p<0.05 relative to wild type (^) or sodA² (*) via two-way ANOVA with Dunnett's post-test, n = 6 – 21 (A, B) n = 6 – 11 (C, D).

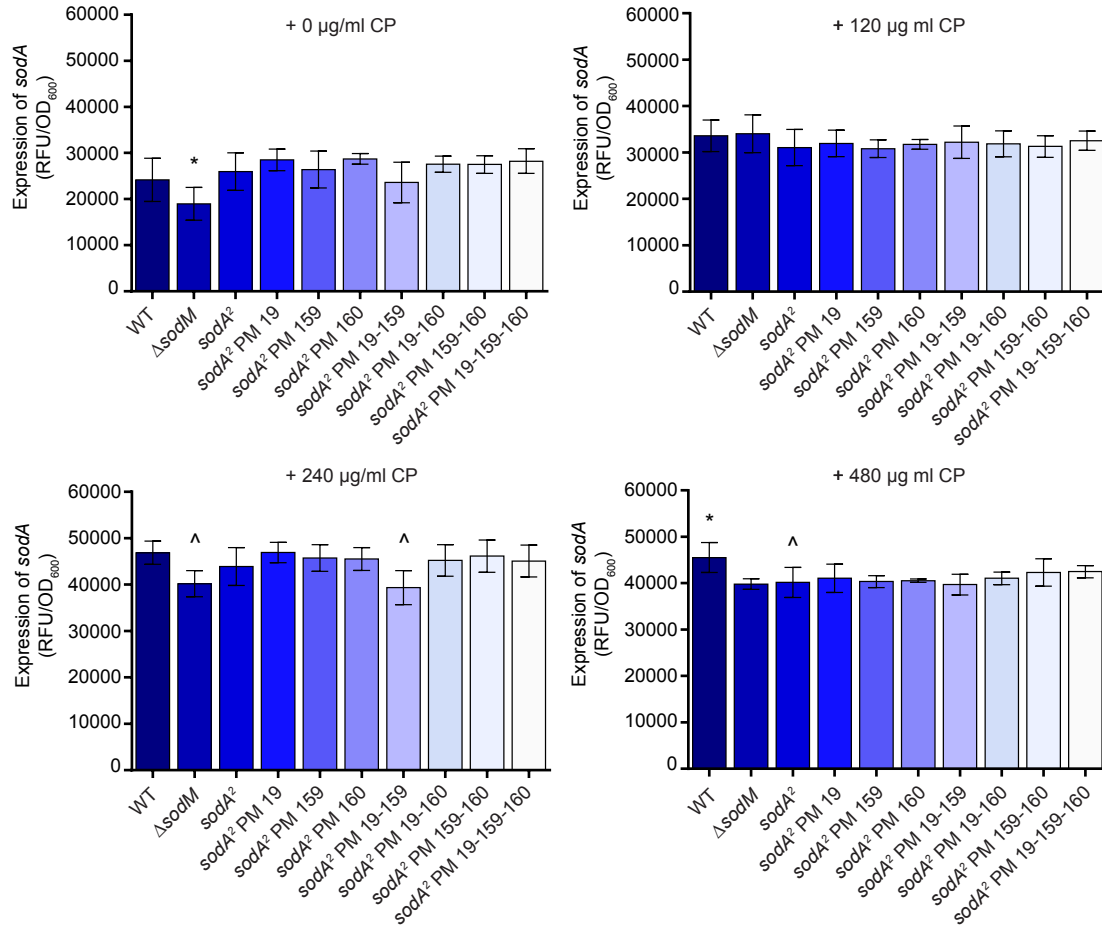


Figure 2.5. Growth differences among strains encoding SodA variants are not due to grossly altered expression of the SodA. *S. aureus* strains carrying a plasmid containing P_{sodA}-YFP reporter constructs were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP), and expression was assessed at 7 hrs. Error bars indicate SEM, p<0.05 relative to wild type (^) or *sodA*² (*) via two-way ANOVA with Dunnett's post-test, n = 3 – 6.

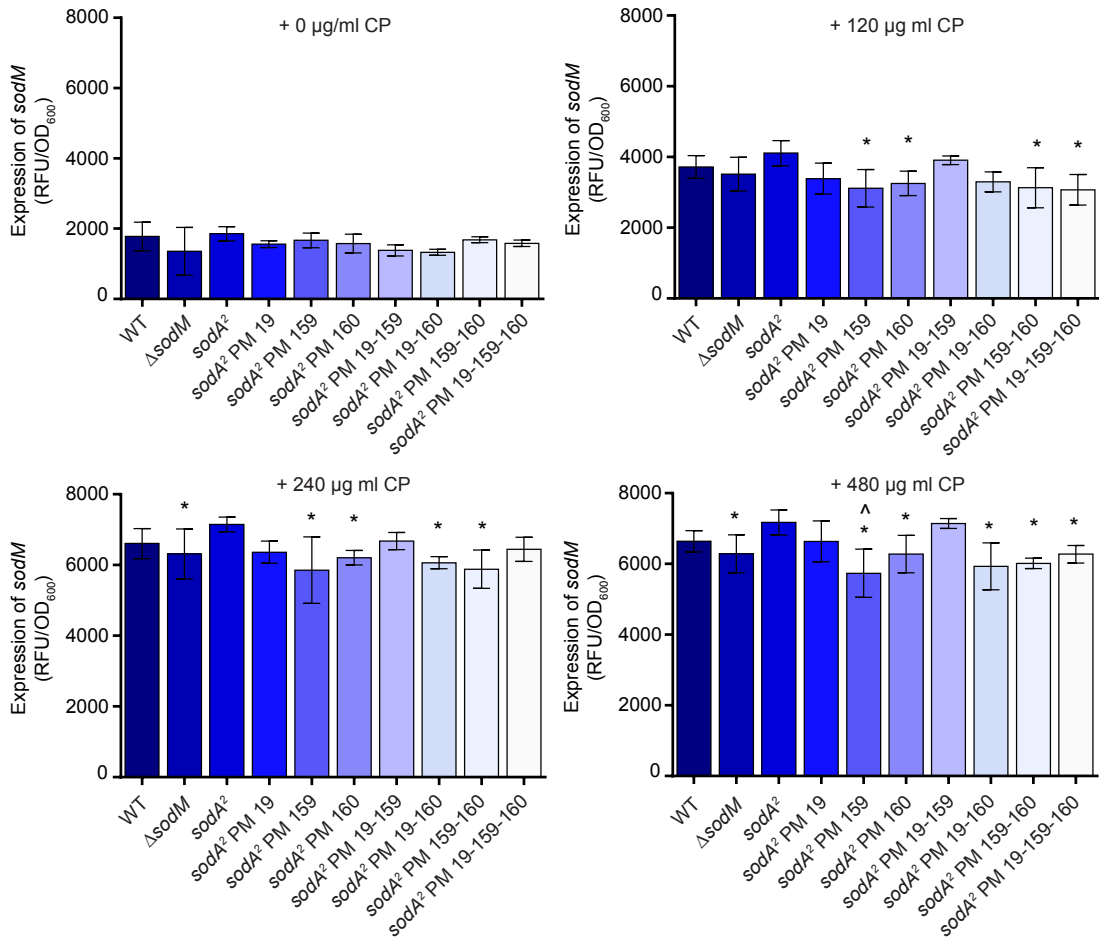


Figure 2.6. Growth differences among strains encoding the *SodA* variants are not due to grossly altered expression of the *SodM* locus. *S. aureus* strains carrying a plasmid containing *P_{sodM}*-YFP reporter constructs were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP), and expression was assessed at 7 hrs. Error bars indicate SEM, p < 0.05 relative to wild type (^) or *sodA*² (*) via two-way ANOVA with Dunnett's post-test, n = 3 – 6.

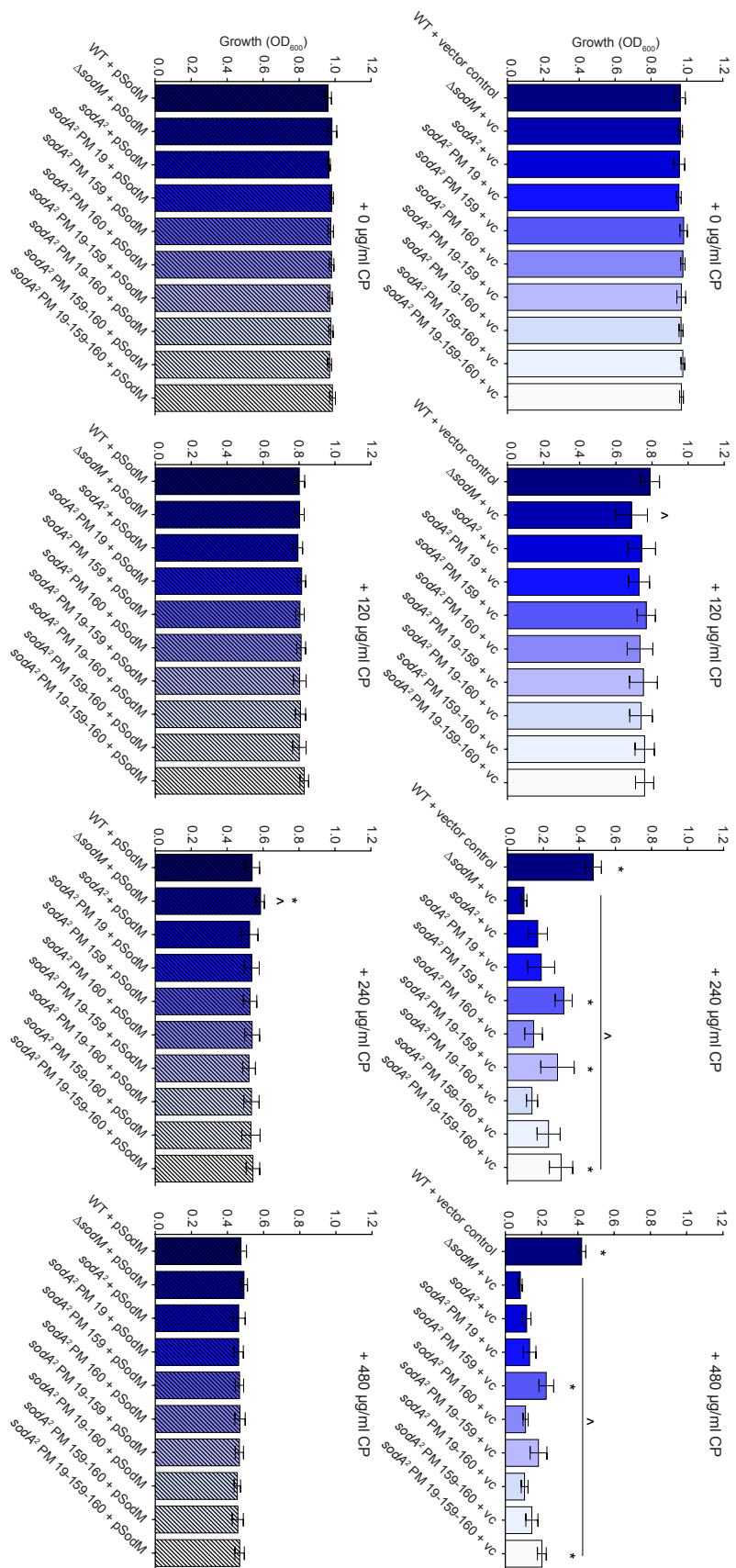


Figure 2.7. Expression of SodM from a plasmid rescues the growth defects of SodA²- and SodA² PM variants. *S. aureus* strains carrying an empty vector control (top panels) or *sodM* under the control of its native promoter (bottom panels) were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP) and 0.01mM PQ. Growth was assessed by measuring optical density (OD₆₀₀) at 8 hrs. Error bars indicate SEM, p<0.05 relative to wild type (Δ) or *sodA*² (*) via two-way ANOVA with Dunnett's post-test, n = 4.

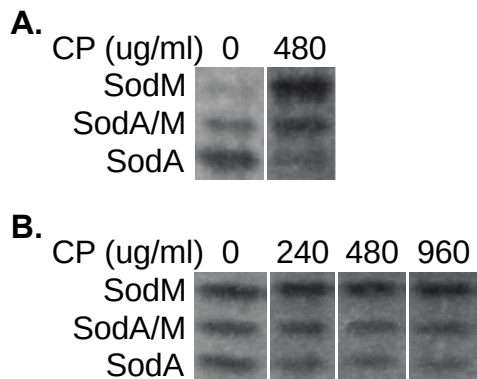


Figure 2.8. Evaluating SOD activity in WT *S. aureus* in various growth conditions. *S. aureus* wildtype was grown in the absence or presence of increasing amounts of calprotectin (CP) and SOD activity in cell lysates was determined. Cells were harvested after overnight growth in Mn-limited media (nRPMI) before growth in rich media (TSB) to an OD₆₀₀ of 0.3, similar to growth conditions used for evaluating growth of the strains (A, Figure 2.2), or after overnight growth in nRPMI, similar to conditions used for evaluating SOD activity (B, Figure 2.16).

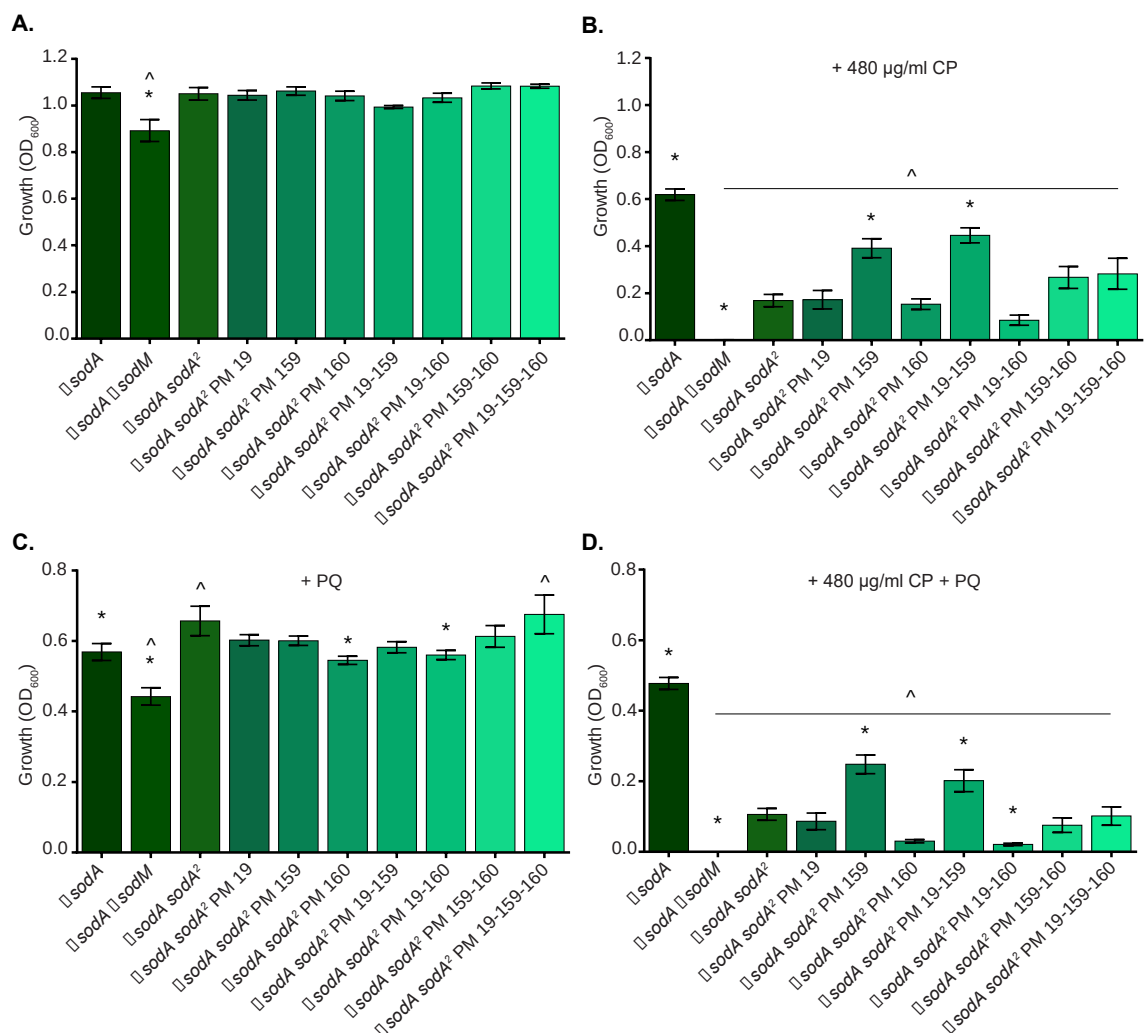


Figure 2.9. SodA variants Gly159Leu and Phe19Ile-Gly159Leu provide a benefit to *S. aureus* in the presence of Mn starvation and oxidative stress in the absence of a native copy of SodA. *S. aureus* $\Delta sodA$, $\Delta sodA \Delta sodM$, $\Delta sodA \Delta sodA^2$, and $\Delta sodA \Delta sodA^2$ point mutant (PM) strains were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP) in the absence (A, B) or presence (C, D) of 0.01mM paraquat (PQ). Growth was assessed by measuring optical density (OD₆₀₀) at 10 hrs. Error bars indicate SEM, $p < 0.05$ relative to $\Delta sodA$ (^) or $\Delta sodA \Delta sodA^2$ (*) via two-way ANOVA with Dunnett's post-test, $n = 3 - 16$ (A, B) $n = 3 - 8$ (C, D).

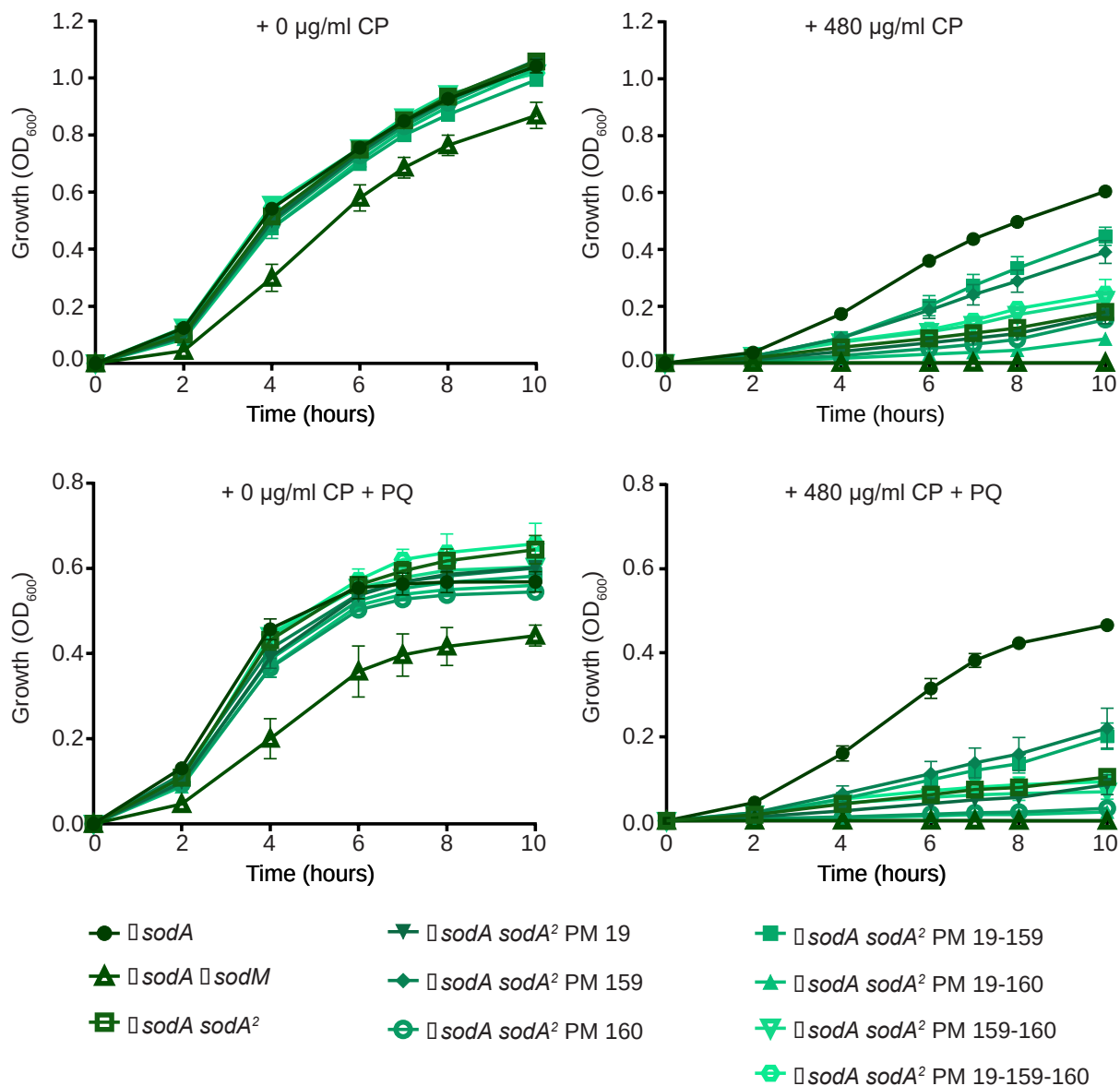


Figure 2.10. SodA variants Gly159Leu and Phe19Ile-Gly159Leu provide a benefit to *S. aureus* in the presence of Mn starvation and oxidative stress in the absence of a native copy of SodA. *S. aureus* ΔsodA , $\Delta\text{sodA}\Delta\text{sodM}$, $\Delta\text{sodA}\text{sodA}^2$, and $\Delta\text{sodA}\text{sodA}^2$ point mutant (PM) strains were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP) in the absence or presence of 0.01mM paraquat (PQ). Growth was assessed by measuring optical density (OD₆₀₀) over 10 hrs. Error bars indicate SEM, n = 3 – 16 (no PQ) and n = 3 – 8 (+ PQ).

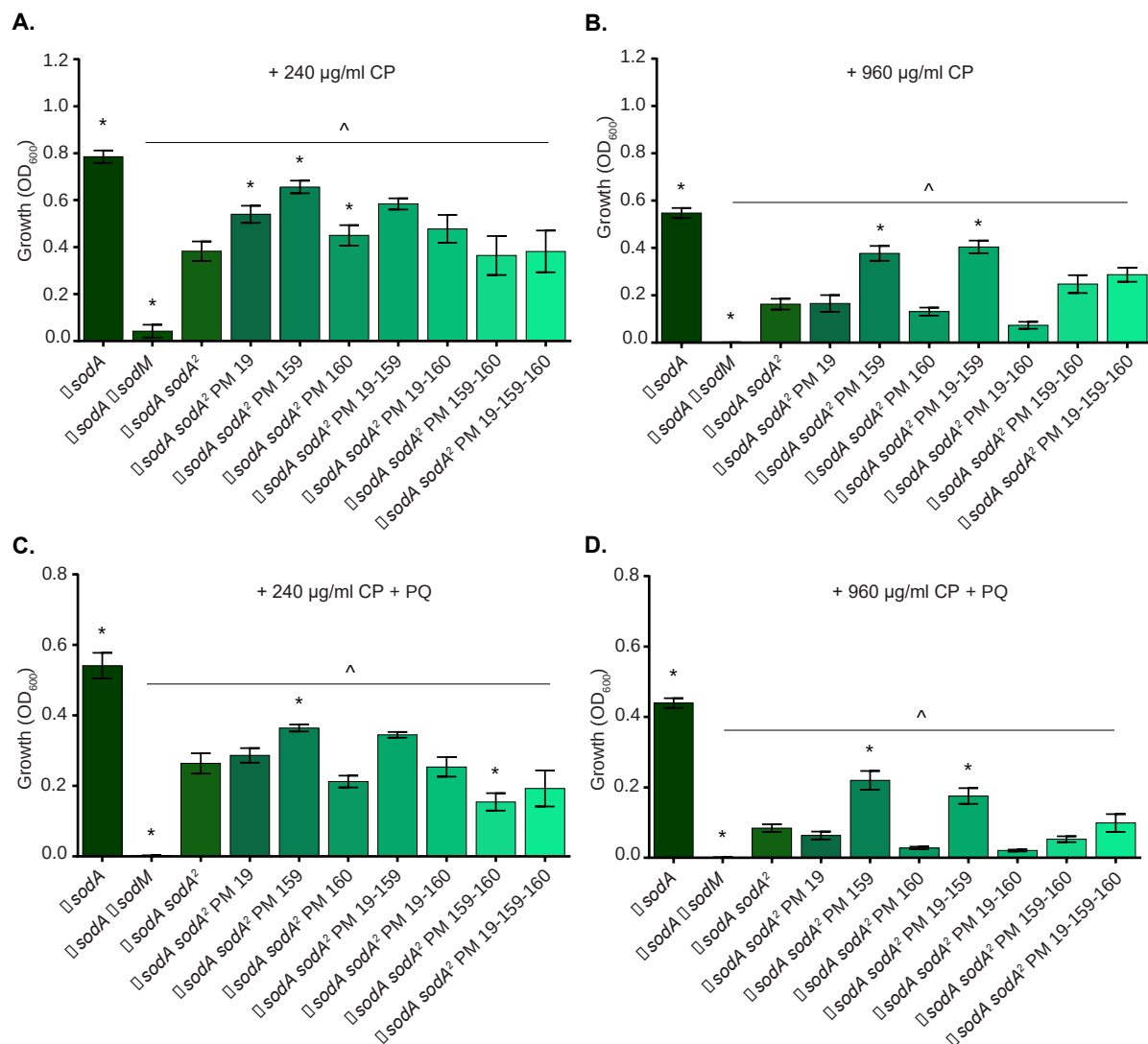
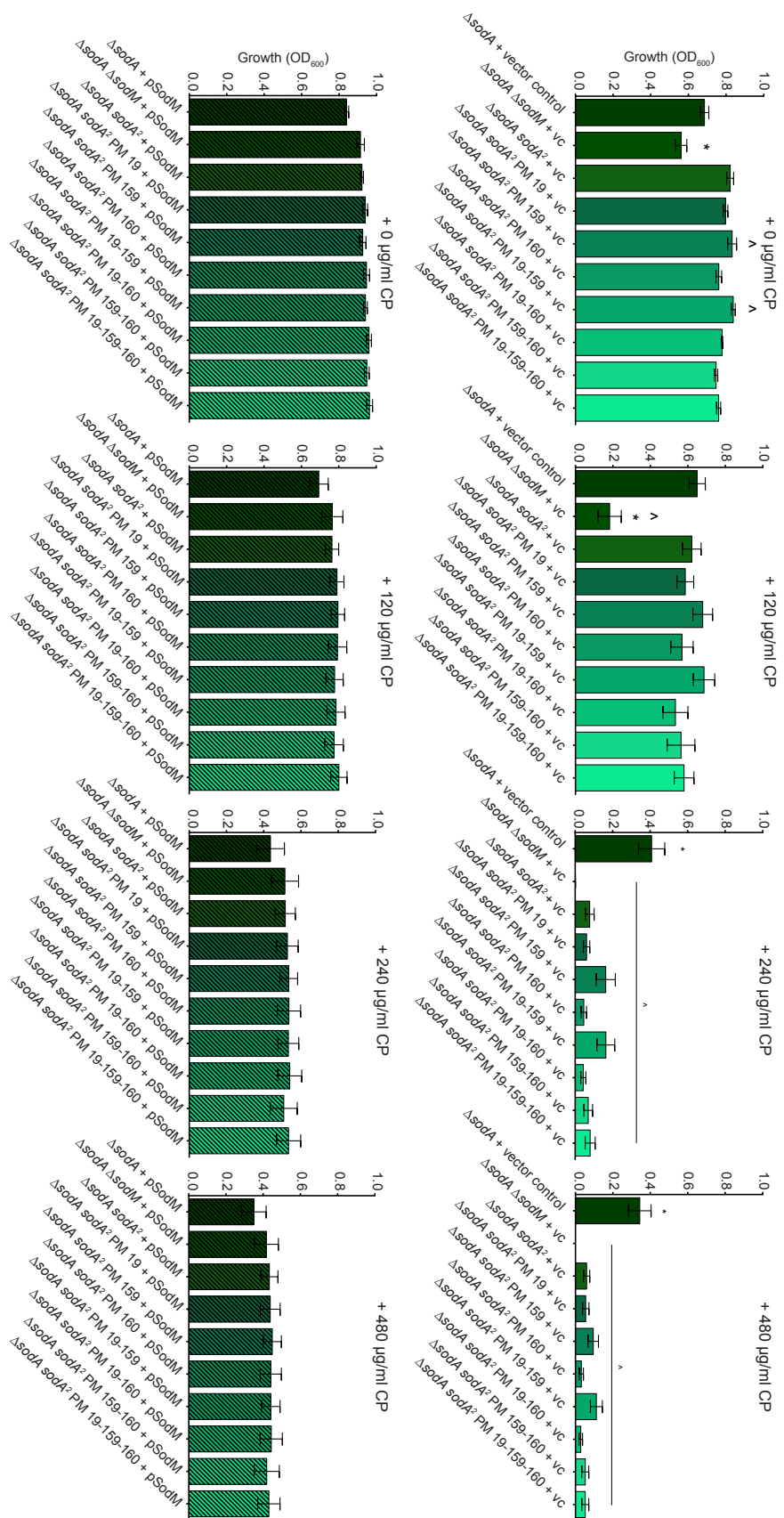


Figure 2.11. SodA variants Gly159Leu and Phe19Ile-Gly159Leu provide a benefit to *S. aureus* in the presence of Mn starvation and oxidative stress in the absence of a native copy of SodA. *S. aureus* Δ sodA, Δ sodA Δ sodM, Δ sodA Δ sodA², and Δ sodA Δ sodA² point mutant (PM) strains were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of 240 µg/ml CP (A, C) or 960 µg/ml CP (B, D) in the absence (A, B) or presence (C, D) of 0.01mM paraquat (PQ). Growth was assessed by measuring optical density (OD₆₀₀) at 10 hrs. Error bars indicate SEM, p<0.05 relative to Δ sodA (^) or Δ sodA sodA² (*) via two-way ANOVA with Dunnett's post-test, n = 3 – 16 (A, B) n = 3 – 8 (C, D).



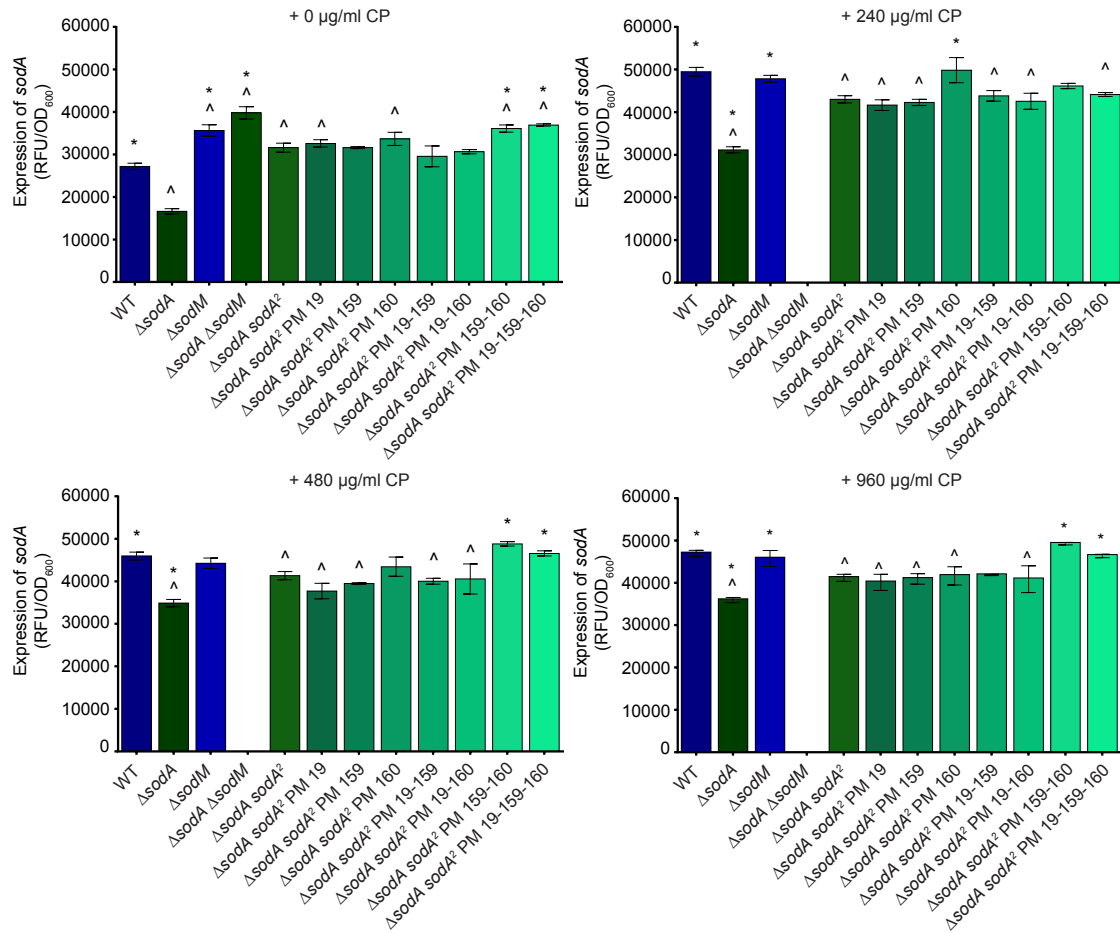


Figure 2.13. Growth differences among *S. aureus* strains encoding SodA variants are not likely due to altered expression from the *sodA* locus. *S. aureus* strains carrying a plasmid containing PsodA-YFP reporter constructs were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP), and expression was assessed at 7 hrs. Expression from strains with growth less than an OD₆₀₀ of 0.2 have been excluded. Error bars indicate SEM, $p < 0.05$ relative to wild type (^) or Δ *sodA* *sodA*² (*) via two-way ANOVA with Dunnett's post-test, $n = 3 - 6$.

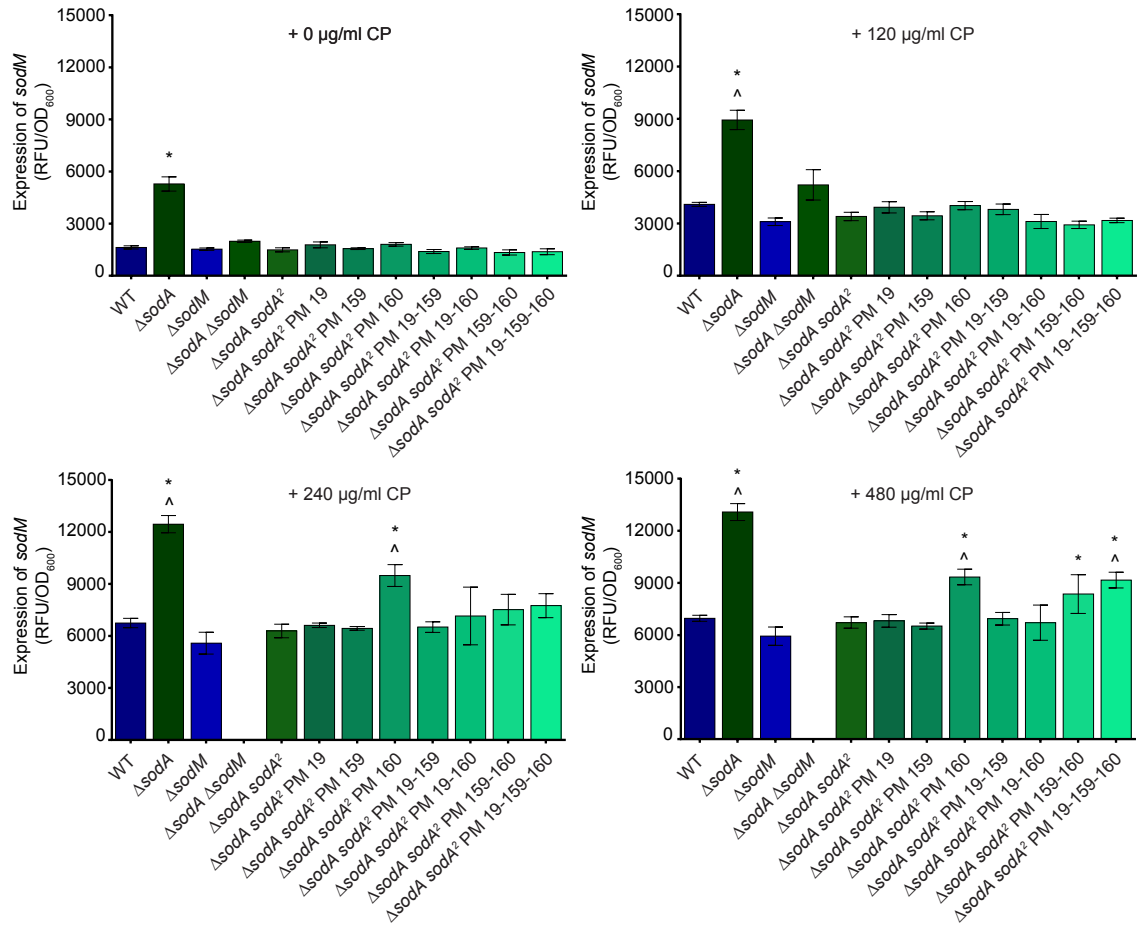


Figure 2.14. Growth differences among *S. aureus* strains encoding *SodA* variants are not likely due to altered expression from the *sodM* locus. *S. aureus* strains carrying a plasmid containing P_{*sodM*}-YFP reporter constructs were cultured overnight in Mn-limited media (nRPMI) before growth in rich media (TSB) in the presence of increasing concentrations of calprotectin (CP), and expression was assessed at 7 hrs. Expression from strains with growth less than an OD₆₀₀ of 0.2 have been excluded. Error bars indicate SEM, p<0.05 relative to wild type (^) or Δ *sodA* *sodA*² (*) via two-way ANOVA with Dunnett's post-test, n = 3 – 6.

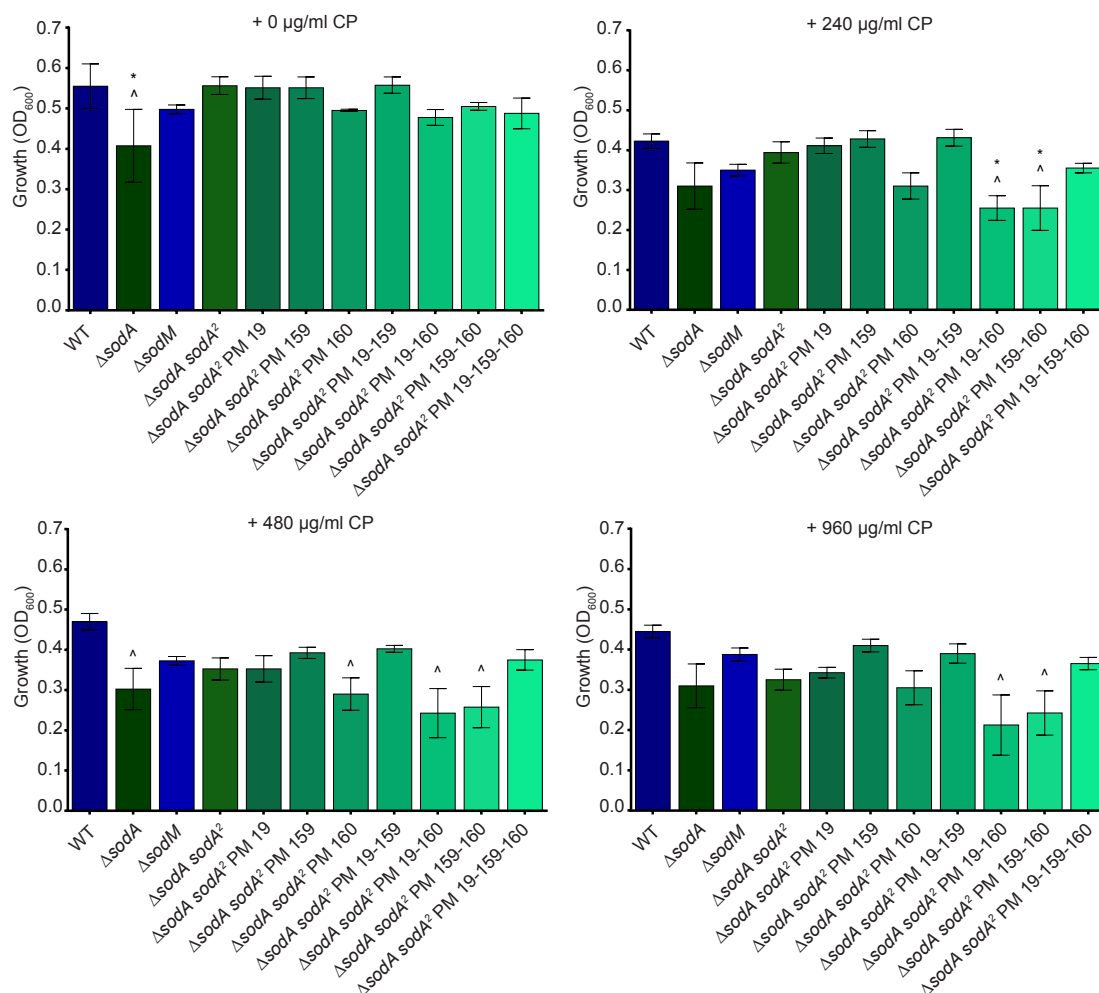


Figure 2.15. Growth of *S. aureus* wild-type, Δ sodA, Δ sodM, Δ sodA sodA², and Δ sodA sodA² SodA variant strains after overnight growth before harvesting for SOD activity assays. *S. aureus* strains were grown overnight in the absence or presence of increasing amounts of calprotectin (CP). Growth was assessed by measuring optical density (OD₆₀₀). Error bars indicate SEM, p < 0.05 relative to wild type (^) or Δ sodA sodA² (*) via two-way ANOVA with Dunnett's post-test, n = 4 – 6.

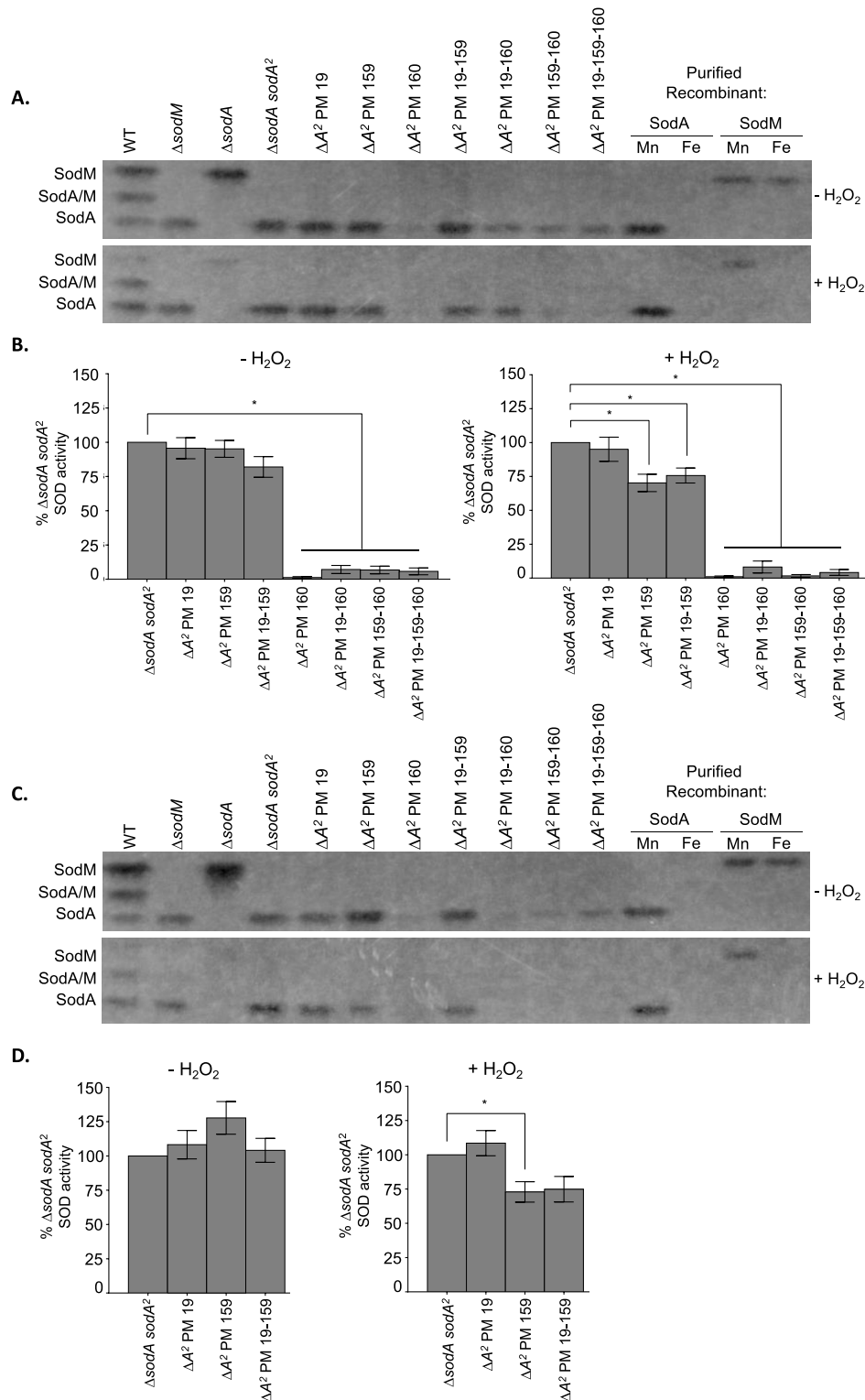


Figure 2.16. SodA variants containing reciprocal mutation Leu160Phe have decreased total SOD activity in the absence of a native copy of SodA and SodA variants Gly159Leu and Phe19Ile-Gly159Leu have increased Fe selectivity as compared to WT SodA. *S. aureus* strains were grown

Figure 2.16. Continued. overnight in the absence (A, B) or presence (C, D) of 240 µg/ml of calprotectin (CP) and SOD activity in cell lysates was determined. Gels were treated with hydrogen peroxide prior to assessing SOD activity to inactivate Fe-containing SODs. For *S. aureus* $\Delta sodA$ $sodA^2$ strains encoding SodA variants, $\Delta sodA$ $sodA^2$ has been abbreviated to ΔA^2 in the figure (ex. ΔA^2 PM 19 is a $\Delta sodA$ $sodA^2$ SodA-PM Phe19Ile strain). Quantification of SOD activity was performed on samples with 4 – 96 µg total protein loaded such that each sample was within the linear range of purified recombinant SodA-Mn. SOD activity values are represented as a percent of SOD activity in $\Delta sodA$ $sodA^2$. Representative gels shown are loaded with 10 µg (0 CP) or 16 µg (240 CP) for each sample in the gel. (D) SOD activity for SodA variants with amino acid reciprocal mutations including residue 160 (Leu160Phe, Phe19Ile-Leu160Phe, Gly159Leu-Leu160Phe, and Phe19Ile-Gly159Leu-Leu160Phe) were below the linear range in samples harvested from cells grown in 240 µg/ml CP and are therefore not included in the quantification. Purified recombinant SodA and SodM loaded with either Mn or Fe were included as controls at 0.1µg total protein in these representative gels. SOD activity was assessed in four to ten biological replicates with representative activity gels shown. Error bars indicate SEM, $p < 0.05$ (*) via one-way ANOVA with Dunnett's post-test.

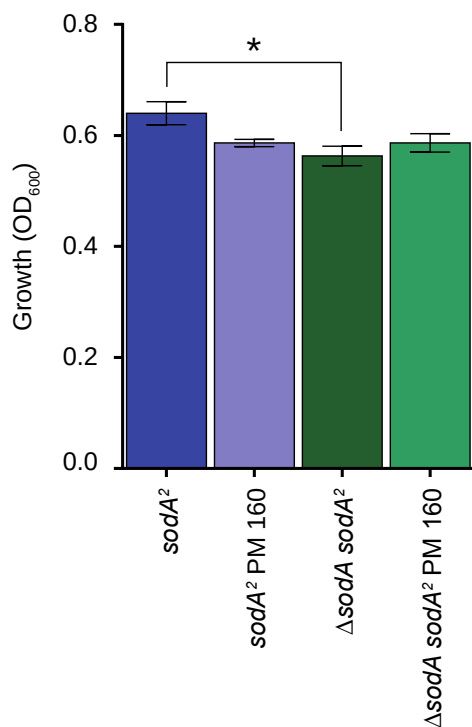


Figure 2.17. *S. aureus* strains grow largely equivalently after growth for harvesting of cell lysates for LC-MS. *S. aureus* cell density after overnight growth was assessed by measuring optical density (OD₆₀₀). Error bars indicate SEM, $p < 0.05$ (*) via two-way ANOVA with Dunnett's post-test, $n = 3$.

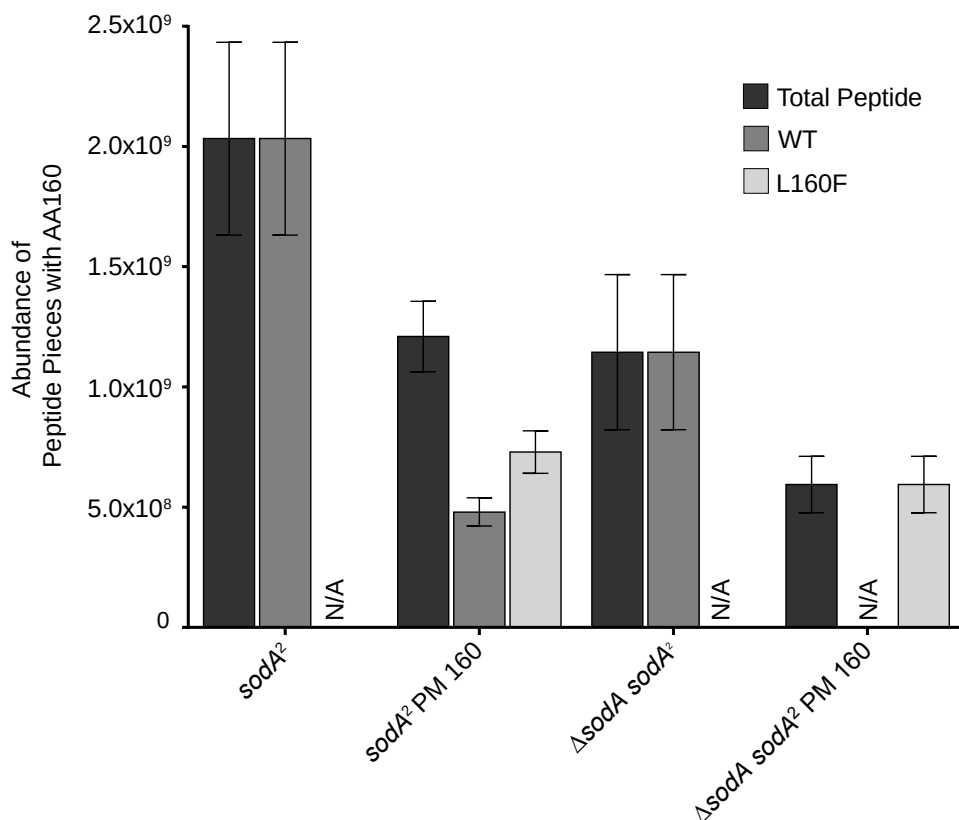


Figure 2.18. Protein abundance of SodA variants encoding the Leu160Phe mutation is not reduced. *S. aureus* *sodA*², *sodA*² point mutant (PM) Leu160Phe, Δ*sodA sodA*², and Δ*sodA sodA*² PM Leu160Phe were grown overnight, and cell lysates were harvested and subjected to LC-MS. Total peptide represents the sum of WT and L160F peptide abundances in that sample. Not applicable (N/A) denotes there was not WT/L160F peptide in that sample, therefore, there is a zero value for peptide abundance. Error bars indicate SEM, n = 3.

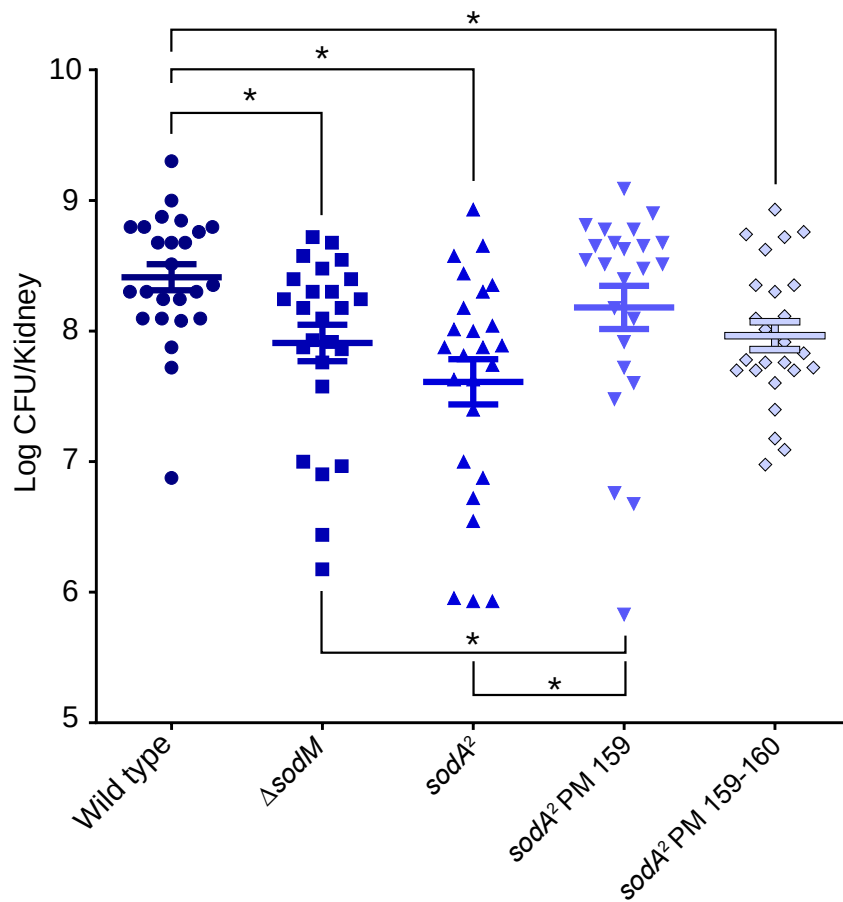


Figure 2.19. A Gly159Leu mutation in SodA restores staphylococcal virulence. Wild type C57BL/6 mice were systemically infected with either wild-type *S. aureus*, Δ sodM, sodA², sodA² point mutant (PM) Gly159Leu, or sodA² PM Gly159Leu-Leu160Phe and bacterial burdens were enumerated after the mice were sacrificed 96 hours post-infection. p<0.05 (*) as determined by Mann-Whitney test. Lines represent the mean and error bars represent the standard error of the mean.

Table 2.1. Enzymatic Activity of WT SodA, WT SodM, and SodA residue swapped mutants.

SOD Enzyme	Mutation	Mn Activity^a	Fe Activity^a	Cambialism Ratio^b
SodA	-	1836 ± 79	4 ± 1	0.002
SodA	F19I	1734 ± 61	5 ± 1	0.003
SodA	G159L	942 ± 69	39 ± 4	0.041
SodA	L160F ^c	1167 ± 92	19 ± 13	0.016
SodA	G159L-L160F	173 ± 21	84 ± 12	0.486
SodA	F19I-G159L-L160F	526 ± 49	57 ± 7	0.108
SodM	-	251 ± 5	250 ± 10	0.996

^a Values represent Units mg⁻¹ protein ± SD.

^b The cambialism ratio (CR) is calculated by dividing the iron-dependent activity by the manganese-dependent activity, with a value of 1 representing a perfectly cambialistic enzyme. Enzymes that are cambialistic are defined herein as having a CR in the range 0.1–10.00.

^c Whereas all other samples were assayed simultaneously to circumvent noticeable day-to-day variability in the absolute activity detected using the commercial assay, the single-mutant variants with mutations at position 160 were assayed separately. To ensure comparability between the datasets, these variants were assayed simultaneously alongside samples of all wild-type forms, and small discrepancies in the absolute values were normalized by scaling using those wild-type activities as controls.

*Data for this table was originally generated for Barwinska-Sendra *et al.*, 2020.

Table 2.2. Strains and plasmids used in this study.

Genotype	Name	Description	Source
Wild type	WT	<i>S. aureus</i> Newman	
<i>sodA::tet</i>	$\Delta sodA$	<i>S. aureus</i> Newman $\Delta sodA$	23
<i>sodM::erm</i>	$\Delta sodM$	<i>S. aureus</i> Newman $\Delta sodM$	23
<i>sodA::tet sodM::erm</i>	$\Delta sodA \Delta sodM$	<i>S. aureus</i> Newman $\Delta sodA \Delta sodM$	31
<i>sodM::sodA</i>	$sodA^2$	<i>S. aureus</i> Newman with <i>sodA</i> replacing <i>sodM</i>	This study
<i>sodM::sodA F19I</i>	$sodA^2$ PM19	<i>S. aureus</i> Newman with <i>sodA F19I</i> replacing <i>sodM</i>	This study
<i>sodM::sodA G159L</i>	$sodA^2$ PM159	<i>S. aureus</i> Newman with <i>sodA G159L</i> replacing <i>sodM</i>	This study
<i>sodM::sodA L160F</i>	$sodA^2$ PM160	<i>S. aureus</i> Newman with <i>sodA L160F</i> replacing <i>sodM</i>	This study
<i>sodM::sodA F19I G159L</i>	$sodA^2$ PM19-159	<i>S. aureus</i> Newman with <i>sodA F19I G159L</i> replacing <i>sodM</i>	This study
<i>sodM::sodA F19I L160F</i>	$sodA^2$ PM19-160	<i>S. aureus</i> Newman with <i>sodA F19I L160F</i> replacing <i>sodM</i>	This study
<i>sodM::sodA G159L L160F</i>	$sodA^2$ PM159-160	<i>S. aureus</i> Newman with <i>sodA G159L L160F</i> replacing <i>sodM</i>	This study
<i>sodM::sodA F19I G159L L160F</i>	$sodA^2$ PM19-159-160	<i>S. aureus</i> Newman with <i>sodA F19I G159L L160F</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA</i>	$\Delta sodA sodA^2$	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA F19I</i>	$\Delta sodA sodA^2$ PM19	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA F19I</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA G159L</i>	$\Delta sodA sodA^2$ PM159	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA G159L</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA L160F</i>	$\Delta sodA sodA^2$ PM160	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA L160F</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA F19I G159L</i>	$\Delta sodA sodA^2$ PM19-159	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA F19I G159L</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA F19I L160F</i>	$\Delta sodA sodA^2$ PM19-160	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA F19I L160F</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA G159L L160F</i>	$\Delta sodA sodA^2$ PM159-160	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA G159L L160F</i> replacing <i>sodM</i>	This study
<i>sodA::tet sodM::sodA F19I G159L L160F</i>	$\Delta sodA sodA^2$ PM19-159-160	<i>S. aureus</i> Newman $\Delta sodA$ with <i>sodA F19I G159L L160F</i> replacing <i>sodM</i>	This study

* Each of the above strains with each of the below plasmids were also created in this study

Table 2.2 Continued.

Plasmid		Description	Source
PAH5		PAH5 with no promoter driving YFP	23
PAH5 psodA-YFP		PAH5 with the sodA promoter driving YFP	23
PAH5 psodM-YFP		PAH5 with the sodM promoter driving YFP	23
pKK30		pKK30 stable vector	35
pKK30 psodM-SodM		pKK30 with sodM under its native promoter	This study

Table 2.3. Primers used in this study.

Primer	Sequence (5' to 3')
sodM_upstream_F	TGACCA TGTAA TACGACTCACTATAGGGGATATCTTTAGTAGGTACATACCAACAAGTTG
sodM_upstream_R	TGGTAATTC AAAAGGCCATAGTATATTCCTCCTTTTATGAATATAC
sodM_downstream_F	TATAATGCAACA AAAATAATATACTTAATAATAATTGAGGTGAG
sodM_downstream_R	GAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAGGAAGGACCTTTCACCAGCAGCTGG
soda_F	AAAAAGGAGGAATATACTTATGGCTTTTGAATTACCAAAAATTAC
soda_R	CAATTATATTAAGTTATATATTTTGTGCAATTATATAATTG
SDM_Soda_19_F	AGAACCACATATTGACAAAGA AAC
SDM_Soda_19_R	GTTTCTTTGTC AATATGTGTTCT
SDM_Soda_159_F	GGGTAAAAACACCTATTTTACTGTTAGACGTATGGACACG
SDM_Soda_159_R	CGTGTCCCATACGTC TAAACAGTAAAA TAGGTGTTTACCC
SDM_Soda_160_F	GGGTAAAAACACCTATTTTAGGTTTCGACGTATGGAAACACG
SDM_Soda_160_R	CGTGTCCCATACGTCGAAAACCTAAAA TAGGTGTTTACCC
SDM_Soda_159_160_F	CACCTATTTTACTGTTGACGTA TGGAAACACGCTTATTAC
SDM_Soda_159_160_R	GTAATAAGCGGTGTTCCCATACGTCGAAACAGTAAAA TAGGTG
sodM_pKK30_F	AGCAAAAGTGACAGCGCATGCGTACCGGTAA TGTCTATAC
sodM_pKK30_R	TGCTTGTAA TTCATGATTGCTTATTTTGTGCTGCTGGTATAATTC

2.7 References

1. Anbar AD. 2008. Oceans: Elements and evolution. *Science*. 322(5907):1481–83.
2. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM. 2008. Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem*. 13(8):1205–18.
3. Bae T, Schneewind O. 2006. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid*. 55(1):58–63.
4. Barra D, Schininà ME, Bossa F, Puget K, Durosay P, et al. 1990. A tetrameric iron superoxide dismutase from the eucaryote *Tetrahymena pyriformis*. *J Biol Chem*. 265(29):17680–87.
5. Barwinska-Sendra A, Garcia YM, Sendra KM, Baslé A, Mackenzie ES, et al. 2020. An evolutionary path to altered cofactor specificity in a metalloenzyme. *Nat Commun*. 11(1):2738.
6. Barwinska-Sendra A, Waldron KJ. 2017. The Role of intermetal competition and mis-metalation in metal toxicity. In *Advances in Microbial Physiology*. 70:315–79.
7. Bekker A, Holland HD, Wang P-L, Rumble D, Stein HJ, et al. 2004. Dating the rise of atmospheric oxygen. *Nature*. 427(6970):117–20.
8. Besold AN, Gilston BA, Radin JN, Ramsoomair C, Culbertson EM, et al. 2018. Role of calprotectin in withholding zinc and copper from *Candida albicans*. *Infection and Immunity*. 86(2):1–16.
9. Beyer WF, Fridovich I. 1991. In vivo competition between iron and manganese for occupancy of the active site region of the manganese-superoxide dismutase of *Escherichia coli*. *Journal of Biological Chemistry*. 266(1):303–8.
10. Brioukhanov AL, Netrusov AI. 2004. Catalase and superoxide dismutase: distribution, properties, and physiological role in cells of strict anaerobes. *Biochemistry (Moscow)*. 69(9):949–62.
11. Brown JR, Doolittle WF. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev*. 61(4):456–502.
12. Broxton CN, Culotta VC. 2016. SOD Enzymes and microbial pathogens: surviving the oxidative storm of infection. *PLoS Pathog*. 12(1):e1005295.

13. Burton NA, Schürmann N, Casse O, Steeb AK, Claudi B, et al. 2014. Disparate impact of oxidative host defenses determines the fate of *Salmonella* during systemic infection in mice. *Cell Host & Microbe*. 15(1):72–83.
14. Capdevila DA, Edmonds KA, Giedroc DP. 2017. Metallochaperones and metalloregulation in bacteria. *Essays In Biochemistry*. 61(2):177–200.
15. Corbin, B. D., Seeley, E. H., Raab, A., Feldmann, J., Miller, M. R., Torres, V. J., Anderson, K. L., Dattilo, B. M., Anderson, K. L., Dunman, P. M., Gerads, R., Caprioli, R., Nacken, W., Chazin WJ, Skaar EP. 2008. Metal chelation and inhibition of bacterial growth in tissue abscesses. *Science*. 319(February):962–66.
16. Culotta VC, Yang M, O'Halloran TV. 2006. Activation of superoxide dismutases: Putting the metal to the pedal. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 1763(7):747–58.
17. Damo SM, Kehl-Fie TE, Sugitani N, Holt ME, Rathi S, et al. 2013. Molecular basis for manganese sequestration by calprotectin and roles in the innate immune response to invading bacterial pathogens. *Proceedings of the National Academy of Sciences*. 110(10):3841–46.
18. Dupont CL, Butcher A, Valas RE, Bourne PE, Caetano-Anolles G. 2010. History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proceedings of the National Academy of Sciences*. 107(23):10567–72.
19. Edwards RA, Baker HM, Whittaker MM, Whittaker JW, Jameson GB, Baker EN. 1998. Crystal structure of *Escherichia coli* manganese superoxide dismutase at 2.1-Å resolution. *J Biol Inorg Chem*. 3(2):161–71.
20. Eom H, Song WJ. 2019. Emergence of metal selectivity and promiscuity in metalloenzymes. *J Biol Inorg Chem*. 24(4):517–31.
21. Fenlon LA, Slauch JM. 2014. Phagocyte roulette in *Salmonella* killing. *Cell Host & Microbe*. 15(1):7–8.
22. Foster AW, Osman D, Robinson NJ. 2014. Metal preferences and metallation. *Journal of Biological Chemistry*. 289(41):28095–103.
23. Frye KA, Sendra KM, Waldron KJ, Kehl-Fie TE. 2022. Old dogs, new tricks: New insights into the iron/manganese superoxide dismutase family. *Journal of Inorganic Biochemistry*. 230:111748.

24. Garcia YM, Barwinska-Sendra A, Tarrant E, Skaar EP, Waldron KJ, Kehl-Fie TE. 2017. A Superoxide dismutase capable of functioning with iron or manganese promotes the resistance of *Staphylococcus aureus* to calprotectin and nutritional immunity. *PLOS Pathogens*. 13(1):e1006125.
25. Grove LE, Xie J, Yikilmaz E, Miller A-F, Brunold TC. 2008. Spectroscopic and computational investigation of second-sphere contributions to redox tuning in *Escherichia coli* iron superoxide dismutase. *Inorg. Chem.* 47(10):3978–92.
26. Hampton MB, Kettle AJ, Winterbourn CC. 1996. Involvement of superoxide and myeloperoxidase in oxygen-dependent killing of *Staphylococcus aureus* by neutrophils. *Infect Immun.* 64(9):3512–17.
27. Hood MI, Skaar EP. 2012. Nutritional immunity: Transition metals at the pathogen-host interface. *Nature Reviews Microbiology*. 10(8):525–37.
28. Imlay JA. 2008. Cellular defenses against superoxide and hydrogen peroxide. *Annual Review of Biochemistry*. 77(1):755–76.
29. Imlay JA. 2015. Common mechanisms of bacterial metal homeostasis. In *Trace Metals and Infectious Diseases*. Cambridge (MA): MIT Press.
30. Irving H, Williams RJP. 1948. Order of stability of metal complexes. *Nature*. 162(4123):746–47.
31. Ji H-F, Chen L, Jiang Y-Y, Zhang H-Y. 2009. Evolutionary formation of new protein folds is linked to metallic cofactor recruitment. *BioEssays*. 31(9):975–80.
32. Kehl-Fie TE, Chitayat S, Hood MI, Damo S, Restrepo N, et al. 2011. Nutrient metal sequestration by calprotectin inhibits bacterial superoxide defense, enhancing neutrophil killing of *Staphylococcus aureus*. *Cell Host and Microbe*. 10(2):158–64.
33. Kehl-Fie TE, Skaar EP. 2010. Nutritional immunity beyond iron: a role for manganese and zinc. *Current Opinion in Chemical Biology*. 14(2):218–24.
34. Kehl-Fie TE, Zhang Y, Moore JL, Farrand AJ, Hood MI, et al. 2013. MntABC and MntH contribute to systemic *Staphylococcus aureus* infection by competing with calprotectin for nutrient manganese. *Infection and Immunity*. 81(9):3395–3405.
35. Khademian M, Imlay JA. 2021. How microbes evolved to tolerate oxygen. *Trends in Microbiology*. 29(5):428–40.

36. Kirschvink JL, Gaidos EJ, Bertani LE, Beukes NJ, Gutzmer J, et al. 2000. Paleoproterozoic snowball Earth: Extreme climatic and geochemical global change and its biological consequences. *Proceedings of the National Academy of Sciences*. 97(4):1400–1405.
37. Krute CN, Krausz KL, Markiewicz MA, Joyner JA, Pokhrel S, et al. 2016. Generation of a stable plasmid for in vitro and in vivo studies of *Staphylococcus* species. *Applied and Environmental Microbiology*. 82(23):6859–69.
38. Lu Y, Yeung N, Sieracki N, Marshall NM. 2009. Design of functional metalloproteins. *Nature*. 460(7257):855–62.
39. Lynch M, Kuramitsu H. 2000. Expression and role of superoxide dismutases (SOD) in pathogenic bacteria. *Microbes and infection / Institut Pasteur*. 2(10):1245–55.
40. Lyons TW, Reinhard CT, Planavsky NJ. 2014. The rise of oxygen in Earth's early ocean and atmosphere. *Nature*. 506(7488):307–15.
41. Malone CL, Boles BR, Lauderdale KJ, Thoendel M, Kavanaugh JS, Horswill AR. 2009. Fluorescent reporters for *Staphylococcus aureus*. *Journal of Microbiological Methods*. 77(3):251–60.
42. Martin ME, Byers BR, Olson MO, Salin ML, Arceneaux JE, Tolbert C. 1986. A *Streptococcus mutans* superoxide dismutase that is active with either manganese or iron as a cofactor. *Journal of Biological Chemistry*. 261(20):9361–67.
43. Massey V, Strickland S, Mayhew SG, Howell LG, Engel PC, et al. 1969. The production of superoxide anion radicals in the reaction of reduced flavins and flavoproteins with molecular oxygen. *Biochemical and Biophysical Research Communications*. 36(6):891–97.
44. McCord JM, Fridovich I. 1969. Superoxide Dismutase: An enzymic function for erythrocuprein (hemocuprein). *Journal of Biological Chemistry*. 244(22):6049–55.
45. Messner KR, Imlay JA. 1999. The identification of primary sites of superoxide and hydrogen peroxide formation in the aerobic respiratory chain and sulfite reductase complex of *Escherichia coli*. *Journal of Biological Chemistry*. 274(15):10119–28.
46. Miller A-F. 2004. Superoxide dismutases: active sites that save, but a protein that kills. *Current Opinion in Chemical Biology*. 8(2):162–68.
47. Miller AF. 2012. Superoxide dismutases: Ancient enzymes and new insights. *FEBS Letters*. 586(5):585–95.

48. Mizuno K, Whittaker MM, Bächinger HP, Whittaker JW. 2004. Calorimetric studies on the tight binding metal interactions of *Escherichia coli* manganese superoxide dismutase. *Journal of Biological Chemistry*. 279(26):27339–44.
49. Moore EK, Jelen BI, Giovannelli D, Raanan H, Falkowski PG. 2017. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nature Geosci*. 10(9):629–36.
50. Osman D, Martini MA, Foster AW, Chen J, Scott AJP, et al. 2019. Bacterial sensors define intracellular free energies for correct enzyme metalation. *Nat Chem Biol*. 15(3):241–49.
51. Peers G, Price NM. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature*. 441(7091):341–44.
52. Radin JN, Kelliher JL, Párraga Solórzano PK, Kehl-Fie TE. 2016. The two-component system ArlRS and alterations in metabolism enable *Staphylococcus aureus* to resist calprotectin-induced manganese starvation. *PLoS Pathogens*. 12(11):1–23.
53. Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc*. 2(8):1896–1906.
54. Reyes-Caballero H, Campanello GC, Giedroc DP. 2011. Metalloregulatory proteins: Metal selectivity and allosteric switching. *Biophysical Chemistry*. 156(2–3):103–14.
55. Rigby KM, DeLeo FR. 2012. Neutrophils in innate host defense against *Staphylococcus aureus* infections. *Seminars in Immunopathology*. 34(2):237–59.
56. Sheng Y, Abreu IA, Cabelli DE, Maroney MJ, Miller AF, et al. 2014. Superoxide dismutases and superoxide reductases. *Chemical Reviews*. 114(7):3854–3918.
57. Stallings WC, Powers TB, Patridge KA, Fee JA, Ludwig ML. 1983. Iron superoxide dismutase from *Escherichia coli* at 3.1-Å resolution: a structure unlike that of copper/zinc protein at both monomer and dimer levels. *Proceedings of the National Academy of Sciences*. 80(13):3884–88.
58. Strzepek RF, Harrison PJ. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature*. 431(7009):689–92.
59. Valderas MW, Hart ME. 2001. Identification and characterization of a second superoxide dismutase gene (*sodM*) from *Staphylococcus aureus*. *Journal of Bacteriology*. 183(11):3399–3407.

60. Vance CK, Miller A-F. 1998. A simple proposal that can explain the inactivity of metal-substituted superoxide dismutases. *J. Am. Chem. Soc.* 120(3):461–67.
61. Vance CK, Miller A-F. 2001. Novel insights into the basis for *Escherichia coli* superoxide dismutase's metal ion specificity from Mn-substituted FeSOD and its very High E_m. *Biochemistry*. 40(43):13079–87.
62. Waldron KJ, Robinson NJ. 2009. How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews Microbiology*. 7(1):25–35.
63. Waldron KJ, Rutherford JC, Ford D, Robinson NJ. 2009. Metalloproteins and metal sensing. *Nature*. 460(7257):823–30.
64. Whittaker JW. 2010. Metal uptake by manganese superoxide dismutase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 1804(2):298–307.
65. Whittaker MM, Whittaker JW. 1998. A glutamate bridge is essential for dimer stability and metal selectivity in manganese superoxide dismutase. *Journal of Biological Chemistry*. 273(35):22188–93.
66. Wright Valderas M, Gatson JW, Wreyford N, Hart ME. 2002. The superoxide dismutase gene *sodM* is unique to *Staphylococcus aureus*: Absence of *sodM* in coagulase-negative staphylococci. *Journal of Bacteriology*. 184(9):2465–72.

CHAPTER 3¹:

Mobilization of vitamin B₁₂ transporters alter competitive dynamics in a human gut microbe

3.1 Summary

The functional and genomic diversity of the human gut microbiome is shaped by horizontal transfer of mobile genetic elements (MGEs). Characterized MGEs can encode genes beneficial for their host's self-defense (e.g., antibiotic resistance) or ability to compete for essential or limited resources (e.g., vitamins). Vitamin B₁₂ and related compounds (corrinoids) are critical nutrients that enable colonization by members of the common gut microbe phylum, the Bacteroidetes. Herein, we identify a distinct class of MGEs in the Bacteroidetes, responsible for the mobilization and exchange of the genes required for transport of corrinoids, a group of cyclic tetrapyrrole cofactors including vitamin B₁₂ (*btuGBFCD*). This class includes two distinct groups of conjugative transposons (CTns) and one group of phage. Conjugative transfer and vitamin B₁₂ transport activity of two of the CTns was confirmed *in vitro* and *in vivo*, demonstrating the important role MGEs play in distribution of corrinoid transporters in the Bacteroidetes.

¹This chapter has been published in *Cell Reports*. Minor modifications have been made to the text to meet formatting requirements for the dissertation. Copyright © 2021 Katie A. Frye, Varadh Piamthai, Ansel Hsiao, and Patrick H Degnan. Citation: Frye KA, Piamthai V, Hsiao A, Degnan PH. 2021. Mobilization of vitamin B₁₂ transporters alters competitive dynamics in a human gut microbe. *Cell Reports*. 37(13):110164. PHD performed work for and contributed to production of Figures 3.11 and 3.10 D-G. Danielle Campbell performed the sequencing for the *Bacteroides thetaiotaomicron* VPI-C11-15 and *Bacteroides dorei* VPI-6598B genomes in Table 3.11.

3.2 Introduction

The Bacteroidetes are prominent members of the human gut microbiota that comprise as much as 80% of the microbiome in some individuals (61). Importantly, the Bacteroidetes play roles in different aspects of human health including metabolism, obesity, inflammatory bowel disease, development of the immune system, and protection against colonization by invading pathogens (7, 31, 33, 34, 37, 49, 54, 58, 68, 77). To colonize the human gut, microbes must compete for essential resources, such as vitamins. This includes vitamin B₁₂, and related compounds termed corrinoids (21, 26). As a cofactor, corrinoids are critical in many enzymatic reactions, including those involved in amino acid synthesis, nucleotide synthesis, and metabolism (12, 48). As such, corrinoid transporters are key for colonization by a majority of gut microbes (19, 26).

Microbially produced corrinoids can be abundant in the large intestines and are structurally diverse with as many as eight distinct corrinoids found in a single patient sample (1). This structural diversity of corrinoids is matched by corrinoid transporter sequence diversity (19). Microbes with distinct corrinoid cofactor preferences can either possess specific transport capabilities, specific synthesis or remodeling capabilities, or a combination of these to accommodate their needs (78). Computational analyses revealed that of the 80% of gut microbes that encode corrinoid dependent genes or riboswitches, only 20% encode complete, *de novo*, biosynthesis pathways without any ability to transport corrinoids (19). Therefore, most gut microbes must rely on transport to obtain and/or remodel these essential cofactors from dietary or microbial sources.

Corrinoid transport loci in the Bacteroidetes have components homologous to those in *Escherichia coli* including an outer membrane transporter (BtuB), periplasmic binding protein (BtuF), and inner membrane ATP-dependent transporters (BtuC/D) (22, 38). Unlike *E. coli*, Bacteroidetes transport operons also include a novel family of surface-exposed lipoproteins (BtuG), critical for binding extracellular corrinoids and essential for efficient transport via BtuB (75). Furthermore, many of the Bacteroidetes encode multiple corrinoid transport loci, with the type strain *B. thetaiotaomicron* VPI-5482 encoding three homologous loci, Locus 1, Locus 2, and Locus 3 (19). Locus 1 encodes only the outer membrane transporter and lipoprotein, *btuGB*, while Locus 2 and 3 encode complete *btu* transport loci (*btuGBFCD*). These homologous loci are not redundant and, in fact, display distinct corrinoid preferences that confer competitive advantages depending upon the corrinoid present (19). Previous work comparing the corrinoid preference specifying outer membrane transporter (BtuB) phylogeny with the species phylogeny revealed that among the Proteobacteria, the BtuB and species phylogeny are similar, whereas the Bacteroidetes BtuB and species phylogenies are incongruent (19). This suggested that there has been duplication, divergence, and horizontal gene transfer (HGT) of these alleles among the Bacteroidetes.

The possibility of spread of these corrinoid transporters via HGT is likely as there is widespread evidence for rampant HGT within the gut (16, 28, 42, 56, 57). In gut Bacteroidetes, conjugative transposons (CTns) and phage are common and are major mechanisms for HGT. This has resulted in substantial

genomic diversity at both the species and strain level (57). When these mobile genetic elements (MGEs) are transferred, part of their success (e.g., persistence) in a new host depends upon their contribution to their host's fitness. For instance, a single MGE, CTnDOT, carries antibiotic resistance genes for erythromycin (*ermF*), tetracycline (*tetQ*, *tetX*), and streptomycin (*aads*). Other Bacteroidetes CTns encode gene products that have alternate means of increasing the fitness of their host by altering cell attachment or engaging in interbacterial antagonism (16) and at least one Bacteroidetes phage can alter bile acid metabolism (9).

In this study, we have identified a distinct class of MGEs including putative CTns and a phage within gut Bacteroidetes that encode putative corrinoid transport genes. Here, we have experimentally verified both the ability for two of the CTns to transfer to naïve hosts and the functionality of the MGE-encoded corrinoid transport locus to transport vitamin B₁₂. Furthermore, these CTns exhibit different fitness advantages when competing against the previously characterized *B. thetaiotaomicron* chromosomally-encoded *btu* loci depending on which loci they are competing against. Finally, we determined that the CTns are capable of transferring within the murine intestine, suggesting these elements can be shared in the human gut and contribute to the diversity of corrinoid acquisition loci found in gut Bacteroidetes. Together, these data expose a distinct class of MGEs within the Bacteroidetes that encode and disseminate transport loci for an essential resource required for colonization by these prominent members of the human gut microbiome.

3.3 Results

3.3.1 Distinct corrinoid transporters are encoded on three genetically divergent MGE families. Our previous work on chromosomally-encoded corrinoid transport loci within *B. thetaiotaomicron* revealed a discordance between the Bacteroidetes species and BtuB phylogenies raising the possibility of horizontal gene transfer of *btuB* alleles among the Bacteroidetes (19). To determine the role of mobile genetic elements (MGEs) in *btuB* allele distribution, we examined 133 gut Bacteroidetes genomes and identified 19 putative MGEs encoding putative corrinoid transport loci (Table 3.1, 3.3, 3.4). Phylogenetic reconstruction of this expanded set of *btuB* alleles determined that all MGE-encoded BtuBs group within the previously delineated BtuB “Family N” (Figure 3.1A, 3.2) (19). Furthermore, the majority of MGE-encoded BtuBs are clustered on the phylogeny and are more similar to the *B. thetaiotaomicron* VPI-5482 chromosomally-encoded BtuB2 rather than BtuB3 (Figure 3.1A, 3.2). Despite BtuB2 and BtuB3 both belonging to Family N, they have been previously demonstrated to have distinct corrinoid preferences (19).

In order to understand the relatedness of these corrinoid transport-encoding MGEs, the nucleotide similarity over the entirety of each of the elements was used to calculate pairwise percent length aligned (PLA) and these elements were clustered with the Markov Cluster Algorithm (Figure 3.1B) (25, 66). This led to the generation of three distinct “families” of predicted corrinoid transport-encoding MGEs (Family 1 – 3, Figure 3.1B). We note that the well-characterized Bacteroidetes CTn, CTnDOT, was included in these relatedness

analyses and found to have similarity to one of the corrinoid transport-encoding MGEs, however, given its lack of a corrinoid transport locus it was not included in Family 2 (Figure 3.1B, 3.3). We also note that this limited similarity of Family 2 mating pair formation and DNA transfer replication genes also extends to CTnBST, CTnHyb and other members of the CTnDOT/ERL family of antibiotic resistant conjugative transposons (32, 74).

In general, the predicted corrinoid transport-encoding MGEs grouped into families that mirror the BtuB phylogeny (Figure 3.1, 3.2). All of the MGE-encoded BtuB alleles in Family 1 (except for *B. plebeius*) and Family 3 group separately from the chromosomally-encoded BtuB2 and BtuB3 in *B. thetaiotaomicron* VPI-5482. The *B. plebeius* and Family 2 MGEs, however, encode BtuB alleles that are nearly identical to BtuB2 (*B. plebeius* BtuB: 95% identity, *B. stercoris* and *B. fingoldii* BtuB: 100% identity; Figure 3.1A, 3.2). In addition, examination of the predicted integrases for MGE Families 1 and 3 reveals a high degree of conservation within each family. This conservation, however, is not detected in the Family 1 MGE from *B. plebeius* and the Family 2 MGEs. The *B. plebeius* MGE encodes an integrase that is greater than 80% similar to the one encoded by the Family 2 *B. fingoldii* MGE, sharing far less (67% – 71%) homology to the integrases encoded by the other members of their respective MGE Families (Figure 3.1B).

Annotation and analysis of the protein domains (PFAMs) and families (TIGRFAMs) present within the predicted MGEs allowed us to determine that Family 1 and 2 elements are likely CTns while the Family 3 elements are likely

integrated phage (Tables 3.5, 3.6). Alignments of the complete set of Family 1 MGEs exhibits extensive homology over the entire length of the MGEs, however, broad diversity within the Family 1 MGEs indicates they are related but unlikely to represent a single, recently transferred element (Figure 3.4). In contrast, 3 of the 5 predicted phages within Family 3 are nearly identical over their entire lengths suggesting recent transfer of the phage among divergent bacterial host species (Figure 3.5). The remaining members, *Tannerella* sp. 6_1_58_FAA_CT1 and *B. coprophilus* DSM 18228, exhibit large 48kb and 19kb replacements, respectively, as well as a large 69kb deletion in *B. coprophilus* DSM 18228 suggesting that one or both of these are inactivated prophage (Figure 3.5). Although the MCL clustering grouped *B. fingoldii* and *B. stercoris* into Family 2, this relationship is largely driven by their extended homology surrounding the *btu* locus (Figure 3.3). It is likely that with further investigation and identification of corrinoid transport-encoding CTNs, this family may be subdivided. It is notable that extended homology (43kb) is observed surrounding the *btu* locus between *B. thetaiotaomicron* VPI-5482 chromosomally-encoded *btu* Locus 2 and the *btu* locus in the *B. stercoris* predicted MGE (Figure 3.3). This further supports the role of HGT and CTNs in the distribution *btu* alleles in diverse gut microbes (19).

3.3.2 Family 1 corrinoid transport-encoding CTNs are mobilizable and transport vitamin B₁₂. Conjugation experiments were carried out with two MGE Family 1 representative strains, *B. ovatus* ATCC 8483 (*B. ovatus*) and

B. thetaiotaomicron 7330 to determine if these predicted elements are capable of transferring via conjugation. These CTn containing strains acted as donors and an erythromycin resistant triple corrinoid transport locus mutant of *B. thetaiotaomicron* as the recipient (*B. thetaiotaomicron* $\Delta\Delta\Delta$ Erm^R, Table 3.2). In minimal medium supplemented with vitamin B₁₂ (MM+B₁₂), *B. thetaiotaomicron* $\Delta\Delta\Delta$ is a methionine auxotroph as it can no longer import corrinoids (19, 67). Therefore, successful transconjugants, but not original recipients or donors, will grow on MM+B₁₂ + Erm. This is in fact what we observed, isolating putative transconjugants derived from both *B. ovatus* and *B. thetaiotaomicron* 7330 predicted CTn-encoding donor strains (Figure 3.6A, Table 3.7). Conjugation frequencies for both strains were calculated by taking the average of the transconjugant (MM+B₁₂) CFUs divided by the CFUs of the transconjugants and recipients (MM+Met) from each experiment. We found that the transfer frequency of the BtCTn (1.02×10^{-5}) was nearly 3 times greater than that of the BoCTn (2.83×10^{-5}) (Figure 3.6A, Table 3.7).

To determine if the isolated transconjugants from the initial experiment (*B. thetaiotaomicron* $\Delta\Delta\Delta$ Erm^R + BoCTn or BtCTn) encoded the entirety of the required transfer machinery, these strains were used as donors and *B. thetaiotaomicron* $\Delta\Delta\Delta$ Tet^R as the recipient in a second round of conjugation experiments. We obtained successful secondary transconjugants for both CTns, indicating the acquisition of all the necessary transfer machinery by the initial transconjugants (Figure 3.6B, Table 3.7). Notably, in the second round of

transfers the CTn transfer frequency of the BoCTn was ~16-times greater than the transfer frequency of the BoCTn from its original host whereas the secondary transfer frequency of the BtCTn was about the same as the original. Correspondingly, the transfer frequency of the BoCTn from the new transconjugants to *B. thetaiotaomicron* $\Delta\Delta\Delta$ Tet^R was ~42-times greater than that of the BtCTn (Figure 3.6B, Table 3.7).

Although transconjugants were isolated on MM+B₁₂ plates, we confirmed the ability of the corrinoid transport locus on the CTn to be expressed, translated, and functional for importing vitamin B₁₂ in their new hosts by assessing growth rate of donors, recipients, and representative transconjugants in MM+B₁₂ and MM+Met. Transconjugants were found to grow at similar rates as the donors in both media types and the original *B. thetaiotaomicron* $\Delta\Delta\Delta$ recipients only grew when methionine was supplemented (Figure 3.6C).

Many Bacteroidetes CTns have relatively broad host ranges (4, 52, 64). To test this, we introduced a Tet^R cassette into BoCTn and BtCTns and attempted conjugations with several Tet^S recipient gut microbes. Overall, we found that the BoCTn and BtCTns were able to transfer to *B. uniformis* ATCC 8492 and *B. vulgatus* ATCC 8482 strains but not *B. fragilis* NCTC9343 and *B. caccae* ATCC 43185 (Table 3.7). There could be a variety of factors inhibiting the transfer of the CTns to *B. fragilis* and *B. caccae* including CRISPR systems, restriction modification systems, masking or occupation of the attachment site, culture conditions, etc. that remain to be explored.

3.3.3 BoCTn and BtCTn have conserved conjugation and integration

mechanisms. As an initial confirmation of CTn insertion, pulsed-field gel electrophoresis (PFGE) and Southern hybridizations were performed. PFGE of recipients, donors, and representative transconjugants clearly revealed an additional band in the transconjugants as compared to the recipient strain (Figure 3.7A). Southern hybridizations of the gels using a biotinylated probe specific for the outer membrane corrinoid transporter, *btuB*, consistently found that the extra bands in the gel hybridized with the *btuB* specific probe (Figure 3.7B). In the majority of cases, the additional, *btuB*-containing band appeared in the same location on the gel for both the BoCTn and BtCTn (~100 kb), raising the possibility that these CTns use the same attachment site. In a small number of cases (7/32), the additional band appeared at a different molecular size or two additional bands were observed, raising the possibility of a second attachment site and/or multiple integrations (see *B. ovatus* transconjugant 1 Figure 3.7A/B).

Whole genome sequencing was performed on 11 representative transconjugants encoding the BoCTn and 11 encoding the BtCTn (Table 3.9). These sequences were used to confirm and refine the initial predictions. Pairwise comparison of BoCTn and BtCTn indicate that these CTns share a common ~80 kb backbone but encode several variable regions (Figure 3.7C). As part of their conserved backbone, the CTns share integrase, mobilization, and transfer genes that are 99-100% conserved at the amino acid level (Figure 3.7C). Despite both the BoCTn and BtCTn *BtuB* alleles nesting within the same phylogenetic clade (Figure 3.1A), the *BtuGBFCD* proteins are on average only 95% similar to one

another and encode varied gene content immediately up and downstream. An analysis of conserved protein domains in the variable CTn regions identified 1 unique TIGRFAM (in BoCTn) and 31 unique PFAMs, 26 of which were detected in BtCTn which is 16 kb longer than the BoCTn (Table 3.10, 3.11). The additional DNA in BtCTn includes a mobilizable restriction modification system integron (12 kb) and an operon possibly involved in lipid transport and methylation (6 kb) (Table 3.8, 3.9, 3.10, 3.11). BoCTn encodes a putative threonine/serine exporter downstream of its *btuGBFCD* operon and several additional genes of unknown function (Table 3.11).

To determine an attachment site (*attB*) for the CTns, the CTn boundaries within the sequenced contigs were aligned with the recipient *B. thetaiotaomicron* strain using Mauve (17). This identified 18 nucleotide *attL* and *attR* direct repeat sequences which were nearly identical to the putative *attB* in *B. thetaiotaomicron* (ACTAAATTTACCAGATTT, Figure 3.8). Confirmation of integration of the CTns at the main *attB* in the intergenic spacer between BT4140 and BT4141 was attempted in 153 of 289 transconjugants via a negative PCR across the main attachment site (*attB*). Integration at this site was PCR confirmed in 118 of the 153 tested. Four of the genome sequenced transconjugants were determined to have integrated in different locations in the *B. thetaiotaomicron* chromosome. Two were identified between BT2505 and BT2506, and one each within BT4088 and BT4227 (Table 3.9). These alternate attachment sites differed by 3, 3, and 5 bp respectively from the primary *attB*, mostly clustered near the 5' end of the sequence (Figure 3.8), indicating some flexibility in the integrase recognition site

specificity. We then examined the remaining Family 1 MGEs for evidence of homologous *attL* and *attR* sites given they encode nearly identical integrases (97% – 100% amino acid identity). This revealed that nearly all Family 1 MGEs encode attachment site sequences that are nearly identical to the attachment sites identified in the BoCTn and BtCTn transconjugants with limited nucleotide differences occurring near the 5' end of some of the *attL* sequences (Figure 3.8). However, even with a high degree of conservation of the *attL/attR* sites among the Family 1 MGEs, there is not conservation in the genomic context that these MGEs integrate within. This differs from what is seen with some other MGEs where integration of the element occurs near a tRNA (8). Regardless, possible *att* sites were detected in the majority of Bacteroidetes genomes analyzed (1 mismatch allowed, 112/133 genomes). As expected, the Family 2 and Family 3 corrinoid transport MGEs, as well as the Family 1 *B. plebeius* MGE, encode divergent integrases and all encode correspondingly distinct predicted attachment sites (Figure 3.1B, Table 3.1, 3.3).

3.3.4 CTn-encoded corrinoid transport loci provide distinct competitive advantages. Previous work on chromosomally-encoded corrinoid transport loci within *B. thetaiotaomicron* revealed that, despite sequence and structural homology, each of the three corrinoid transport loci in *B. thetaiotaomicron* have distinct corrinoid preferences (19). These preferences occur even between BtuB2 and BtuB3, which are 73% identical at the amino acid level. Of the three chromosomal corrinoid transport loci, Locus 2, encoding BtuB2, is the most competitively advantageous both *in vitro* as well as in a

mouse model (19). Therefore, similar competition assays were carried out to determine the relative fitness of strains encoding the corrinoid transport-encoding CTns compared to the chromosomally-encoded corrinoid transport alleles.

B. thetaiotaomicron corrinoid transport mutant strains encoding the BoCTn, BtCTn, or individual chromosomal corrinoid transport loci were marked with unique oligonucleotide barcodes and competed (Table 3.2, Figure 3.9) (41). *In vitro* competitions were carried out as in Degnan *et al.* (19) comparing growth of strains over 5 days of serial transfer in both competitive (MM+B₁₂) and control conditions (MM+Met). As expected, when the CTn containing strains were competed against the triple locus deletion, *B. thetaiotaomicron* $\Delta\Delta\Delta$ BC, both CTn containing strains completely outcompeted this strain defective in vitamin B₁₂ transport in competitive but not control conditions (Figure 3.9, 3.10). However, the competitive abilities of the CTn containing strains when paired head-to-head with the individual chromosomal corrinoid transport loci were not as predictable. We determined that both CTn containing strains were outcompeted by both the Locus 1 and Locus 2 corrinoid transport loci. Although we note that strains with the BoCTn are greatly outcompeted by *B. thetaiotaomicron* $\Delta\text{locus1}\Delta\text{locus3}$ BC (99:1) while strains with the BtCTn are closer to a 50:50 ratio (Figure 3.9). In contrast, both the *B. thetaiotaomicron* $\Delta\Delta\Delta$ + BoCTn BC and *B. thetaiotaomicron* $\Delta\Delta\Delta$ + BtCTn BC strains greatly outcompete *B. thetaiotaomicron* $\Delta\text{locus1}\Delta\text{locus2}$ BC, which only has the chromosomal Locus 3 intact (Figure 3.9). Competitive performance of the CTn strains in the control medium was variable with strains containing the BoCTn being outcompeted,

regardless of which chromosomal corrinoid transport locus mutant they were competed against while strains containing the BtCTn greatly outcompeted *B. thetaiotaomicron* $\Delta locus1\Delta locus3$ BC strains, were about equal with *B. thetaiotaomicron* $\Delta BT1953\Delta BT2094$ BC strains, and were outcompeted by *B. thetaiotaomicron* $\Delta locus1\Delta locus2$ BC strains (Figure 3.10).

The competitive differences between the CTn and chromosomally-encoded loci cannot be wholly explained by differences in individual growth rates as barcoded strains were assessed for their ability to grow in both the competitive and control media and no significant growth defects were identified (Figure 3.10C). We note that CTn transfer during the competition would have an effect on the interpretation of these results, however, this possibility was assessed by attempting liquid vs. solid phase conjugation assays. As reported elsewhere, CTn transfer requires cell-to-cell contact (2, 35, 53, 60) and we found it exceedingly rare ($\sim 10^{-8}$) in liquid cultures and therefore it is unlikely that transfer of the CTns was occurring in the course of these competitions.

Although successful conjugation is rare, CTns are still able to excise and re-integrate into the chromosome when cultured in liquid medium. Using qPCR, we determined that the relative quantity of the integrated and excised forms of BtCTn and BoCTn are not different during mid-log growth in competitive (MM+B₁₂) and control conditions (MM+Met) (Figure 3.10D & E). We did find that

the excised BtCTn copy number is ~1.8 times greater than BoCTn (Figure 3.10F & G) but it is unclear if this has any influence on the major competitive differences observed with the *B. thetaiotaomicron* $\Delta locus1\Delta locus3$ BC strain (Figure 3.9).

3.3.5 BoCTn and BtCTn mobilize in the mammalian intestinal tract.

MGE transfer is widespread within the gut, suggesting that the gut acts more as a structured environment and less like a liquid culture (16, 80). Therefore, we endeavored to determine if the BoCTn and BtCTn corrinoid transport-encoding elements could mobilize *in vivo*. To this end, we pre-colonized Germ-free mice with excess of the recipient *B. thetaiotaomicron* $\Delta\Delta\Delta$ Erm^R strain, which is normally outcompeted by strains competent in corrinoid transport. We then gavaged the mice with equal portions of donor strains *B. thetaiotaomicron* $\Delta\Delta\Delta$ + BoCTn_ Tet^R and *B. thetaiotaomicron* $\Delta\Delta\Delta$ + BtCTn_ Tet^R and plated fecal pellets to detect transconjugants (Erm^R + Tet^R). After three days, no transconjugants (recipient cells (Erm^R) that had gained a Tet^R marked CTn (BoCTn_ Tet^R or BtCTn_ Tet^R)) were observed (Figure 3.11). We next hypothesized that MGE transfer occurs in a density dependent fashion within the gut, with the initial inoculum of $\sim 3.2 \times 10^5$ CFUs for the donor cells being too low to support element transfer, especially as the level of donor cells dropped to $\sim 2.6 \times 10^2$ on Days 1-3 (Figure 3.11). This drop in the donor population suggests that despite the fact that *B. thetaiotaomicron* $\Delta\Delta\Delta$ Erm^R lacks functional corrinoid transport, the initial

establishment of this strain within the host likely provided it with colonization resistance (40). Therefore, on Day 4 the mice were gavaged a second time with $\sim 3.2 \times 10^{10}$ CFUs of equal portions of the two donor strains.

Immediately following the second gavage, transconjugants encoding either the BoCTn or BtCTn were isolated from the mouse gut (Figure 3.11). Once detected, the abundance of the transconjugants remained fairly steady, showing a slight increase over the course of the experiment. The total transfer frequencies for these CTNs *in vivo* ranged from $\sim 9.7 \times 10^{-9}$ on Day 5 to the highest on Day 9 of $\sim 1.0 \times 10^{-7}$ (Figure 3.11). However, calculating precise transfer frequencies is complicated by the presence of both BoCTn and BtCTn transconjugants and the inability to distinguish unique transconjugant events from the maintenance of already established transconjugants. We note that these *in vivo* conjugation rates are much lower than those observed for the BoCTn and BtCTn *in vitro* (Figure 3.6A,B). As efficient conjugative transfer heavily relies on physical contact it is possible that the structured gut environment, gastric motility, the host immune system, or other factors may have contributed to the lower transfer rates observed (2, 35, 53, 60). We expect that new transconjugants have equal fitness with the donor strains. Therefore, given the 10-fold increase in transconjugants and donors during the final 7 days of the experiment, if continued we would predict the corrinoid transport deficient recipients would eventually be replaced by donors and transconjugants.

Following isolation, a total of 92 transconjugants from days 5, 7, 9 and 12 were confirmed by PCR for the presence of BoCTn (n=59) or BtCTn (n=33) and

the site-specific Erm^R insertion. A subset of 24 of these isolates were further confirmed for successful growth in MM+B₁₂ (Figure 3.12). These results demonstrate that these predicted corrinoid transport-encoding CTNs are functional in both transfer and restoration of growth on corrinoid species *in vivo*. Furthermore, this suggests that these elements play a key role in re-assorting corrinoid transporters and influencing competition and colonization in the mammalian gut.

3.4 Discussion

The human gut microbiome is incredibly important to human health and its functional diversity is shaped by MGEs. While many characterized MGEs increase their hosts' fitness through the spread of antibiotic resistance (3, 27, 51, 57, 63, 71–74) and antagonistic type 6 secretion systems (15, 16) any gene that is beneficial to a host's survival and colonization should facilitate the spread of an MGE within a population (46, 62). In this study, we identified 19 putative MGEs integrated into human gut *Bacteroidetes* genomes that encode putative corrinoid transport loci homologous to those we previously described in *B. thetaiotaomicron* VPI-5482 that are critical for colonization and competition in the mammalian gut (Table 3.1) (19, 26). We were able to classify these MGEs into three distinct families, two that are comprised of CTNs and the third are putative phage (Figure 3.1). Furthermore, we were able to demonstrate the ability of two of the CTNs to mobilize and functionally complement vitamin B₁₂ transport (Figure 3.6). Finally, conjugation was also detected by both CTNs during

colonization of a gnotobiotic mouse model, confirming the role of these MGEs in disseminating corrinoid transporters and influencing competition and colonization in the mammalian gut (Figure 3.11).

Corrinoids are required small molecule cofactors for the majority of sequenced gut microbes (19, 79). While some gut microbes are capable of synthesizing these cofactors *de novo*, a large proportion rely on transporters to access these critical molecules (19). Our previous analysis of corrinoid transport gene clusters among diverse Bacteroidetes led us to speculate that part of the diversity and distribution of these loci was due to horizontal gene transfer among these species (19). In examining the genetic context of the 33 Family N BtuB alleles previously identified, we found that 11 were closely flanked by genes with strong homology to CTn and phage protein domains. Subsequent comparisons with additional gut microbial genomes led to the refinement of the MGE boundaries and identified eight more predicted MGEs encoding additional Family N BtuB alleles. Like the previously reported chromosomal loci, all of the MGE-encoded BtuBs were localized in operons preceded by a B₁₂ binding riboswitch and encoding the additional transport proteins BtuGFCD (19, 75).

Despite the similarity of the corrinoid transport loci, the MGEs encoding these loci were readily resolved into three distinct families (Figure 3.1). Given the abundance and importance of CTns in Bacteroidetes genomes, the identification of two CTn families responsible for mobilizing corrinoid transport loci was not surprising (16, 45, 57, 76). In contrast, the identification of a group of integrated prophage with corrinoid transport loci was at first unexpected, however, is

consistent with the ability of phage to encode genes that can directly contribute to their bacterial host's fitness (e.g., Shiga toxin, cholera toxin) (44, 69). We note that these MGEs are distributed among diverse species and strains isolated from a combination of both healthy and sick patients (e.g., ulcerative colitis, Crohn's disease) dating between 1933 and 2014 from the United States, Canada, and Japan (Table 3.3). This suggests the broad circulation and maintenance of these MGE families among human gut *Bacteroidetes* communities.

Alignments of Family 1 and Family 3 MGEs show broad synteny within these integrated element families (Figure 3.4, 3.5) while this synteny is considerably less in alignments of Family 2 MGEs (Figure 3.3). In Family 1, the synteny is disrupted in several cases by an isolated Group II intron insertion or an insertion sequence element. Furthermore, we detected four distinct putative restriction modification gene integrons, all of which are flanked by integrase family genes suggesting their independent acquisitions by these CTns. For example, in the case of the sequenced BtCTn transconjugants 5 of the 11 isolates spontaneously lost this integron (Table 3.9). In Family 3, three virtually identical phage were detected in three distinct *Bacteroides* species, however, all were isolated from separate patient biopsy samples collected in Calgary, Alberta, CA. A fourth representative found in a *Tannerella* species, also from a patient sample from the same research group, has an apparent non-orthologous replacement of 47 kb of phage DNA (Figure 3.5). This type of mosaicism is common among phage and can result in phage structural changes and/or permit infection of novel hosts (11, 13, 36). Of the 19 corrinoid transport-encoding

MGEs we identified in this study, only the final member of Family 3 has clear evidence of substantial genetic loss and rearrangement that would suggest it is no longer an active phage. However, bioinformatics alone cannot rule out the possibility that genomic regions are mobilizable given the presence of helper phage or CTNs (10, 23).

We experimentally characterized two of the predicted Family 1 CTNs, confirming their ability to independently transfer to and from *B. thetaiotaomicron* both *in vitro* as well as in a gnotobiotic animal model (Figure 3.6, 3.7, 3.11). Furthermore, we demonstrated that the CTN-encoded corrinoid transporters were capable of functionally complementing strains otherwise incapable of transporting vitamin B₁₂ (Figure 3.6C). Similar to our previous observations (19), although the BoCTN and BtCTN encoded BtuB fall in the same corrinoid transporter Family (Family N) as two of the chromosomally-encoded transporters in *B. thetaiotaomicron* VPI-5482, they exhibit distinct competitive abilities in acquiring vitamin B₁₂ (Figure 3.9). Indeed, the acquisition of BoCTN or BtCTN by a gut microbe only encoding a *btuGBFCD* operon like chromosomal Locus 3 would presumably provide it with a marked fitness improvement. Future experiments are still required to assess both if there are distinct preferences among the CTN-encoded transporters for other corrinoids and what factors (e.g., binding affinities, expression levels) contribute to these preferences.

Consistent with the diversity of hosts initially identified harboring a Family 1 CTN, we demonstrated that both BoCTN and BtCTN were capable of transferring to divergent *Bacteroides* species (*B. uniformis* and *B. vulgatus*)

(Table 3.7). We note that we were not able to detect transconjugants in all of the species tested (*B. fragilis* and *B. caccae*), however, there are likely a multitude of factors that could interfere with successful conjugation of these elements. For example, we've previously shown that gut colonization was critical for detecting transduction of the *Bacteroides* phage, BV01 (9). We find it likely that conjugation frequency and compatibility may also be modulated by growth in the intestinal environment. Additionally, we saw a marked increase in BoCTn transfer frequency during the secondary round of transfer when this MGE was transferring among *B. thetaiotaomicron* species as compared to when the MGE was transferring from its original *B. ovatus* host to *B. thetaiotaomicron* (Figure 3.6A, B, Table 3.7). This potentially indicates a slight species barrier that leads to a lower transfer frequency between more distantly related organisms. Interestingly, six of the 12 members of Family 1 appear to have integrated either within or immediately adjacent to another mobile-like region (Figure 3.4). Although nearly all of the Family 1 MGEs share a conserved attachment site (Figure 3.8), it is clear that genomic locations of these attachment sites are not conserved. Furthermore, it is unclear what effect, if any, the integration of the Family 1 CTn has on the activity of the nearby element. It is possible that the function of the nearby element is disturbed or that the elements are now capable of transferring as a single conjugative element.

In stark contrast to many of the characterized MGEs encoded by the Bacteroidetes, the MGEs identified in this study do not carry a tetracycline resistance gene, or any genes homologous to any known antibiotic resistance

genes, and have limited similarity to CTnDOT (3, 6, 27, 70, 71). These antibiotic resistance MGEs within the Bacteroidetes spread as a result of the widespread clinical use of tetracycline in the '80s and '90s (57). The MGEs in this study represent a different class of MGEs that do not contain antibiotic resistance cassettes, but rather loci that allow the host to gain or increase access to an essential metabolite (16). They are akin to the integrative and conjugative elements encoding polysaccharide utilization loci (PUL), one of which that is responsible for catabolizing algal polysaccharides and is common among individuals in Japan that regularly consume seaweed products (29). The success and spread of these MGEs in populations is expected if the corrinoid transporter expands the corrinoid repertoire of the host or encodes a more effective transporter, thereby increasing fitness of the host.

Together, this work has identified a distinct class of MGEs that encode corrinoid transporters and are widespread among gut Bacteroidetes. For these gut Bacteroidetes, the presence of corrinoid transporters is crucial as many are unable to synthesize this essential cofactor. Therefore, these MGE-encoded transporters have the potential to expand the corrinoid repertoire of their host, leading to a competitive advantage in the human gut where the ability to transport corrinoids has been shown to be essential for colonization. Future investigations into how MGEs contribute to the spread, maintenance, and distribution of corrinoid transporters among the Bacteroidetes are essential as they provide insight to novel factors that affect colonization, and thus composition, of the human gut microbiome.

3.5 Acknowledgements

We thank Alvaro Hernandez from the W. M. Keck Center for Comparative and Functional Genomics at UIUC for technical services; Jonathan Mitchell for maintenance and animal care at the UCR vivarium; Rachel Whitaker at UIUC for use of the PFGE equipment; and Samson Avena at UCR for assistance screening transconjugants. This research was supported by initial complement funding to P.H.D from UIUC and UCR and K.F. was supported by the Department of Microbiology at UIUC. Gnotobiotic mouse work and A.H. were supported by NIH/NIGMS grant R35GM124724 and NIH/NIAID grant R01AI157106, to A.H.

3.6 Methods

3.6.1 Resource availability

Lead Contact. Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Patrick Degnan (patrick.degnan@ucr.edu).

Materials Availability. Bacterial strains and plasmids generated in this study are available upon request to the Lead Contact, Dr. Patrick Degnan (patrick.degnan@ucr.edu).

Data and Code Availability. All NCBI Accession numbers for genomes used and sequenced in this study can be found in Table 3.4. DNA sequencing reads from the transconjugants generated in this study are deposited in NCBI under PRJNA715234. Information regarding these sequences can be found in

Table 3.8 and 3.9. New assembled genomes are also deposited in NCBI under PRJNA715234. The code used for the genome assembly is available here <https://github.com/phdegnan/A5ud> (DOI: 10.5281/zenodo.5712008). The code for the generation of pairwise percent length alignment (PLA) and reformatting these files for Markov Cluster algorithm analysis can be found here <https://github.com/IGBillinois/VICSIN> (DOI: 10.5281/zenodo.5138177). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

3.6.2 Experimental model and subject details

Bacterial Culturing and Genetic Manipulation. Bacterial strains used in this work can be found in Table 3.2 and 3.12. Routine culturing of *Bacteroides* species occurred anaerobically at 37°C in liquid tryptone-yeast extract-glucose (TYG, (30)) medium or Difco Brain Heart Infusion (BHI, Quad Five, Ryegate, MT) agar with the addition of 10% defibrinated horse blood as previously described (14). Anaerobic culturing was performed in a vinyl anaerobic chamber containing a 70% N₂, 20% CO₂, and 10% H₂ gas mixture (Coy Laboratory Products, Grass Lake, MI). *Escherichia coli* S17-1 lambda pir strains used for cloning and conjugation of suicide vectors were grown in Lysogeny Broth (LB) at 37°C aerobically with shaking (9, 14, 19). When needed, cultures were resuspended or grown in minimal medium (MM, (19, 41)). As noted, MM was supplemented with methionine (MM+Met, 500μM) or vitamin B₁₂ (MM+B₁₂, 0.4nM or 4nM) and agar (20 g/L) was added for plate preparation as needed. Antibiotics were added to

the medium when appropriate in the following concentrations: ampicillin, 100 µg/ml; gentamicin, 200 µg/ml; erythromycin, 25 µg/ml; tetracycline, 2µg/ml; and 5-fluoro-2'-deoxyuridine (FudR), 200 µg/ml.

All vectors and primers used in this study can be found in Tables 3.13 and 3.14, respectively. Insertion of the *tetQ* allele into BoCTn and BtCTn was accomplished via allelic exchange conjugated from donor *E. coli* S17-1 lambda pir strains into recipient *B. thetaiotaomicron* strains before being confirmed by PCR (14, 19, 39). Briefly, the ~1kb left and right flanks of the MGE regions surrounding BACOVA4685 (BoCTn) or Bthe2347 (BtCTn) were amplified via HiFi PCR (Kapa HiFi Taq MasterMix, Kapa Biosystems, Wilmington, MA) and combined by splicing overlap extension (SOE) PCR designed to insert BamHI and SpeI restriction sites in the middle of the region. The *tetQ* allele was also amplified via HiFi PCR. The purified SOE products and *tetQ* were digested with restriction enzymes, cloned into pExchange_tdk vector, and transconjugants screened by standard methods (9, 14, 39, 41). Barcoded pNBU2 vectors were constructed previously and were introduced into the genome in a single copy using methods similar to pExchange_tdk (14, 39).

Gnotobiotic Mice. All experiments using mice were performed using protocols approved by the University of California Riverside Institutional Animal Care and Use Committee. Germfree C57BL/6J mice were maintained in flexible plastic gnotobiotic isolators with a 12-hr light/dark cycle. Animals caged individually (n=1, female) or in pairs of litter mates (n=2 males, n=2 females) were provided with standard, autoclaved mouse chow (5K67 LabDiet, Purina)

ad libitum. Since there was no *a priori* reason to expect animal age to influence conjugation rates, animals were 11 weeks old at time of colonization. Individually antibiotic resistance marked bacterial strains were grown individually for ~20h in TYG medium with appropriate antibiotics and frozen at -80°C in anaerobic cryovials. Cell viability was tested by plating. Viable CFU counts were used to prep the recipient cells *B. thetaiotaomicron* $\Delta\Delta\Delta$ Erm^R for oral gavage on Day -2 at $\sim 3.2 \times 10^6$ CFUs per animal. On Day 0, equal amounts of two donor strains *B. thetaiotaomicron* $\Delta\Delta\Delta$ + BoCTn_Tet^R and *B. thetaiotaomicron* $\Delta\Delta\Delta$ + BtCTn_Tet^R were combined and gavaged at $\sim 3.2 \times 10^5$ CFUs. A second gavage of equal amounts of the two donor strains were given on Day 4 at a concentration of $\sim 3.2 \times 10^{10}$ CFUs per animal. Fecal samples were collected on days 0 prior to gavage of the donors followed by days 1, 3, 5, 7, 9, and 12 from each animal. Fecal pellets were processed by adding 500 μ l of TYG + 20% vol/vol glycerol to each tube and vigorously shaking in a beadbeater without beads for 1m 30s. Fecal slurries were spun down at 2,000 x g for 10s, followed by serial dilution on selective media (BHI+Tet+Gn, BHI+Erm+Gn, BHI+Erm+Tet+Gn) to determine CFUs. Animals were sacrificed on day 12 following final fecal collection. A maximum of 5 individual isolates per animal were struck out on BHI+Erm+Tet+Gn media, before being stocked at -80°C and prepped for gDNA and PCR screening. Isolates predicted to be successful transconjugants were screened as above (screening of isolates, and Liquid Growth Assays).

3.6.3 Method details

Identification of Predicted MGEs. Bacteroidetes genomes (Table 3.4) were screened for the presence of genomic regions encoding mobile-like genes using HMMER v3 (hmmer.org) with trusted cutoffs (--cut_tc) for PFAMs (El-Gebali *et al.*, 2019) and TIGRFAMs (Haft, 2001) (Table 3.5, 3.6). Boundaries for the predicted mobile regions were identified using a combination of BLASTn and Mauve (17) to compare strains containing the predicted element to those of the same, or closely related, species without the element. Predicted attachment sites (*attL* and *attR*) for the elements were determined by examination of alignments of the element boundaries for short, repeated sequences. Additionally, once attachment sites were determined for the BoCTn and BtCTn transconjugants sequenced in this study, Fuzznuc (47) was utilized to search for similar attachment sites in other Bacteroidetes genomes.

Relationships and Relatedness of Predicted MGEs. MGE-encoded BtuB alleles were initially detected by BLASTp using BtuB1, BtuB2, and BtuB3 from *B. thetaiotaomicron* VPI-5482 as queries. These BtuB alleles were then aligned with a subset of those detected previously (19) using MUSCLE (24) and subjected to maximum likelihood based phylogenetic reconstruction with RAxML on the CIPRES server (43, 59). BLASTp was also used to calculate protein similarity among the MGE-encoded integrases (e-value cutoff of 0.1).

Nucleotide sequences for each predicted MGE were aligned using Mauve (17) and compared pairwise using BLASTn with an e-value cutoff of 0.0001. These results were analyzed to calculate the percent length aligned (PLA), total

length aligned, and total bit score for each pair using Blast_to_MCL.1.py. PLA values below 0.2 were excluded and the Markov Cluster algorithm (MCL algorithm) was used to generate a relatedness network and family groupings of the predicted MGEs using MCLdump2clusters.py (25, 66). The network was visualized in Cytoscape (55).

Conjugation/Screening of Isolates with Bo/BtCTn. Cells were grown in TYG to exponential phase (0.3 – 0.6 O.D.₆₀₀) and then pelleted and washed in PBS. Donor and recipient strains were then combined in PBS before plating on BHI agar + 10% horse blood for 16 hours. Cell lawn was then resuspended in PBS and washed four times in MM. Cells were then plated on selective media (BHI + Ab (Erm or Tet) + Gent, MM+Met + Ab (Erm or Tet) + Gent, and MM+B₁₂ (4nM) + Ab (Erm or Tet) + Gent) and grown anaerobically for 16 – 72 hours.

Colonies that were isolated from conjugations after plating on selective media were then screened to confirm they were transconjugants. Cells were confirmed for their ability to grow on selective media (Erm^R or Tet^R depending on the identity of the recipient strain), the presence of a *btuB* in their genome (as detected by PCR, Primers in Table 3.14), the correct sequence of the *btuB* in their genome (confirmed by Sanger Sequencing), and their ability to grow in MM+B₁₂.

Liquid Growth Assays. Cells were grown in TYG to stationary phase and then pelleted and washed four times in MM (19). Cells were then normalized, and cultures were inoculated in a 96-well plate in triplicate at 0.003 O.D.₆₀₀ in TYG, MM+Met, and MM+B₁₂ (4nM). When performing the growth assays comparing

barcoded and non-barcoded strains (Figure 3.10C) growth was also assessed in MM+B₁₂ (0.4nM). Growth was then monitored every 30 minutes for 36 hours using a BioTek Synergy HTX Multi-Mode microplate reader in conjunction with a BioStack3 microplate stacker (BioTek, Winooski, VT, (14) or on a Clariostar microplate reader with a Stack III microplate stacker (BMG Labtech, Cary, NY). Doubling times were calculated from cells in early to mid-log phase (O.D.₆₀₀ of 0.01 – 0.12) using the least-squares fitting method (<http://www.doubling-time.com/compute.php?lang=en>).

PFGE and Southern Hybridizations. Whole-cell agarose plugs were made from overnight stationary cultures. Cultures were harvested and resuspended in TE Buffer (10 mM Tris, 1 mM EDTA, pH 8.0) before being combined 1:1 with 1.5% InCERT/LMP agarose. Once set, plugs were digested with Proteinase K (20mg/mL) before being washed and stored at 4°C in Low TE Buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) (20). Plugs were then digested with NotI before being run on a pulsed field gel electrophoresis (PFGE) rig for 24 hours to resolve bands 50-1000 kb. Conditions were as follows: 1% PFGE agarose; initial switch time, 24 seconds; final switch time, 2 min 30 seconds; Buffer, 0.5x Tris-borate-EDTA (TBE); V/cm, 6.0. Yeast Chromosomal PFG-Marker Mid-Range ladder was included for band size comparison (NEB). After the gel was visualized, the DNA was transferred overnight to a nitrocellulose membrane then probed overnight with a biotinylated, universal *btuB* specific probe at 42°C (Table 3.12). Detection was performed per manufacturer's protocol (Chemiluminescent Nucleic Acid Detection Module Kit, Thermo Fisher Scientific) using autoradiography film.

Whole Genome Sequencing. 22 representative transconjugant strains from eight independent conjugation experiments were selected for sequencing. In addition, 13 historical *Bacteroides* strains from the Virginia Polytechnic Institute Anaerobe Laboratory (VPI) and Wadsworth Anaerobe Laboratory (WAL) strain collections were sequenced. Sequencing was done as performed previously (9). Briefly, total high molecular weight gDNA was extracted as performed previously (9). This involves resuspension of overnight cell pellets in TE buffer (10mM Tris, 1mM EDTA) followed by lysis with sodium dodecyl sulfate (SDS; 0.07% w/v final) and proteinase K (300µg/mL final) and incubation at 55°C for 2 hours. Lysates were then mixed and centrifuged twice with phenol:chloroform:iso-amyl alcohol (25:24:1) (v/v/v), and the aqueous phase was retained and precipitated with 2 volumes of 100% EtOH and 1/10th volume of 0.3M sodium acetate at –20°C. DNA was pelleted with centrifugation at max speed for 30 min, then washed with 70% EtOH, dried and resuspended in TE buffer. DNA libraries were prepped with the Nextera XT Library Preparation Kit and Index Kit (Illumina, San Diego, CA), and the samples were pooled and submitted for sequencing on a HiSeq 2500 or HiSeq 5000 Illumina sequencer. Genomes were then assembled as performed previously (9) and mapped back to a *B. thetaiotaomicron* reference genome, as well as the genomic regions of *B. thetaiotaomicron* 7330 or *B. ovatus* encoding the CTn using BRESEQ v0.33.2 (18).

In vitro Competition Assays. Competitions between strains carrying unique oligonucleotide barcodes were performed as described previously (14, 19). Briefly, barcoded cells were grown for 16 hrs in TYG then pelleted and washed four times with MM. O.D.₆₀₀ values were normalized to 0.4 before being combined at a 1:1 ratio and inoculated 1:1,000 into an Axygen deep-well plate (Corning, Inc., Corning, NY) in triplicate for each conjugation pair in MM+Met and MM+B₁₂ (0.4nM). Cultures were incubated anaerobically at 37°C for 24 hrs and passaged 1:1,000 into the same medium they were previously grown in (14, 19, 39). Time points were taken every 24 hrs and samples were saved in medium plus 20% glycerol at -80°C in 96-well plates. Samples were prepared and analyzed using qPCR as previously described (5, 14) on a Bio-Rad CFX Connect instrument (Bio-Rad, Hercules, CA) with SYBR Fast MasterMix 2X Universal (Kapa Biosystems, Wilmington, MA). Briefly, this employed the HotSHOT 96-well plate method (65), which lyses 5μL of cultured cells in 50μL of the lysis buffer (25mM NaOH, 0.2mM EDTA) during a 30 min incubation at 95°C. A fraction of the lysate (16μL) is then combined with 144μL of 10mM Tris pH 8.0 and 16μL of neutralization buffer (40mM Tris-HCl). Each sample was run in technical triplicate using the manufacturer's instructions for the SYBR Fast Master Mix for each relevant set of barcode primers. A single touchdown PCR protocol was previously optimized (19) and used on the Bio-Rad CFX Connect instrument for all primer combinations: 95°C for 30 sec, 6 cycles of 95°C for 3 sec, 68°C for 20 sec (-1°C per cycle), 72°C for 1 sec, and 34 cycles of 95°C for 3 sec, 62°C for 20 sec, 72°C for 1 sec, followed by a melt curve starting at 65°C and ending at 95°C.

Mean strain quantities (in ng) were calculated using a standard curves derived from 10-fold dilution curves of 100ng – 0.1pg of gDNA for each barcoded *B. thetaiotomicron* strain and the efficiency-corrected ΔC_q method (5). Quantities were then transformed into percentages of each barcode within a sample and plotted.

Excision qPCR. *B. thetaiotaomicron* transconjugants with BtCTn and BoCTn were precultured in TYG and grown as above in the liquid growth assays in 5mL of MM+Met and MM+B₁₂ (4nM) and biological triplicate until they reached mid-log phase. Cells were recovered by centrifugation for 10 minutes at 3,000 x *g* and total genomic DNA was isolated as performed for genomic sequencing above. Samples were analyzed in technical triplicate with qPCR as described above using primer pairs for the ‘empty’ bacterial attachment site (*attB*), the ‘closed’ CTn confirmation (*attC*) and the integrated left and right attachment sites (*attL*, *attR*) (Table 3.12). Fold differences for each amplicon in each strain between the two media conditions were evaluated using the $\Delta\Delta C_q$ method (5) with *BT1087* as the reference gene. The LinRegPCR v2021.1 program (50) was used to calculate the starting concentration of each amplicon (N_0), which were in turn used to evaluate the frequency of excision ($N_0 \text{ attC} / N_0 \text{ attL}$ and $N_0 \text{ attC} / N_0 \text{ attR}$) using a two-way ANOVA.

3.6.4 Quantification and statistical analysis. Statistical tests, number of events quantified, standard error of the mean or deviations, and statistical significance is reported in figure legends.

3.7 Figures and Tables

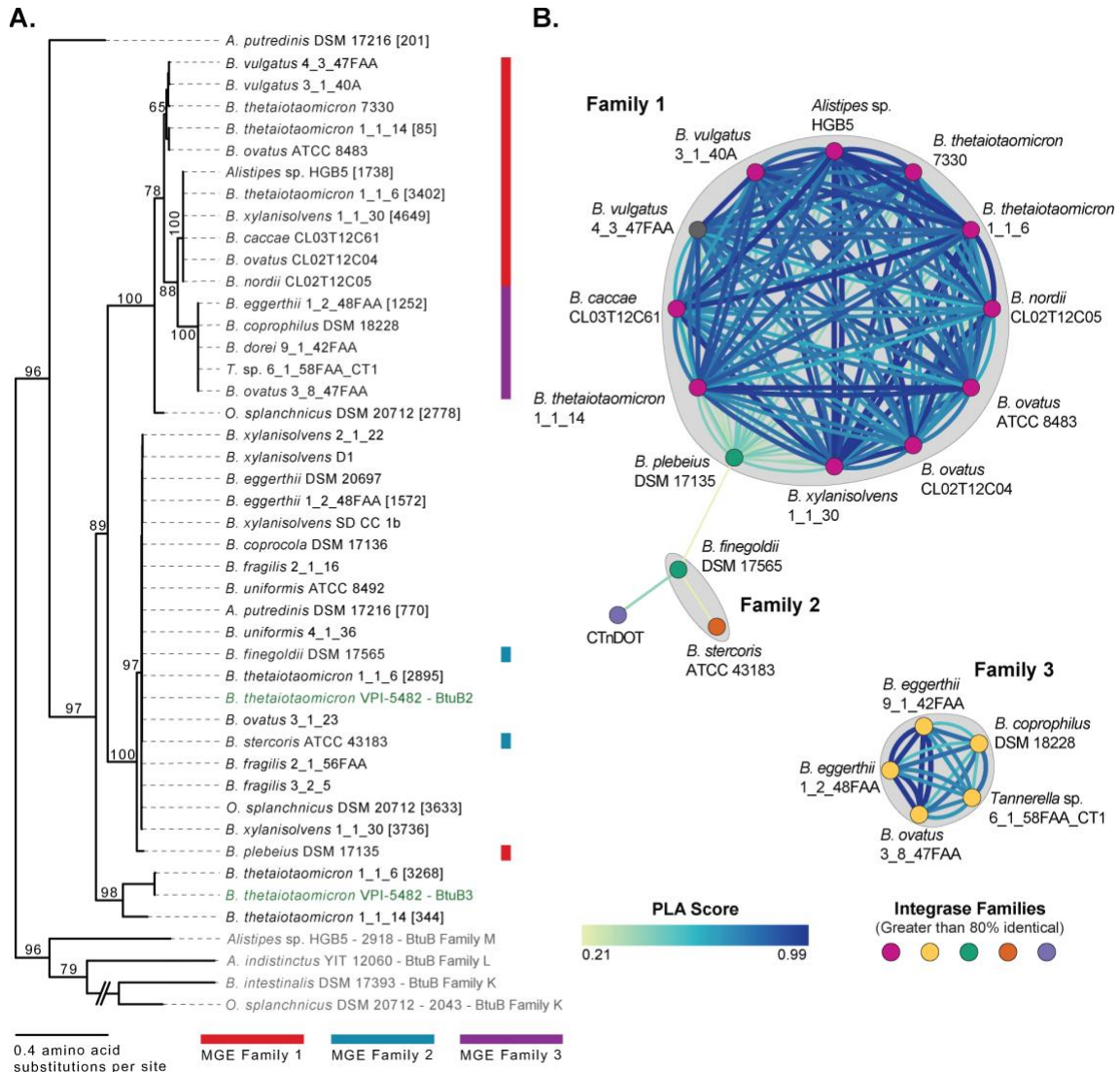


Figure 3.1. Predicted Corrinoid Transport-Encoding MGEs Encode Similar BtuB Alleles and Group into Three Distinct Families. **A.** Maximum likelihood phylogeny of chromosomal and mobilizable gut Bacteroidetes BtuB alleles. All MGE encoded BtuB alleles cluster within BtuB Family N. Representatives of Families K, L, and M are included as outgroups (grey). Chromosomally-encoded BtuB2 and BtuB3 from *B. thetaiotaomicron* VPI-5482 are in green and the MGE family groupings are noted by colored blocks. Species with multiple BtuB alleles represented contain gene identifiers in brackets. Bootstrap values are shown. **B.** Predicted MGE families were inferred by the MCL algorithm based on the percent length aligned (PLA) between each MGE. Edge widths and colors connecting each of the nodes represent the similarity between MGEs. CTnDOT was included for reference. Nodes are colored by the amino acid identity of the MGE integrase with a grey node indicating no integrase was detected. See also Figures 3.2, 3.4, 3.5 and 3.3.

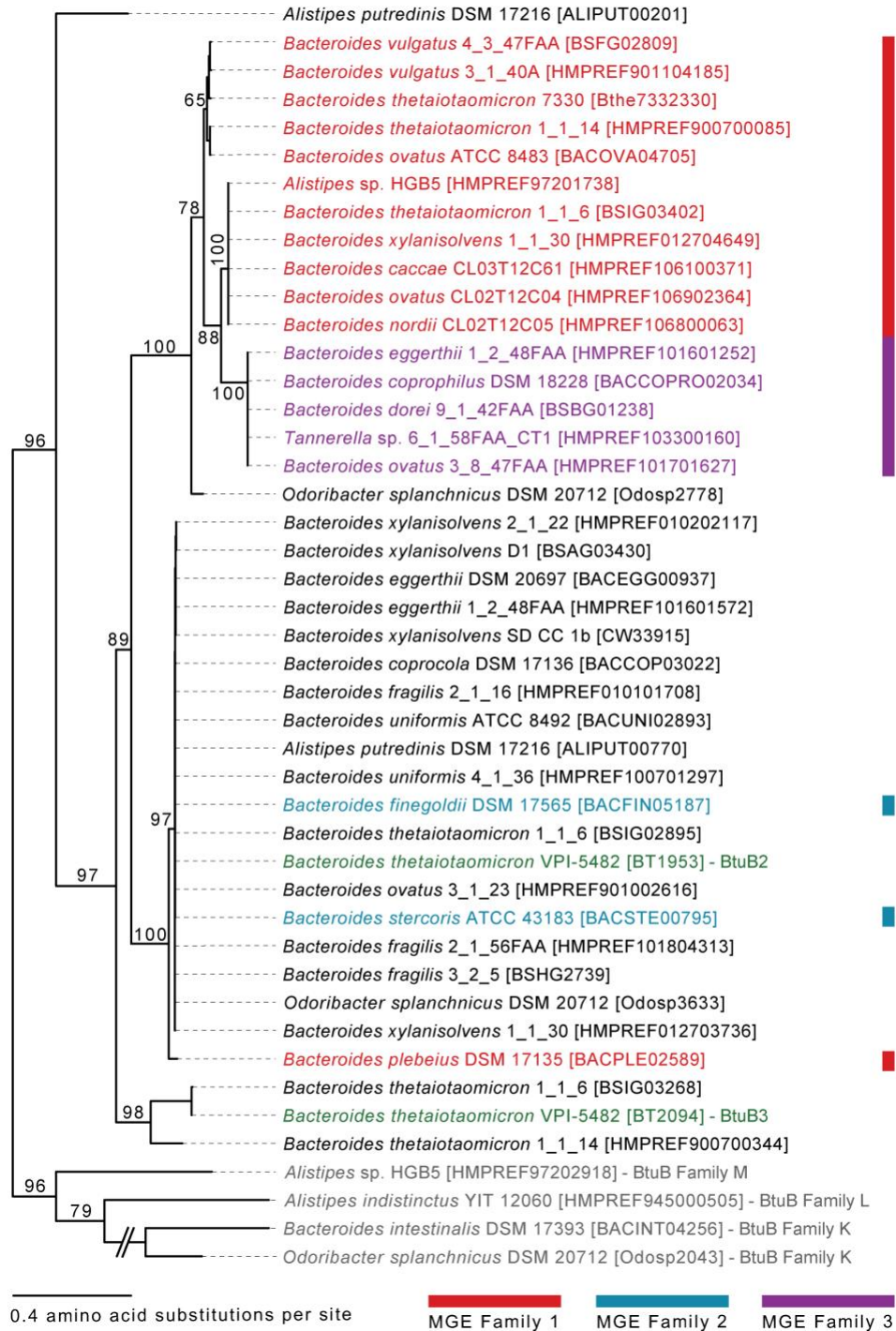


Figure 3.2. All BtuB Alleles Encoded on Predicted Corrinoid Transport-Encoding MGEs are Members of Family N, Related to Figure 3.1. Maximum likelihood phylogeny of BtuB alleles from gut Bacteroidetes from BtuB Family N, K, L, and M. Outgroups are shown in grey. Chromosomally-encoded BtuB2 and BtuB3 from *B. thetaiotaomicron* VPI-5482 are in green and the MGE family groupings are noted as colored blocks. Bootstrap values are shown. Full genus, species, and gene loci names included from Figure 3.1.

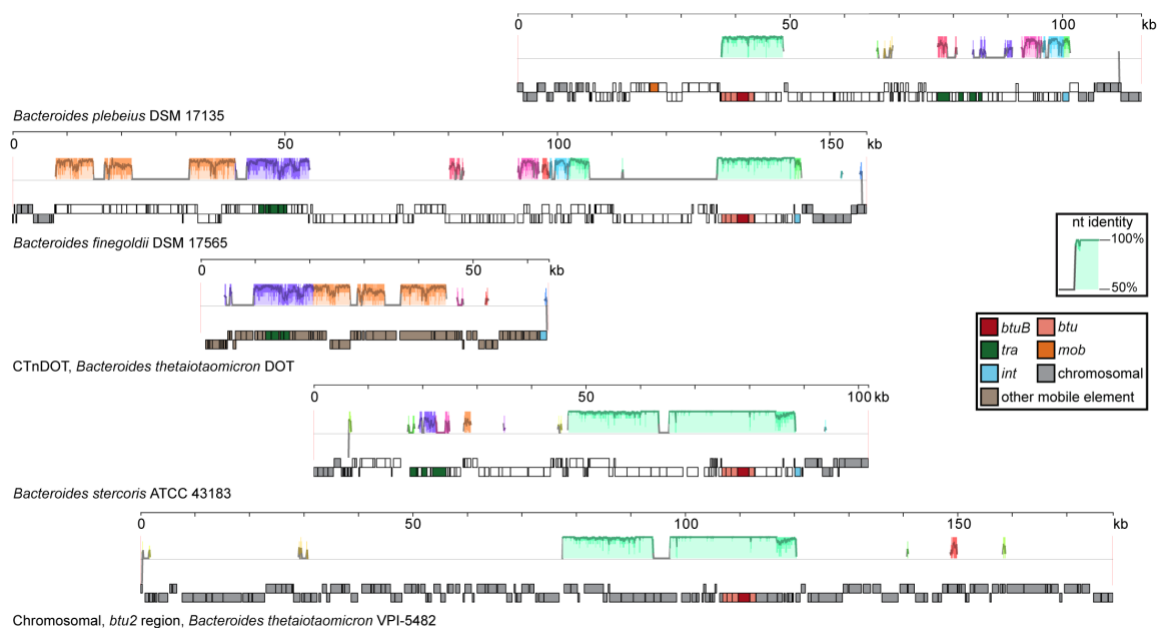


Figure 3.3. Family 2 MGEs Have Limited Regions of High Sequence Similarity, Related to Figure 3.1. Nucleotide alignments of predicted MGEs with ten additional genes upstream and downstream generated by Mauve (Darling *et al.*, 2004). Open reading frames are denoted below each alignment as boxes and are colored according to their identity as one of the following: the *btu* operon (red), predicted transfer genes (green), predicted mobilization gene (orange), predicted integrase (blue), CTnDOT (brown), chromosomal (grey). The height of the colored plots represent the nucleotide (nt) identity at a given location according to the key and the coloring represents different sequence homology groups.

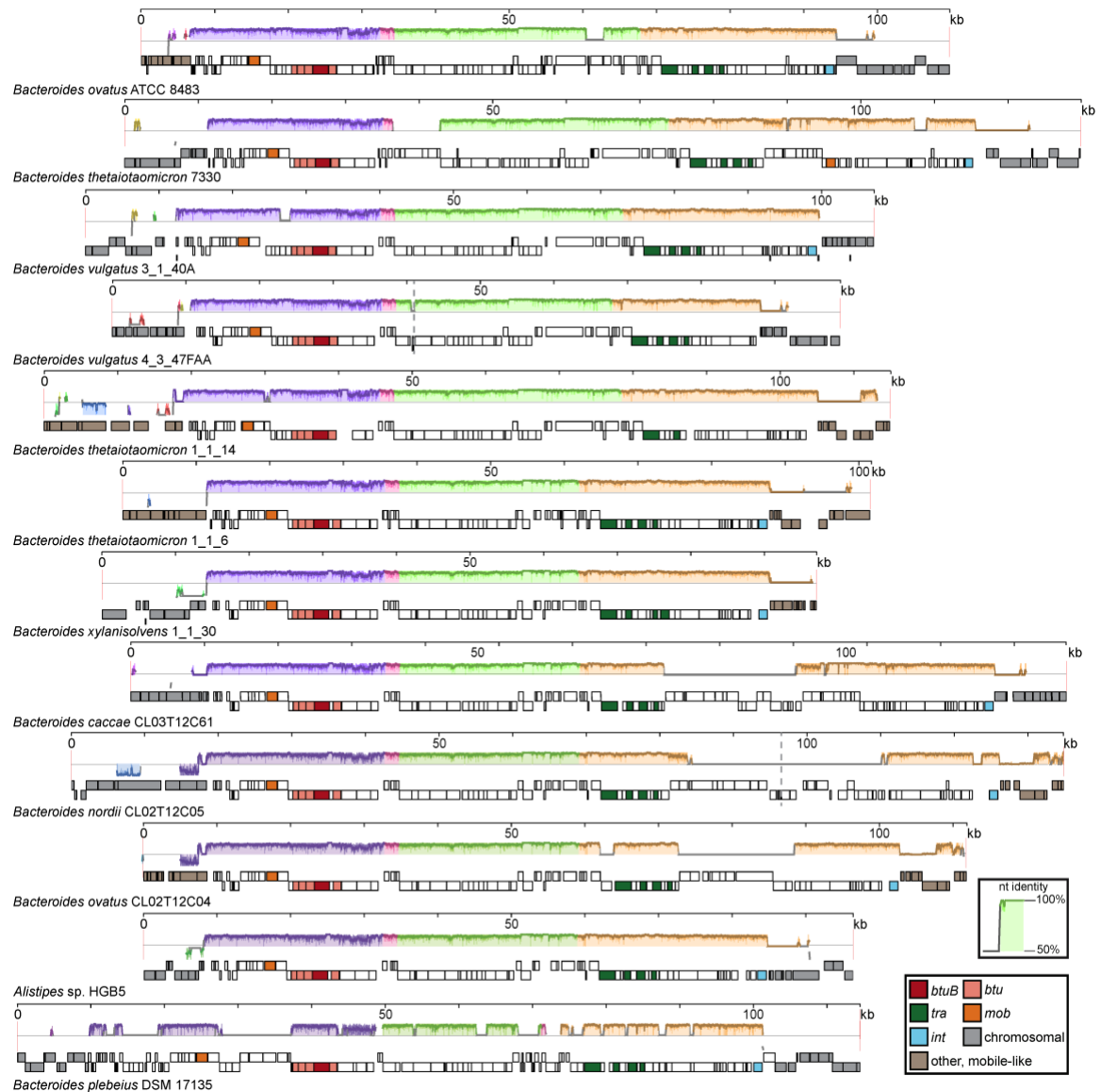


Figure 3.4. Family 1 Corrinoid Transport-Encoding MGEs Have Broad Conservation, Related to Figure 3.1. Nucleotide alignments of predicted MGEs with ten additional genes upstream and downstream generated by Mauve (Darling *et al.*, 2004). Open reading frames are denoted below each alignment as boxes and are colored according to their identity as one of the following: the *btu* operon (red), predicted transfer genes (green), predicted mobilization gene (orange), predicted integrase (blue), chromosomal (grey). The height of the colored plots represent the nucleotide (nt) identity at a given location according to the key. Colors represent the different homology groups. Note, the *B. nordii* and *B. vulgatus* 4_3_47FAA MGEs are split over two contigs and the break points are marked by a vertical grey dashed line.

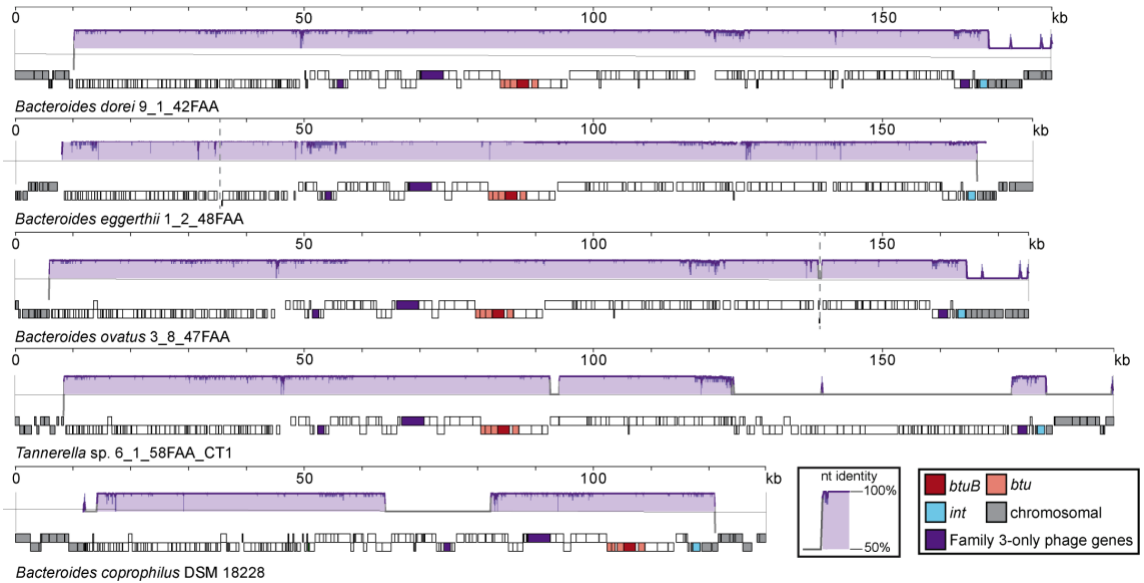


Figure 3.5. Family 3 Corrinoi Transport-Encoding Phage are Highly Related, Related to Figure 3.1. Nucleotide alignments of predicted MGEs with ten additional genes upstream and downstream generated by Mauve (Darling *et al.*, 2004). Open reading frames are denoted below each alignment as boxes and are colored according to their identity as one of the following: the *btu* operon (red), predicted phage genes conserved across Family 3 and not present in Families 1 and 2 (purple), predicted integrase (blue), chromosomal (grey). The height of the purple plots represent the nucleotide (nt) identity at a given location according to the key for single identified homology group. Predicted phage in *B. eggerthii* and *B. ovatus* are split over two contigs with the break points marked by a vertical grey dashed line.

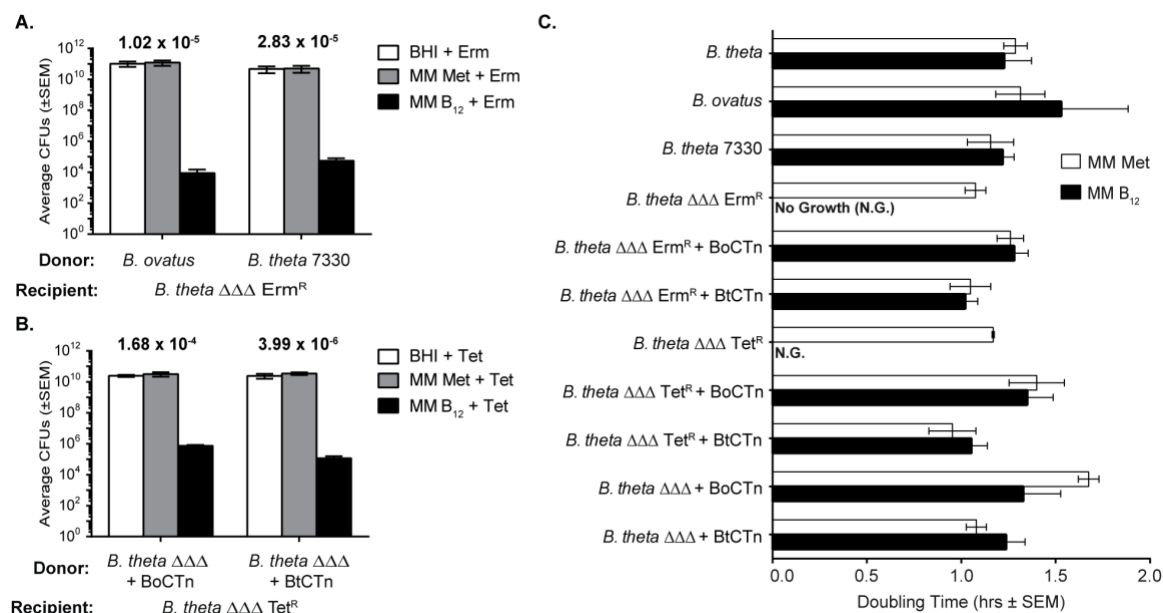


Figure 3.6. Predicted MGEs BoCTn and BtCTn are Mobilizable and Transport Vitamin B₁₂. **A.** Initial conjugations were performed between native MGE-host strains *B. ovatus* ATCC 8483 and *B. thetaiotaomicron* 7330 and recipient strain *B. thetaiotaomicron* $\Delta\Delta\Delta$ Erm^R . Successful *B. thetaiotaomicron* $\Delta\Delta\Delta$ transconjugants for both BoCTn and BtCTn were isolated in minimal media containing vitamin B₁₂ and erythromycin. **B.** Secondary conjugations between BoCTn or BtCTn encoding transconjugants and *B. thetaiotaomicron* $\Delta\Delta\Delta$ Tet^R recipients were performed to investigate if the first transfer of the CTn included all the necessary elements for repeated transfer of the element. Successful *B. thetaiotaomicron* $\Delta\Delta\Delta$ transconjugants for both BoCTn and BtCTn were isolated in minimal media containing vitamin B₁₂ and tetracycline. Transfer frequencies for each conjugation are denoted above each section of bars. **C.** Parent, recipient, and representative initial and secondary transconjugant strains were grown in MM+Met and MM+B₁₂. Recipient triple locus deletion strains (*B. theta* $\Delta\Delta\Delta$ Erm^R and Tet^R) are incapable of growing in the presence of vitamin B₁₂ while all parental strains and transconjugants are capable. Values represent the average and standard error of the mean for four biological replicates (A, B) or biological triplicates (C).

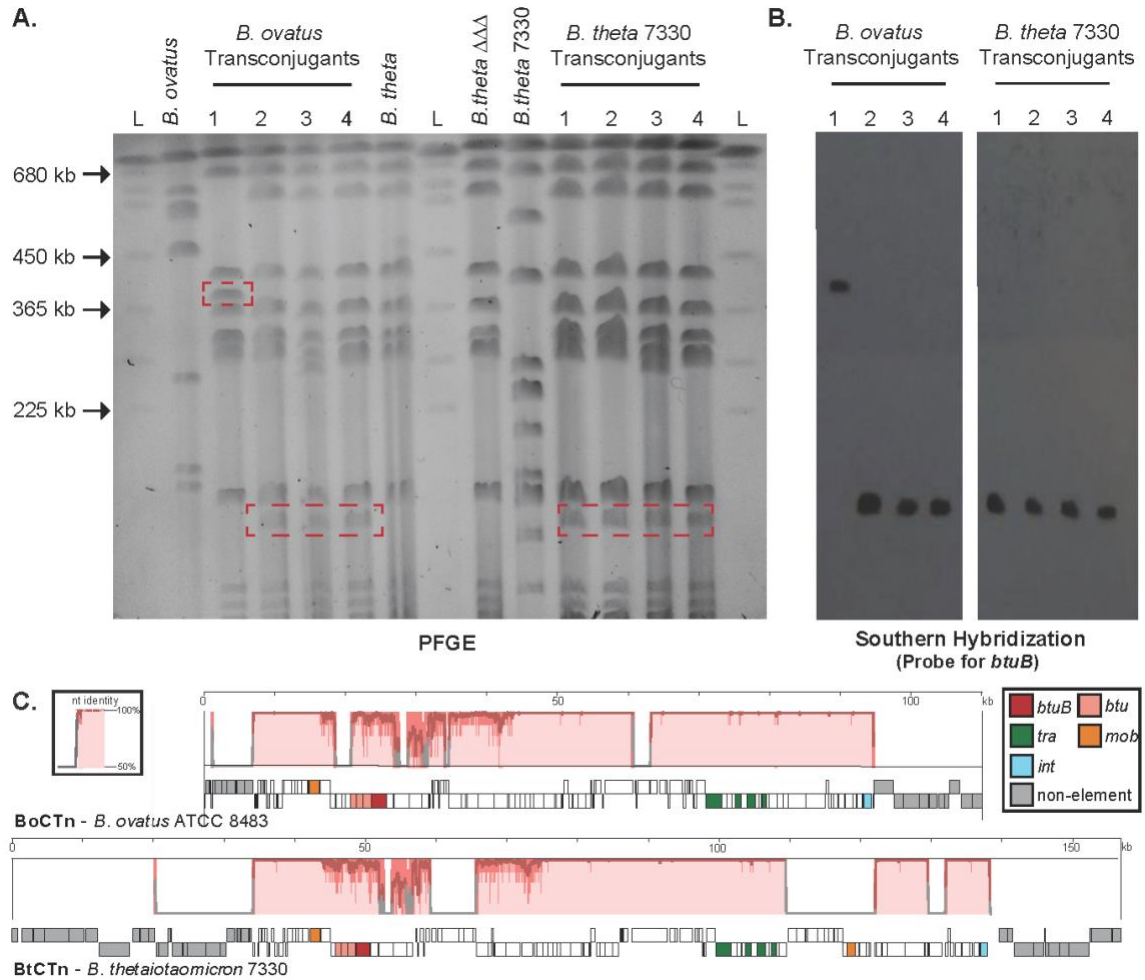


Figure 3.7. BoCTn and BtCTn Transfers are Detected by PFGE and Southern Hybridization. **A.** Pulsed field gel electrophoresis (PFGE) of wildtype *B. thetaiotaomicron*, donor (*B. ovatus* and *B. thetaiotaomicron* 7330), recipient (*B. thetaiotaomicron* $\Delta\Delta\Delta$), and representative transconjugants. Transconjugant genome fingerprints match the recipient with the exception of evidence of CTn insertion (red dotted boxes). L signifies the Ladder lanes. **B.** Corresponding Southern hybridizations of PFGE gel in A with probes specific to *btuB*. **C.** Nucleotide alignments of BoCTn and BtCTn (including their flanking regions) have a single locally conserved block as aligned by Mauve. Shared regions have $\geq 80\%$ nucleotide sequence identity as indicated by the similarity profile. Genes of note are colored according to the key and further details are available in Table 3.10, 3.11. See Figure 3.8 for the consensus attachment site sequences.

Sequenced att Sites	<i>B. ovatus attL</i> :	ACTAAATTTTCCAGCTTC
	<i>B. ovatus attR</i> :	AATAAAGTTTCCAGATTT
	<i>B. theta</i> 7330 <i>attL</i> :	ACTAAATTTACCAGATTT
	<i>B. theta</i> 7330 <i>attR</i> :	ACTAAATTTACCAGATTT
	<i>B. theta attB</i> 1:	ACTAAATTTACCAGATTT
	<i>B. theta attB</i> 2:	GTTTAATTTACCAGATTT
	<i>B. theta attB</i> 3:	CATAAAATTACCAGATTT
	<i>B. theta attB</i> 4:	TATAAATTTGCCAGTTTA
	Transconjugant <i>attL</i> (12):	ACTAAATTTACCAGATTT
	Transconjugant <i>attR</i> (12):	ACTAAATTTACCAGATTT
	Transconjugant <i>attL</i> (2):	ACTAAATTTACCAGATTT
	Transconjugant <i>attR</i> (2):	ACTAAATTTTCCAGATTT
	Transconjugant <i>attL</i> (1):	CATAAAATTTCAGATTT
	Transconjugant <i>attR</i> (1):	ACTAAATTTACCAGATTT
	Transconjugant <i>attL</i> (1):	ACTAAATTTGCCAGTTTA
	Transconjugant <i>attR</i> (1):	TATAAATTTTCCAGATTT
	<i>Alistipes</i> sp. HGB5 <i>attL</i> :	AAAAAATTTACCAGATTT
	<i>Alistipes</i> sp. HGB5 <i>attR</i> :	ACTAAATTTACCAGATTC
Predicted att Sites	<i>B. caccae</i> CL03T12C61 <i>attL</i> :	GTATAATTTATCAGATTT
	<i>B. caccae</i> CL03T12C61 <i>attR</i> :	ACTAAATTTAGCAGATTT
	<i>B. nordii</i> CL02T12C05 <i>attL</i> :	TTCCAATTTACCAGATTT
	<i>B. nordii</i> CL02T12C05 <i>attR</i> :	ACTAAATTTACCAGATTT
	<i>B. ovatus</i> CL02T12C04 <i>attL</i> :	TGTAAATTTACCAGATTT
	<i>B. ovatus</i> CL02T12C04 <i>attR</i> :	ACTAAATTTACCAGATTT
	<i>B. thetaiotaomicron</i> 1_1_14 <i>attL</i> :	TTCCAATTTACCAGATTT
	<i>B. thetaiotaomicron</i> 1_1_14 <i>attR</i> :	ACTAAATTTACCAGATTT
	<i>B. thetaiotaomicron</i> 1_1_6 <i>attL</i> :	TATCAATTTACCAGATTT
	<i>B. thetaiotaomicron</i> 1_1_6 <i>attR</i> :	ACTAAATTTACCAGATTT
	<i>B. vulgatus</i> 4_3_47FAA <i>attR</i> :	ACTAAATTTACCAGATTT
	<i>B. vulgatus</i> 3_1_40A <i>attR</i> :	ACTAAATTTACCAGATTT
	<i>B. xylanisolvens</i> 1_1_30 <i>attL</i> :	AATAAATTTACCAGATTT
	<i>B. xylanisolvens</i> 1_1_30 <i>attR</i> :	ACTAAATTTACCAGATTT
	Consensus:	ACTAAATTTACCAGATTT

Figure 3.8. Attachment Site Sequences for the Predicted Corrinoide Transport-Encoding CTns are Conserved, Related to Figure 3.7. Attachment sites for the BoCTn and BtCTn were determined based on genome sequencing of transconjugants. Attachment sites for the remaining predicted CTns as well as the *attB* site in the recipient strain (*B. thetaiotaomicron*) were subsequently identified. The consensus *att* site is shown in bold at the bottom.

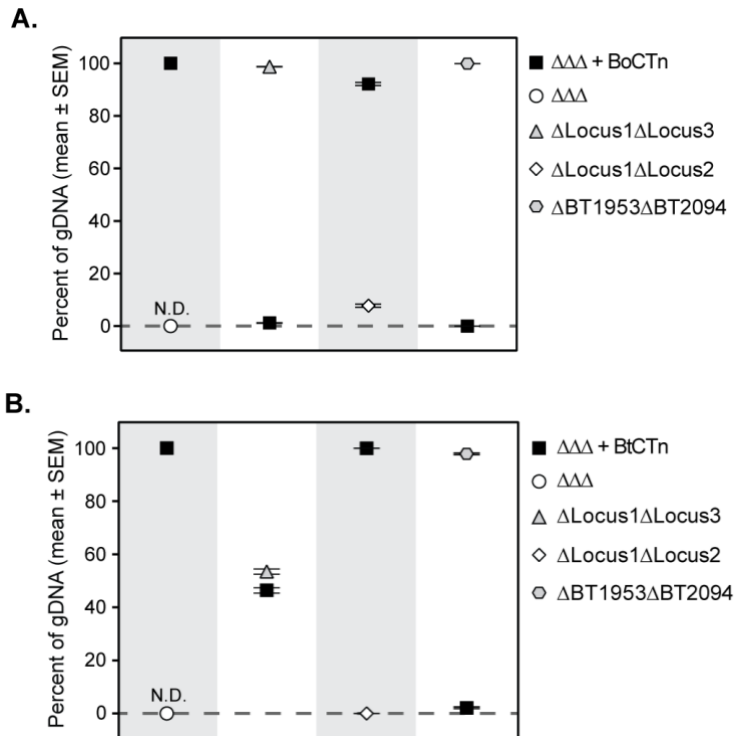


Figure 3.9. Transconjugants Have Intermediate Competitive Advantages Compared to *B. thetaiotaomicron* VPI-5482 Chromosomal *btu* Loci.

Competitions were performed between *B. thetaiotaomicron* $\Delta\Delta\Delta$ + BoCTn (A) or BtCTn (B) and *B. thetaiotaomicron* $\Delta\Delta\Delta$, Δ Locus1 Δ Locus3 (Locus 2), Δ Locus1 Δ Locus2 (Locus 3), or Δ BT1953 Δ BT2094 (*btuB1*, *btuFCD2/3*). Competitions were performed over five days in MM+B₁₂ with serial passaging every 24 hours. Strain abundances were determined by qPCR and are represented by percent of gDNA. Error bars represent the standard error of the mean from three technical qPCR replicates of three technical competition replicates. See also Figure 3.10.

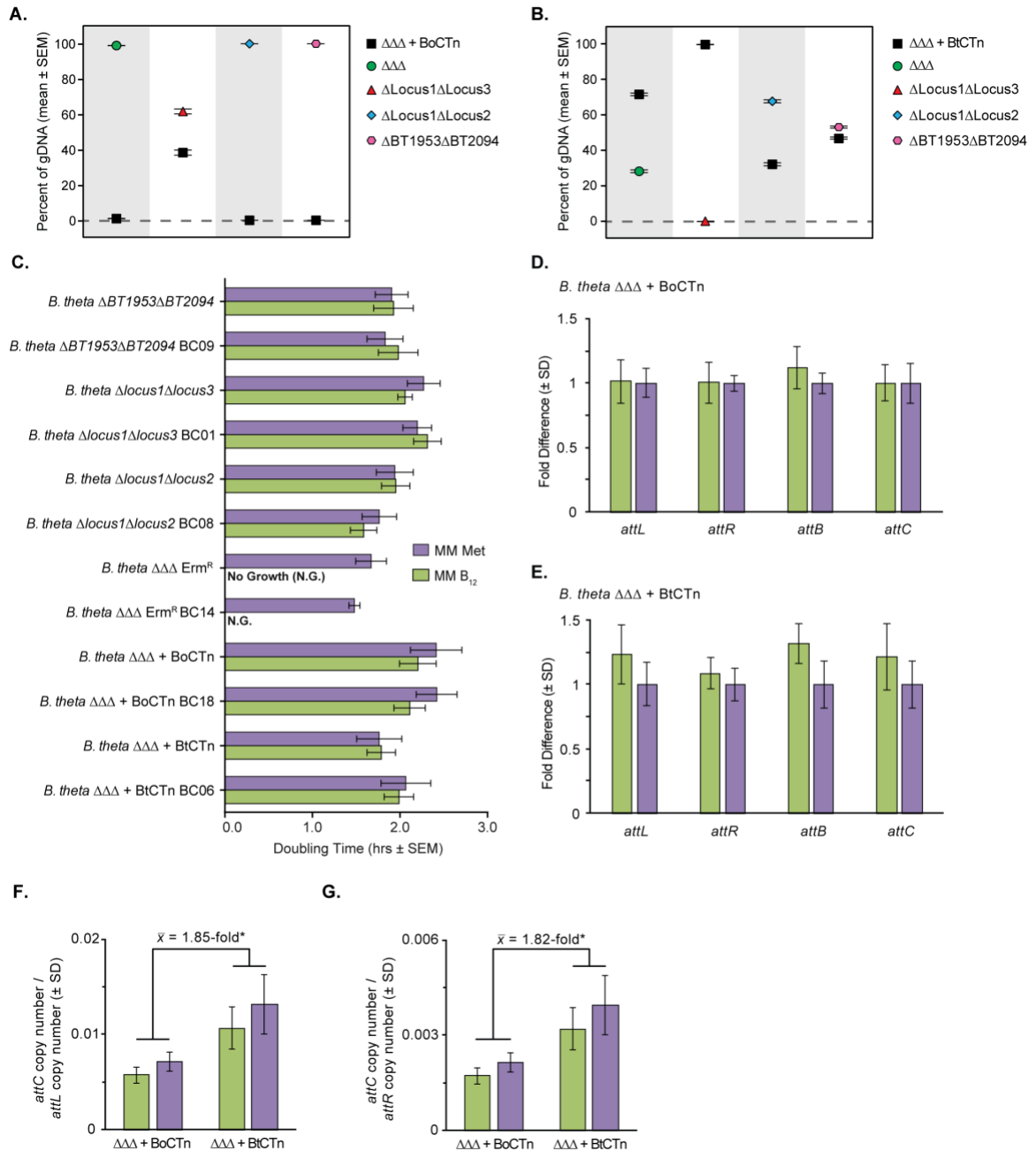


Figure 3.10. Control Competitions of Transconjugants and Single Locus Mutants, Related to Figure 3.9. Competitions were performed between *B. theta* Δ tao Δ micron $\Delta\Delta\Delta$ + BoCTn (A) or BtCTn (B) and *B. theta* Δ tao Δ micron $\Delta\Delta\Delta$, Δ Locus1 Δ Locus3 (Locus 2), Δ Locus1 Δ Locus2 (Locus3), or Δ BT1953 Δ BT2094 (*btuB1*, *btuFCD2/3*). Competitions were performed over five days in MM+Met with serial passaging every 24 hours. Strain abundances were determined by qPCR and are represented by percent of gDNA. Error bars represent the standard error of the mean from three technical qPCR replicates of three technical competition replicates.

Figure 3.10 Continued. (C) Plasmid-free and barcode plasmid containing (BC) strains were grown in MM+Met and MM+B12 at both 4nM and 0.4nM with growth in 0.4nM B12 shown. All barcoded strains showed growth in either MM+Met or MM+B12 near identical to their non-barcoded counterpart. Apart from each strain being significantly different from growth of the *B. theta* $\Delta\Delta\Delta$ strains in MM+B12, the only significant differences observed were between that of *B. theta* $\Delta\Delta\Delta$ ErmR BC14 and *B. theta* $\Delta\Delta\Delta$ + BoCTn BC18 in MM+Met. Values represent the average and standard error of six technical replicates from biological duplicates. Statistical analysis was performed with a one-way ANOVA. Fold differences in DNA copy number for (D) *B. theta* $\Delta\Delta\Delta$ + BoCTn and (E) *B. theta* $\Delta\Delta\Delta$ + BtCTn attachment sites were calculated using the $\Delta\Delta$ Cq method for cells grown to mid-log phase in MM+B12 or MM+Met. Error bars represent the standard deviations from three biological replicates. Excision frequencies were determined by calculating the starting concentration of each amplicon (*NO* ~ copy number) and taking the ratio of the CTn attachment site (*attC*) and dividing it by (F) the left (*NO attC* / *NO attL*) and (G) the right (*NO attC* / *NO attR*) attachment sites. Means and standard deviations are shown and were used in two-way ANOVA which identified significant differences (*) between the strains (F) $F(1,8)=22.03$ $p = 0.0016$ and (G) $F(1,8)=26.83$ $p = 0.0008$.

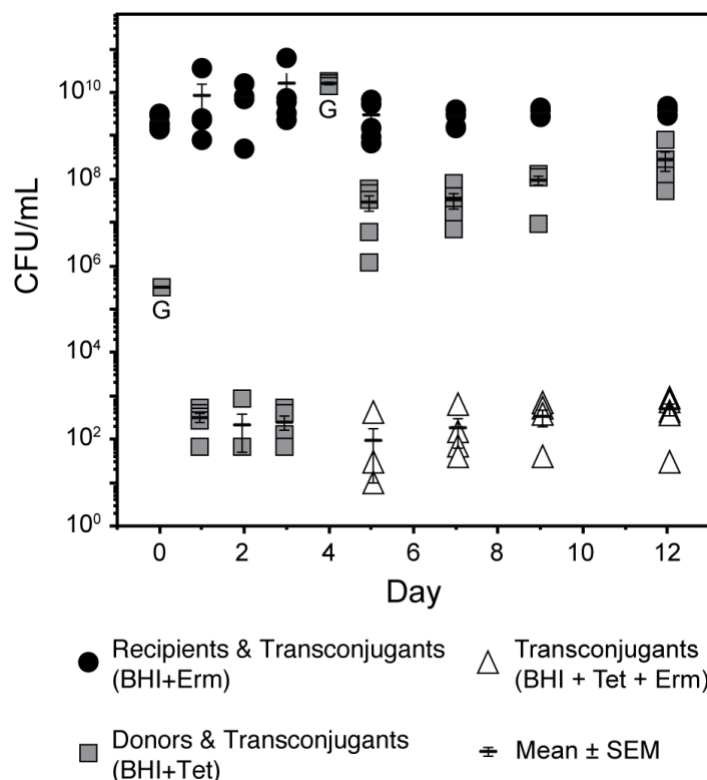


Figure 3.11. BoCTn and BtCTn can Mobilize Within the Mammalian Intestine. Germ-free mice were initially colonized with recipient *B.*

thetaitaomicron $\Delta\Delta\Delta$ Erm^R two days prior to introduction of donor cells (Day -2).

On Day 0, equal portions (G, gavage sample) of donor strains *B.*

thetaitaomicron $\Delta\Delta\Delta$ + BoCTn_Tet^R and *B. thetaitaomicron* $\Delta\Delta\Delta$ + BtCTn_Tet^R were introduced to the mice colonized by *B. thetaitaomicron* $\Delta\Delta\Delta$ Erm^R recipient cells.

At days 0, 1, 3, 5, 7, 9 and 12, fecal samples were collected and plated on BHI+Tet+Gn, BHI+Erm+Gn, and BHI+Erm+Tet+Gn for CFUs. On Day 4, a

second gavage (G) of the donor strains was administered. Starting on Day 5,

transconjugants were observed (triangles). Mean calculated transfer frequencies

for each day were: Day 5 $\sim 9.7 \times 10^{-9}$, Day 7 $\sim 3.3 \times 10^{-8}$, Day 9 $\sim 1.0 \times 10^{-7}$ and

Day 12 $\sim 8.2 \times 10^{-8}$. Mean CFU/mL based on 2 or 3 replicates; standard error of

the mean based on five mice. See also Figure 3.12.

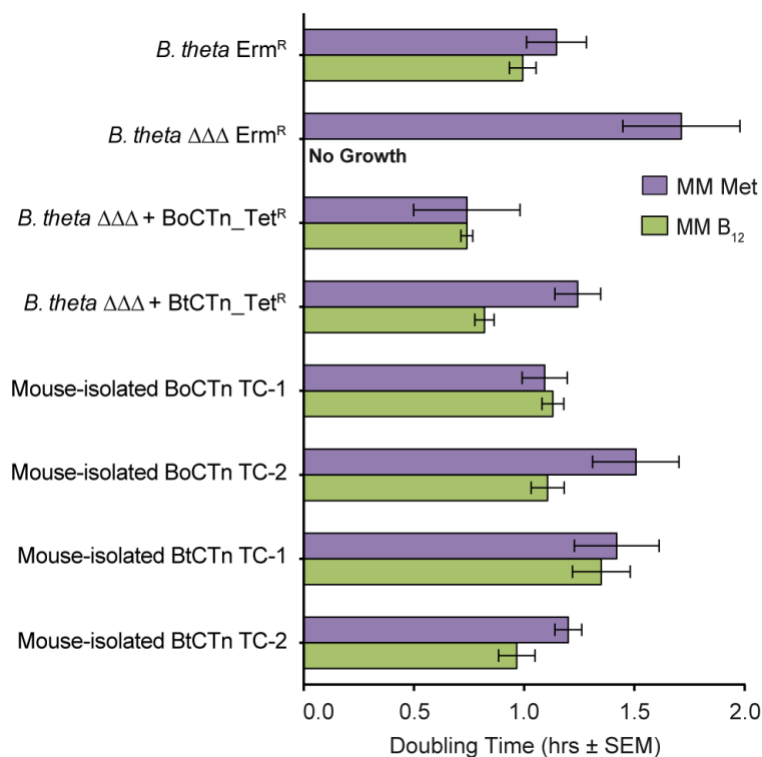


Figure 3.12. Gnotobiotic Transconjugants Have Restored Growth in Minimal Media with Vitamin B12, Related to Figure 3.11. Parent, recipient, and representative transconjugant strains were grown in MM+Met and MM+B12. All of the transconjugants (Mouse-isolated BoCTn TC 1 & 2 and Mouse-isolated BtCTn TC 1 & 2) are capable of growing in the presence of vitamin B12. Apart from each strain being significantly different from growth of *B. theta* ΔΔΔ Erm^R in MM B12, no other significant differences were observed. Values represent the average and standard error of biological triplicates. Statistical analysis was performed with a one-way ANOVA.

Table 3.1. Predicted Corrinoid Transport-Encoding MGEs.

Organism	BtuB	BtuB Phylogeny Group	Total Length (bp)
<u>MGE Family 1 (ACTAAATTTACCAGATTT/ACTAAATTTACCAGATTT)*</u>			
<i>Bacteroides ovatus</i> ATCC 8483	BACOVA04705	1	87,773
<i>Bacteroides thetaiotaomicron</i> 7330	Bthe7332330	1	104,197
<i>Bacteroides thetaiotaomicron</i> 1_1_14	HMPREF900700085	1	86,150
<i>Bacteroides vulgatus</i> 3_1_40A	HMPREF901104185	1	86,906
<i>Bacteroides vulgatus</i> 4_3_47FAA^	BSFG02809	1	77,507
<i>Alistipes</i> sp. HGB5	HMPREF97201738	2	76,410
<i>Bacteroides caccae</i> CL03T12C61	HMPREF106100371	2	106,936
<i>Bacteroides nordii</i> CL02T12C05^	HMPREF106800063	2	107,262
<i>Bacteroides ovatus</i> CL02T12C04	HMPREF106902364	2	94,150
<i>Bacteroides thetaiotaomicron</i> 1_1_6	BSIG03402	2	76,411
<i>Bacteroides xylanisolvens</i> 1_1_30	HMPREF012704649	2	76,412
<i>Bacteroides plebeius</i> DSM 17135 [#]	BACPLE02589	3	93,364
<u>MGE Family 2 (GCGG/GCGG)*</u>			
<i>Bacteroides fingoldii</i> DSM 17565	BACFIN05187	3	137,152
<i>Bacteroides stercoris</i> ATCC 43183	BACSTE00795	3	82,529
<u>MGE Family 3 (GGCARAC/CTR TGCC)*</u>			
<i>Bacteroides coprophilus</i> DSM 18228	BACCOPRO02034	4	106,761
<i>Bacteroides dorei</i> 9_1_42FAA	BSBG01238	4	158,110
<i>Bacteroides eggerthii</i> 1_2_48FAA^	HMPREF101601252	4	158,130
<i>Bacteroides ovatus</i> 3_8_47FAA^	HMPREF101701627	4	158,569
<i>Tannerella</i> sp. 6_1_58FAA_CT1	HMPREF103300160	4	169,829

*Consensus predicted *attL/attR* sites, [#]Possesses non-consensus *attL/attR*

^MGE split on two different contigs

Table 3.2. Complete information on Predicted Corrinoid Transport-Encoding MGEs.

Organism	BtUB	Predicted Integrase ¹	BtUB Family	Predicted MGE Family	BtUB Phylogeny Group	ORF Coordinates	
						Start	Stop
<i>Bacteroides ovatus</i> ATCC 8483	BACOV040705	BACOV040789	BtUB family N - Addition	1	1	BACOV040675	BACOV040790
<i>Bacteroides thetaiotaomicron</i> 7330	Bthe7332330	Bthe7332235	BtUB family N - Addition	1	1	Bthe7332235	Bthe7332355
<i>Bacteroides thetaiotaomicron</i> 1_1_14	HMPREF900700085	HMPREF900700152	BtUB family N - Addition	1	1	HMPREF900700059	HMPREF900700152
<i>Bacteroides vulgatus</i> 3_1_40A	HMPREF901104185	HMPREF901104253	BtUB family N	1	1	HMPREF901104160	HMPREF901104253
<i>Bacteroides vulgatus</i> 4_3_47FAA	BSFG02809	n.d.	BtUB family N	1	1	BSFG04789	BSFG04790
						BSFG02821	BSFG02784
<i>Alistipes</i> sp. HGB5	HMPREF97201738	HMPREF97201676	BtUB family N	1	2	HMPREF97201676	HMPREF97201757
<i>Bacteroides caccae</i> CL03T12C61	HMPREF106100371	HMPREF106100273	BtUB family N - Addition	1	2	HMPREF106100273	HMPREF106100390
<i>Bacteroides nordii</i> CL02T12C05	HMPREF106800063	HMPREF106804326	BtUB family N - Addition	1	2	HMPREF106804326	HMPREF106804354
						HMPREF106800001	HMPREF106800082
<i>Bacteroides ovatus</i> CL02T12C04	HMPREF106902364	HMPREF106902289	BtUB family N - Addition	1	2	HMPREF106902289	HMPREF106902384
<i>Bacteroides thetaiotaomicron</i> 1_1_6	BSIG03402	BSIG03475	BtUB family N	1	2	BSIG03378	BSIG03475
<i>Bacteroides xylosoxylovens</i> 1_1_30	HMPREF012704649	HMPREF012704710	BtUB family N	1	2	HMPREF012704630	HMPREF012704710
<i>Bacteroides plebeius</i> DSM 17135	BACPLE02589	BACPLE02519	BtUB family N	1	3	BACPLE02518	BACPLE02632
<i>Bacteroides finegoldii</i> DSM 17565	BACFIN05187	BACFIN05178	BtUB family N - Addition	2	3	BACFIN05178	BACFIN05327
<i>Bacteroides stercoris</i> ATCC 43183	BACSTE00795	BACSTE00804	BtUB family N	2	3	BACSTE00723	BACSTE00804
<i>Bacteroides coprophilus</i> DSM 18228	BACOPRO02034	BACOPRO02023	BtUB family N	3	4	BACOPRO02023	BACOPRO02171
<i>Bacteroides dorei</i> 9_1_42FAA	BSBG01238	BSBG01316	BtUB family N	3	4	BSBG01134	BSBG01316
<i>Bacteroides eggertii</i> 1_2_48FAA	HMPREF101601252	HMPREF101601179	BtUB family N	3	4	HMPREF101601179	HMPREF101601308
						HMPREF101601309	HMPREF101601353
<i>Bacteroides ovatus</i> 3_8_47FAA	HMPREF101701627		BtUB family N	3	4	HMPREF101701529	HMPREF101701675
		HMPREF101701702				HMPREF101701676	HMPREF101701702
<i>Tamarella</i> sp. 6_1_58FAA CT1	HMPREF103300160	HMPREF103300253	BtUB family N	3	4	HMPREF103300065	HMPREF103300253

¹ n.d. = not detected; ² MC = megacolon; UC = ulcerative colitis; IBS = irritable bowel syndrome; CD = Crohn's disease; ³ Dates indicated as "s" only have date of genome sequence deposit available; ⁴ n.a. = not available

Table 3.2. Continued.

Nucleotide Coordinates				Host	Type ²	Year ³	Location ⁴
MGE Predicted Start	MGE Predicted Stop	Predicted Att Sites ¹					
5,369,288	5,457,061	AAATCTGGAACCTTTATT/GAAGCTGGAAAATTTAGT		human	feces	1933	New York, USA
2,730,344	2,834,541	AAATCTGGTAATTTAGT/AAATCTGGTAATTTAGT		human	feces	2009	Woods Hole, MA, USA
75,403	161,553	AAATCTGGTAATTTAGT/AAATCTGGTAATTTGGAA		human	MC colon biopsy, 40yo male	≤2010	Calgary, Alberta, CA
5,034,353	5,121,259	AAATCTGGTAATTTAGT/ n.d. (ends on contig)		human	UC intestinal biopsy, 38yo female	≤2010	Calgary, Alberta, CA
3,589,168	3,636,593	n.d.		human	CD colon biopsy, 25yo female	≤2011	Calgary, Alberta, CA
3,282,436	3,312,518	<--/ACTAATTTACCAGATTT					
3,006,985	3,083,395	AAATCTGGTAATTTT/ GAATCTGGTAATTTAGT		human	feces	≤2010	n.a.
361,333	468,269	AAATCTGATAAATTATAC/ AAATCTGCTAAATTTAGT		human	feces, adult	2014	Massachusetts, USA
5,589,661	5,619,011	AAATCTGGTAATTTGAA/ -->		human	feces, adult	2014	Massachusetts, USA
1	77,913	<--/AAATCTGGTAATTTAGT					
2,638,955	2,733,105	AAATCTGGTAATTTTACA/ AAATCTGGTAATTTAGT		human	feces, adult	2014	Massachusetts, USA
4,270,738	4,347,149	AAATCTGGTAATTTAGT/ AAATCTGGTAATTTGATA		human	healthy colon biopsy, 56yo male	≤2009	Calgary, Alberta, CA
5,744,576	5,820,988	GTAATCTGTAATTTTA/ GTAATCTGTAATTTTA		human	healthy intestinal biopsy	2007	Calgary, Alberta, CA
2,825,098	2,918,462	CGG/ CCG		human	feces, adult	2005	Japan
555,497	692,649	n.d.		human	feces, healthy 23yo male	2006	Japan
694,681	777,210	GCGG/ GCGG		human	feces	1978	Texas, USA
1,954,187	2,060,948	GGCARAC/CTRTGCC		human	feces, healthy 52yo male	2007	Japan
1,430,335	1,588,445	GGCARAC/CTRTGCC		human	MC duodenum biopsy, 47yo female	2007	Calgary, Alberta, CA
1,528,206	1,658,600	GGCARAC/ -->		human	IBS intestinal biopsy, 25yo male	2007	Calgary, Alberta, CA
1,658,601	1,686,337	<--/CTRTGCC					
1,898,044	2,031,120	GGCARAC/ -->		human	CD intestinal biopsy, 25yo female	≤2011	Calgary, Alberta, CA
2,031,121	2,056,614	<--/CTRTGCC					
60,730	230,559	GGCARAC/CTRTGCC		human	UC intestinal biopsy, 39yo female	≤2011	Calgary, Alberta, CA

¹ n.d. = not detected; ² MC = megacolon, UC = ulcerative colitis, IBS = irritable bowel syndrome, CD = Crohn's disease; ³ Dates indicated as "s" only have date of genome sequence deposit available; ⁴ n.a. = not available

Table 3.3. Genomes Screened for the Presence of MGEs.

Organism	RefSeq Accession	Predicted <i>btu</i> MGE?
<i>Alistipes finegoldii</i> DSM 17242	NC_018011	
<i>Alistipes indistinctus</i> YIT 12060	NZ_ADLD000000000	
<i>Alistipes putredinis</i> DSM 17216	NZ_ABFK000000000	
<i>Alistipes shahii</i> WAL 8301	NC_021030	
<i>Alistipes</i> sp. HGB5	NZ_AENZ000000000	Yes
<i>Alistipes</i> sp. JC136	NZ_CAEG000000000	
<i>Bacteroides caccae</i> ATCC 43185	NZ_AAVM000000000	
<i>Bacteroides caccae</i> CL03T12C61	NZ_AGXF000000000	Yes
<i>Bacteroides cellulosilyticus</i> CL02T12C19	NZ_AGXG000000000	
<i>Bacteroides cellulosilyticus</i> DSM 14838	NZ_ACCH000000000	
<i>Bacteroides cellulosilyticus</i> WH2	NZ_ATFI000000000	
<i>Bacteroides clarus</i> YIT 12056	NZ_AFBM000000000	
<i>Bacteroides coprocola</i> DSM 17136	NZ_ABIY000000000	
<i>Bacteroides coprophilus</i> DSM 18228	NZ_ACBW000000000	Yes
<i>Bacteroides dorei</i> 5_1_36/D4	NZ_ACDI000000000	
<i>Bacteroides dorei</i> 9_1_42FAA	NZ_ACAA000000000	Yes
<i>Bacteroides dorei</i> CL02T00C15	NZ_AGXH000000000	
<i>Bacteroides dorei</i> CL02T12C06	NZ_AGXJ000000000	
<i>Bacteroides dorei</i> CL03T12C01	NZ_AGXI000000000	
<i>Bacteroides dorei</i> DSM 17855	NZ_ABWZ000000000	
<i>Bacteroides eggerthii</i> 1_2_48FAA	NZ_ACWG000000000	Yes
<i>Bacteroides eggerthii</i> DSM 20697	NZ_ABVO000000000	
<i>Bacteroides faecis</i> MAJ27	NZ_AGDG000000000	
<i>Bacteroides finegoldii</i> CL09T03C10	NZ_AGXW000000000	
<i>Bacteroides finegoldii</i> DSM 17565	NZ_ABXI000000000	Yes
<i>Bacteroides fluxus</i> YIT 12057	NZ_AFBN000000000	
<i>Bacteroides fragilis</i> 2_1_16	NZ_ACPP000000000	
<i>Bacteroides fragilis</i> 2_1_56FAA	NZ_ACWI000000000	
<i>Bacteroides fragilis</i> 3_1_12	NZ_ABZX000000000	
<i>Bacteroides fragilis</i> 3_2_5	NZ_ACIB000000000	
<i>Bacteroides fragilis</i> 638R	NC_016776	
<i>Bacteroides fragilis</i> CL03T00C08	NZ_AGXK000000000	
<i>Bacteroides fragilis</i> CL03T12C07	NZ_AGXL000000000	
<i>Bacteroides fragilis</i> CL05T00C42	NZ_AGXO000000000	
<i>Bacteroides fragilis</i> CL05T12C13	NZ_AGXP000000000	
<i>Bacteroides fragilis</i> CL07T00C01	NZ_AGXM000000000	
<i>Bacteroides fragilis</i> CL07T12C05	NZ_AGXN000000000	
<i>Bacteroides fragilis</i> HMW 610	NZ_AGXQ000000000	

Table 3.3. Continued.

Organism	RefSeq Accession	Predicted <i>btu</i> MGE?
<i>Bacteroides fragilis</i> HMW 615	NZ_AGXR000000000	
<i>Bacteroides fragilis</i> HMW 616	NZ_ALOB000000000	
<i>Bacteroides fragilis</i> NCTC 9343	NC_003228	
<i>Bacteroides fragilis</i> YCH46	NC_006347	
<i>Bacteroides helcogenes</i> P 36-108	NC_014933	
<i>Bacteroides intestinalis</i> DSM 17393	NZ_ABJL000000000	
<i>Bacteroides nordii</i> CL02T12C05	NZ_AGXS000000000	Yes
<i>Bacteroides oleiciplenus</i> YIT 12058	NZ_ADLF000000000	
<i>Bacteroides ovatus</i> 3_1_23	NZ_ACRS000000000	
<i>Bacteroides ovatus</i> 3_8_47FAA	NZ_ACWH000000000	Yes
<i>Bacteroides ovatus</i> ATCC 8483	NZ_AAXF000000000	Yes
<i>Bacteroides ovatus</i> CL02T12C04	NZ_AGXT000000000	Yes
<i>Bacteroides ovatus</i> CL03T12C18	NZ_AGXU000000000	
<i>Bacteroides ovatus</i> SD CC 2a	NZ_ADMP000000000	
<i>Bacteroides ovatus</i> SD CMC 3f	NZ_ADMO000000000	
<i>Bacteroides plebeius</i> DSM 17135	NZ_ABQC000000000	Yes
<i>Bacteroides salanitronis</i> DSM 18170	NC_015164	
<i>Bacteroides salyersiae</i> CL02T12C01	NZ_AGXV000000000	
<i>Bacteroides</i> sp. 2_1_33B	NZ_ACPR000000000	
<i>Bacteroides</i> sp. 2_1_7	NZ_ABZY000000000	
<i>Bacteroides</i> sp. 2_2_4	NZ_ABZZ000000000	
<i>Bacteroides</i> sp. 20_3	NZ_ACRQ000000000	
<i>Bacteroides</i> sp. 3_1_19	NZ_ADCJ000000000	
<i>Bacteroides</i> sp. 3_1_33FAA	NZ_ACPS000000000	
<i>Bacteroides</i> sp. CF50	NC_022526	
<i>Bacteroides</i> sp. D2	NZ_ACGA000000000	
<i>Bacteroides</i> sp. D20	NZ_ACPT000000000	
<i>Bacteroides</i> sp. D22	NZ_ADCK000000000	
<i>Bacteroides stercoris</i> ATCC 43183	NZ_ABFZ000000000	Yes
<i>Bacteroides thetaiotaomicron</i> 1_1_6	NZ_ACIC000000000	Yes
<i>Bacteroides thetaiotaomicron</i> 1_1_14	NZ_ACRP000000000	Yes
<i>Bacteroides thetaiotaomicron</i> 3731	NC_Bthetaitaomicron3731	
<i>Bacteroides thetaiotaomicron</i> 7330	NC_Bthetaitaomicron7330	Yes
<i>Bacteroides thetaiotaomicron</i> VPI-C11-15	SAMN18350527	
<i>Bacteroides thetaiotaomicron</i> VPI-0633-1	SAMN18350524	
<i>Bacteroides thetaiotaomicron</i> VPI-0940-1	SAMN18350525	
<i>Bacteroides thetaiotaomicron</i> VPI-5482	NC_004703	
<i>Bacteroides thetaiotaomicron</i> VPI-J19-343	SAMN18350526	

Table 3.3. Continued.

Organism	RefSeq Accession	Predicted <i>btu</i> MGE?
<i>Bacteroides uniformis</i> VPI-61	SAMN18350528	
<i>Bacteroides uniformis</i> VPI-909	SAMN18350529	
<i>Bacteroides uniformis</i> VPI-C20-25	SAMN18350530	
<i>Bacteroides uniformis</i> VPI-C7-17	SAMN18350531	
<i>Bacteroides dorei</i> VPI-2277 ^o	SAMN18350535	
<i>Bacteroides dorei</i> VPI-3776A ^o	SAMN18350536	
<i>Bacteroides nordii</i> WAL7936*	SAMN18350534	
<i>Bacteroides uniformis</i> 4_1_36	NZ_ACTC000000000	
<i>Bacteroides uniformis</i> ATCC 8492	NZ_AAYH000000000	
<i>Bacteroides uniformis</i> CL03T00C23	NZ_AGXX000000000	
<i>Bacteroides uniformis</i> CL03T12C37	NZ_AGXY000000000	
<i>Bacteroides uniformis</i> str. VPI-R5-33	SAMN18350532	
<i>Bacteroides vulgatus</i> 3_1_40A	NZ_ACRT000000000	Yes
<i>Bacteroides vulgatus</i> 4_3_47FAA	NZ_ACDR000000000	
<i>Bacteroides vulgatus</i> ATCC 8482	NC_009614	
<i>Bacteroides vulgatus</i> CL09T03C04	NZ_AGXZ000000000	
<i>Bacteroides vulgatus</i> PC510	NZ_ADKO000000000	
<i>Bacteroides vulgatus</i> str. VPI-2365	SAMN14525940	
<i>Bacteroides vulgatus</i> str. VPI-4025	SAMN14525944	
<i>Bacteroides vulgatus</i> str. VPI-4496.2	SAMN14525945	
<i>Bacteroides vulgatus</i> str. VPI-5710	SAMN14525942	
<i>Bacteroides dorei</i> str. VPI-6598B	SAMN18350537	
<i>Bacteroides vulgatus</i> VPI-4245	SAMN14525939	
<i>Bacteroides vulgatus</i> VPI-4506	SAMN14525946	
<i>Bacteroides vulgatus</i> VPI-6186	SAMN14525941	
<i>Bacteroides</i> WH2	NC_BWH2	
<i>Bacteroides xylanisolvens</i> 1_1_30	NZ_ADCL000000000	Yes
<i>Bacteroides xylanisolvens</i> 2_1_22	NZ_ACPQ000000000	
<i>Bacteroides xylanisolvens</i> CL03T12C04	NZ_AGXE000000000	
<i>Bacteroides xylanisolvens</i> D1	NZ_ACAB000000000	
<i>Bacteroides xylanisolvens</i> SD CC 1b	NZ_ADKP000000000	
<i>Bacteroides xylanisolvens</i> XB1A	NC_BxylanisolvensXB1A	
<i>Dysgonomonas gadei</i> ATCC BAA-286	NZ_ADLV000000000	
<i>Dysgonomonas mossii</i> DSM 22836	NZ_ADLW000000000	
<i>Odoribacter laneus</i> YIT 12061	NZ_ADMC000000000	
<i>Odoribacter splanchnicus</i> DSM 20712	NC_015160	
<i>Parabacteroides distasonis</i> ATCC 8503	NC_009615	
<i>Parabacteroides distasonis</i> CL03T12C09	NZ_AGZM000000000	

Table 3.3. Continued.

Organism	RefSeq Accession	Predicted <i>btu</i> MGE?
<i>Parabacteroides distasonis</i> CL09T03C24	NZ_AGZN000000000	
<i>Parabacteroides goldsteinii</i> CL02T12C30	NZ_AGZO000000000	
<i>Parabacteroides johnsonii</i> CL02T12C29	NZ_AGZP000000000	
<i>Parabacteroides johnsonii</i> DSM 18315	NZ_ABYH000000000	
<i>Parabacteroides merdae</i> ATCC 43184	NZ_AAXE000000000	
<i>Parabacteroides merdae</i> CL03T12C32	NZ_AGZQ000000000	
<i>Parabacteroides merdae</i> CL09T00C40	NZ_AJPU000000000	
<i>Parabacteroides</i> sp. D13	NZ_ACPW000000000	
<i>Parabacteroides</i> sp. D25	NZ_ACUA000000000	
<i>Paraprevotella clara</i> YIT 11840	NZ_AFFY000000000	
<i>Paraprevotella xylaniphila</i> YIT 11841	NZ_AFBR000000000	
<i>Porphyromonas gingivalis</i> ATCC 33277	NC_010729	
<i>Porphyromonas gingivalis</i> TDC60	NC_015571	
<i>Porphyromonas gingivalis</i> W83	NC_002950	
<i>Prevotella copri</i> DSM 18205	NZ_ACBX000000000	
<i>Prevotella salivae</i> DSM 15606	NZ_AEQO000000000	
<i>Prevotella stercorea</i> DSM 18206	NZ_AFZZ000000000	
<i>Tannerella forsythia</i> ATCC 43037	NC_016610	
<i>Tannerella</i> sp. 6_1_58FAA_CT1	NZ_ACWX000000000	Yes

* Originally deposited in the VPI strain collection as *B. uniformis*

° Originally deposited in VPI strain collection as *B. vulgatus*

Table 3.4. CTn or Phage Associated

PFAMs. PFAM	Description
CTn/ICE PFAMs	
PF00078	Reverse transcriptase (RNA-dependent DNA polymerase)
PF00142	NifH/frxC family
PF00216	Bacterial DNA-binding protein
PF00239	Resolvase; N terminal domain
PF00436	Single-strand binding protein family
PF00589	Phage integrase family
PF01076	Plasmid recombination enzyme
PF01131	DNA topoisomerase
PF01348	Type II intron maturase
PF01555	DNA methylase
PF01610	Transposase
PF01656	CobQ/CobB/MinD/ParA nucleotide binding domain
PF01751	Toprim domain
PF02195	ParB-like nuclease domain
PF02534	Type IV secretory system Conjugative DNA transfer
PF03118	Bacterial RNA polymerase; alpha chain C terminal
PF03432	Relaxase/Mobilisation nuclease domain
PF04335	VirB8 protein
PF04610	TrbL/VirB6 plasmid conjugal transfer protein
PF05857	TraX protein
PF06745	B. Theta Gene description (DUF3875)
PF07751	Abi-like protein
PF07863	Homologues of TraJ from Bacteroides conjugative transposon
PF09140	ATPase MipZ
PF09357	RteC protein
PF10412	Type IV secretion-system coupling protein DNA-binding domain
PF10623	Plasmid conjugative transfer protein Pill
PF10626	Conjugative transposon protein TraO
PF11888	Protein of unknown function (DUF3408)
PF12508	Conjugative transposon; TraM
PF12696	TraM recognition site of TraD and TraG
PF12728	Helix-turn-helix domain
PF12846	AAA-like domain
PF12988	Domain of unknown function; B. Theta Gene descript
PF12991	Domain of unknown function
PF12992	Domain of unknown function; B. Theta Gene descript
PF13102	Phage integrase SAM-like domain
PF13150	Protein of unknown function (DUF3989)

Table 3.4. Continued.

PFAM	Description
PF13155	Toprim-like
PF13481	AAA domain
PF13571	Domain of unknown function (DUF4133)
PF13572	Domain of unknown function (DUF4134)
PF13595	Domain of unknown function (DUF4138)
PF13605	Domain of unknown function (DUF4141)
PF13614	AAA domain
PF17293	Arm DNA-binding domain
Phage PFAMs	
PF00004	ATPase family associated with various cellular activities (AAA)
PF00005	ABC transporter
PF00006	ATP synthase alpha/beta family, nucleotide-binding domain
PF00009	Elongation factor Tu GTP binding domain
PF00011	Hsp20/alpha crystallin family
PF00012	Hsp70 protein
PF00015	Methyl-accepting chemotaxis protein (MCP) signalling domain
PF00023	Ankyrin repeat
PF00037	4Fe-4S binding domain
PF00041	Fibronectin type III domain
PF00067	Cytochrome P450
PF00069	Protein kinase domain
PF00072	Response regulator receiver domain
PF00075	RNase H
PF00078	Reverse transcriptase (RNA-dependent DNA polymerase)
PF00080	Copper/zinc superoxide dismutase (SODC)
PF00082	Subtilase family
PF00083	Sugar (and other) transporter
PF00085	Thioredoxin
PF00089	Trypsin
PF00091	Tubulin/FtsZ family, GTPase domain
PF00092	von Willebrand factor type A domain
PF00106	short chain dehydrogenase
PF00109	Beta-ketoacyl synthase, N-terminal domain
PF00111	2Fe-2S iron-sulfur cluster binding domain
PF00112	Papain family cysteine protease
PF00113	Enolase, C-terminal TIM barrel domain
PF00117	Glutamine amidotransferase class-I
PF00118	TCP-1/cpn60 chaperonin family
PF00119	ATP synthase A chain
PF00122	E1-E2 ATPase
PF00124	Photosynthetic reaction centre protein

Table 3.4. Continued.

PFAM	Description
PF00126	Bacterial regulatory helix-turn-helix protein, lysR family
PF00127	Copper binding proteins, plastocyanin/azurin family
PF00132	Bacterial transferase hexapeptide (six repeats)
PF00133	tRNA synthetases class I (I, L, M and V)
PF00136	DNA polymerase family B
PF00144	Beta-lactamase
PF00145	C-5 cytosine-specific DNA methylase
PF00149	Calcineurin-like phosphoesterase
PF00150	Cellulase (glycosyl hydrolase family 5)
PF00152	tRNA synthetases class II (D, K and N)
PF00154	recA bacterial DNA recombination protein
PF00155	Aminotransferase class I and II
PF00156	Phosphoribosyl transferase domain
PF00161	Ribosome inactivating protein
PF00165	Bacterial regulatory helix-turn-helix proteins, AraC family
PF00166	Chaperonin 10 Kd subunit
PF00171	Aldehyde dehydrogenase family
PF00176	SNF2 family N-terminal domain
PF00180	Isocitrate/isopropylmalate dehydrogenase
PF00182	Chitinase class I
PF00183	Hsp90 protein
PF00186	Dihydrofolate reductase
PF00188	Cysteine-rich secretory protein family
PF00195	Chalcone and stilbene synthases, N-terminal domain
PF00196	Bacterial regulatory proteins, luxR family
PF00202	Aminotransferase class-III
PF00204	DNA gyrase B
PF00206	Lyase
PF00210	Ferritin-like domain
PF00213	ATP synthase delta (OSCP) subunit
PF00215	Orotidine 5'-phosphate decarboxylase / HUMPS family
PF00216	Bacterial DNA-binding protein
PF00218	Indole-3-glycerol phosphate synthase
PF00223	Photosystem I psaA/psaB protein
PF00224	Pyruvate kinase, barrel domain
PF00226	DnaJ domain
PF00231	ATP synthase
PF00232	Glycosyl hydrolase family 1
PF00239	Resolvase, N terminal domain
PF00246	Zinc carboxypeptidase
PF00248	Aldo/keto reductase family

Table 3.4. Continued.

PFAM	Description
PF00251	Glycosyl hydrolases family 32 N-terminal domain
PF00258	Flavodoxin
PF00263	Bacterial type II and III secretion system protein
PF00264	Common central domain of tyrosinase
PF00265	Thymidine kinase
PF00266	Aminotransferase class-V
PF00267	Gram-negative porin
PF00268	Ribonucleotide reductase, small chain
PF00269	Small, acid-soluble spore proteins, alpha/beta type
PF00270	DEAD/DEAH box helicase
PF00271	Helicase conserved C-terminal domain
PF00274	Fructose-bisphosphate aldolase class-I
PF00285	Citrate synthase, C-terminal domain
PF00288	GHMP kinases N terminal domain
PF00291	Pyridoxal-phosphate dependent enzyme
PF00293	NUDIX domain
PF00294	pfkB family carbohydrate kinase
PF00300	Histidine phosphatase superfamily (branch 1)
PF00302	Chloramphenicol acetyltransferase
PF00303	Thymidylate synthase
PF00308	Bacterial dnaA protein
PF00313	'Cold-shock' DNA-binding domain
PF00317	Ribonucleotide reductase, all-alpha domain
PF00324	Amino acid permease
PF00334	Nucleoside diphosphate kinase
PF00352	Transcription factor TFIID (or TATA-binding protein, TBP)
PF00353	RTX calcium-binding nonapeptide repeat (4 copies)
PF00358	phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 1
PF00359	Phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 2
PF00367	phosphotransferase system, EIIB
PF00376	MerR family regulatory protein
PF00381	PTS HPr component phosphorylation site
PF00382	Transcription factor TFIIB repeat
PF00383	Cytidine and deoxycytidylate deaminase zinc-binding region
PF00384	Molybdopterin oxidoreductase
PF00389	D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain
PF00391	PEP-utilising enzyme, mobile domain
PF00392	Bacterial regulatory proteins, gntR family
PF00398	Ribosomal RNA adenine dimethylase
PF00406	Adenylate kinase
PF00415	Regulator of chromosome condensation (RCC1) repeat

Table 3.4. Continued.

PFAM	Description
PF00430	ATP synthase B/B' CF(0)
PF00436	Single-strand binding protein family
PF00438	S-adenosylmethionine synthetase, N-terminal domain
PF00440	Bacterial regulatory proteins, tetR family
PF00448	SRP54-type protein, GTPase domain
PF00453	Ribosomal protein L20
PF00455	DeoR C terminal sensor domain
PF00462	Glutaredoxin
PF00464	Serine hydroxymethyltransferase
PF00468	Ribosomal protein L34
PF00471	Ribosomal protein L33
PF00472	RF-1 domain
PF00474	Sodium:solute symporter family
PF00476	DNA polymerase family A
PF00478	IMP dehydrogenase / GMP reductase domain
PF00483	Nucleotidyl transferase
PF00486	Transcriptional regulatory protein, C terminal
PF00487	Fatty acid desaturase
PF00493	MCM P-loop domain
PF00501	AMP-binding enzyme
PF00502	Phycobilisome protein
PF00520	Ion transport protein
PF00521	DNA gyrase/topoisomerase IV, subunit A
PF00528	Binding-protein-dependent transport system inner membrane component
PF00529	HlyD membrane-fusion protein of T1SS
PF00533	BRCA1 C Terminus (BRCT) domain
PF00534	Glycosyl transferases group 1
PF00535	Glycosyl transferase family 2
PF00542	Ribosomal protein L7/L12 C-terminal domain
PF00550	Phosphopantetheine attachment site
PF00551	Formyl transferase
PF00557	Metallopeptidase family M24
PF00561	alpha/beta hydrolase fold
PF00562	RNA polymerase Rpb2, domain 6
PF00563	EAL domain
PF00565	Staphylococcal nuclease homologue
PF00571	CBS domain
PF00574	Clp protease
PF00575	S1 RNA binding domain
PF00578	AhpC/TSA family
PF00579	tRNA synthetases class I (W and Y)

Table 3.4. Continued.

PFAM	Description
PF00580	UvrD/REP helicase N-terminal domain
PF00581	Rhodanese-like domain
PF00582	Universal stress protein family
PF00583	Acetyltransferase (GNAT) family
PF00586	AIR synthase related protein, N-terminal domain
PF00587	tRNA synthetase class II core domain (G, H, P, S and T)
PF00588	SpoU rRNA Methylase family
PF00589	Phage integrase family
PF00590	Tetrapyrrole (Corrin/Porphyrin) Methylases
PF00591	Glycosyl transferase family, a/b domain
PF00593	TonB dependent receptor
PF00596	Class II Aldolase and Adducin N-terminal domain
PF00622	SPRY domain
PF00623	RNA polymerase Rpb1, domain 2
PF00625	Guanylate kinase
PF00644	Poly(ADP-ribose) polymerase catalytic domain
PF00654	Voltage gated chloride channel
PF00664	ABC transporter transmembrane region
PF00665	Integrase core domain
PF00668	Condensation domain
PF00669	Bacterial flagellin N-terminal helical region
PF00675	Insulinase (Peptidase family M16)
PF00676	Dehydrogenase E1 component
PF00677	Lumazine binding domain
PF00680	Viral RNA-dependent RNA polymerase
PF00691	OmpA family
PF00692	dUTPase
PF00696	Amino acid kinase family
PF00704	Glycosyl hydrolases family 18
PF00705	Proliferating cell nuclear antigen, N-terminal domain
PF00707	Translation initiation factor IF-3, C-terminal domain
PF00709	Adenylosuccinate synthetase
PF00712	DNA polymerase III beta subunit, N-terminal domain
PF00717	Peptidase S24-like
PF00719	Inorganic pyrophosphatase
PF00722	Glycosyl hydrolases family 16
PF00728	Glycosyl hydrolase family 20, catalytic domain
PF00730	HhH-GPD superfamily base excision DNA repair protein
PF00733	Asparagine synthase
PF00749	tRNA synthetases class I (E and Q), catalytic domain
PF00753	Metallo-beta-lactamase superfamily

Table 3.4. Continued.

PFAM	Description
PF00754	F5/8 type C domain
PF00756	Putative esterase
PF00765	Autoinducer synthase
PF00768	D-alanyl-D-alanine carboxypeptidase
PF00772	DnaB-like helicase N terminal domain
PF00773	RNB domain
PF00782	Dual specificity phosphatase, catalytic domain
PF00793	DAHPh synthetase I family
PF00801	PKD domain
PF00805	Pentapeptide repeats (8 copies)
PF00809	Pterin binding enzyme
PF00814	tRNA N6-adenosine threonylcarbamoyltransferase
PF00816	H-NS histone family
PF00817	impB/mucB/samB family
PF00823	PPE family
PF00825	Ribonuclease P
PF00830	Ribosomal L28 family
PF00834	Ribulose-phosphate 3 epimerase family
PF00847	AP2 domain
PF00849	RNA pseudouridylate synthase
PF00850	Histone deacetylase domain
PF00852	Glycosyltransferase family 10 (fucosyltransferase) C-term
PF00854	POT family
PF00856	SET domain
PF00857	Isochorismatase family
PF00860	Permease family
PF00872	Transposase, Mutator family
PF00873	AcrB/AcrD/AcrF family
PF00874	PRD domain
PF00877	NlpC/P60 family
PF00881	Nitroreductase family
PF00883	Cytosol aminopeptidase family, catalytic domain
PF00884	Sulfatase
PF00885	6,7-dimethyl-8-ribityllumazine synthase
PF00890	FAD binding domain
PF00892	EamA-like transporter family
PF00893	Small Multidrug Resistance protein
PF00899	ThiF family
PF00908	dTDP-4-dehydrorhamnose 3,5-epimerase
PF00912	Transglycosylase
PF00916	Sulfate permease family

Table 3.4. Continued.

PFAM	Description
PF00923	Transaldolase/Fructose-6-phosphate aldolase
PF00926	3,4-dihydroxy-2-butanone 4-phosphate synthase
PF00929	Exonuclease
PF00939	Sodium:sulfate symporter transmembrane region
PF00940	DNA-dependent RNA polymerase
PF00941	FAD binding domain in molybdopterin dehydrogenase
PF00958	GMP synthase C terminal domain
PF00959	Phage lysozyme
PF00984	UDP-glucose/GDP-mannose dehydrogenase family, central domain
PF00986	DNA gyrase B subunit, carboxyl terminus
PF00990	Diguanylate cyclase, GGDEF domain
PF00999	Sodium/hydrogen exchanger family
PF01011	PQQ enzyme repeat
PF01022	Bacterial regulatory protein, arsR family
PF01024	Colicin pore forming domain
PF01025	GrpE
PF01027	Inhibitor of apoptosis-promoting Bax1
PF01028	Eukaryotic DNA topoisomerase I, catalytic core
PF01032	FecCD transport family
PF01035	6-O-methylguanine DNA methyltransferase, DNA binding domain
PF01037	Lrp/AsnC ligand binding domain
PF01041	DegT/DnrJ/EryC1/StrS aminotransferase family
PF01042	Endoribonuclease L-PSP
PF01048	Phosphorylase superfamily
PF01050	Mannose-6-phosphate isomerase
PF01051	Initiator Replication protein
PF01055	Glycosyl hydrolases family 31
PF01063	Amino-transferase class IV
PF01068	ATP dependent DNA ligase domain
PF01070	FMN-dependent dehydrogenase
PF01075	Glycosyltransferase family 9 (heptosyltransferase)
PF01076	Plasmid recombination enzyme
PF01077	Nitrite and sulphite reductase 4Fe-4S domain
PF01081	KDPG and KHG aldolase
PF01106	NifU-like domain
PF01119	DNA mismatch repair protein, C-terminal domain
PF01124	MAPEG family
PF01126	Heme oxygenase
PF01129	NAD:arginine ADP-ribosyltransferase
PF01131	DNA topoisomerase
PF01134	Glucose inhibited division protein A

Table 3.4. Continued.

PFAM	Description
PF01135	Protein-L-isoaspartate(D-aspartate) O-methyltransferase (PCMT)
PF01138	3' exoribonuclease family, domain 1
PF01139	tRNA-splicing ligase RtcB
PF01145	SPFH domain / Band 7 family
PF01149	Formamidopyrimidine-DNA glycosylase N-terminal domain
PF01165	Ribosomal protein S21
PF01171	PP-loop family
PF01176	Translation initiation factor 1A / IF-1
PF01177	Asp/Glu/Hydantoin racemase
PF01179	Copper amine oxidase, enzyme domain
PF01182	Glucosamine-6-phosphate isomerases/6-phosphogluconolactonase
PF01183	Glycosyl hydrolases family 25
PF01192	RNA polymerase Rpb6
PF01195	Peptidyl-tRNA hydrolase
PF01205	Uncharacterized protein family UPF0029
PF01207	Dihydrouridine synthase (Dus)
PF01212	Beta-eliminating lyase
PF01223	DNA/RNA non-specific endonuclease
PF01225	Mur ligase family, catalytic domain
PF01226	Formate/nitrite transporter
PF01227	GTP cyclohydrolase I
PF01228	Glycine radical
PF01230	HIT domain
PF01242	6-pyruvoyl tetrahydropterin synthase
PF01253	Translation initiation factor SUI1
PF01255	Putative undecaprenyl diphosphate synthase
PF01258	Prokaryotic dksA/traR C4-type zinc finger
PF01259	SAICAR synthetase
PF01261	Xylose isomerase-like TIM barrel
PF01266	FAD dependent oxidoreductase
PF01268	Formate-tetrahydrofolate ligase
PF01272	Transcription elongation factor, GreA/GreB, C-term
PF01292	Prokaryotic cytochrome b561
PF01293	Phosphoenolpyruvate carboxykinase
PF01306	LacY proton/sugar symporter
PF01314	Aldehyde ferredoxin oxidoreductase, domains 2 & 3
PF01326	Pyruvate phosphate dikinase, AMP/ATP-binding domain
PF01327	Polypeptide deformylase
PF01343	Peptidase family S49
PF01344	Kelch motif
PF01353	Green fluorescent protein

Table 3.4. Continued.

PFAM	Description
PF01367	5'-3' exonuclease, C-terminal SAM fold
PF01368	DHH family
PF01370	NAD dependent epimerase/dehydratase family
PF01374	Glycosyl hydrolase family 46
PF01375	Heat-labile enterotoxin alpha chain
PF01376	Heat-labile enterotoxin beta chain
PF01380	SIS domain
PF01381	Helix-turn-helix
PF01385	Probable transposase
PF01391	Collagen triple helix repeat (20 copies)
PF01402	Ribbon-helix-helix protein, copG family
PF01420	Type I restriction modification DNA specificity domain
PF01432	Peptidase family M3
PF01433	Peptidase family M1 domain
PF01434	Peptidase family M41
PF01435	Peptidase family M48
PF01442	Apolipoprotein A1/A4/E domain
PF01446	Replication protein
PF01451	Low molecular weight phosphotyrosine protein phosphatase
PF01458	SUF system FeS cluster assembly, SufBD
PF01464	Transglycosylase SLT domain
PF01467	Cytidylyltransferase-like
PF01469	Pentapeptide repeats (8 copies)
PF01471	Putative peptidoglycan binding domain
PF01473	Putative cell wall binding repeat
PF01476	LysM domain
PF01487	Type I 3-dehydroquinase
PF01494	FAD binding domain
PF01501	Glycosyl transferase family 8
PF01503	Phosphoribosyl-ATP pyrophosphohydrolase
PF01507	Phosphoadenosine phosphosulfate reductase family
PF01510	N-acetylmuramoyl-L-alanine amidase
PF01513	ATP-NAD kinase
PF01515	Phosphate acetyl/butaryl transferase
PF01520	N-acetylmuramoyl-L-alanine amidase
PF01521	Iron-sulphur cluster biosynthesis
PF01522	Polysaccharide deacetylase
PF01526	Tn3 transposase DDE domain
PF01527	Transposase
PF01531	Glycosyl transferase family 11
PF01541	GIY-YIG catalytic domain

Table 3.4. Continued.

PFAM	Description
PF01544	CorA-like Mg ²⁺ transporter protein
PF01545	Cation efflux family
PF01546	Peptidase family M20/M25/M40
PF01548	Transposase
PF01551	Peptidase family M23
PF01555	DNA methylase
PF01556	DnaJ C terminal domain
PF01569	PAP2 superfamily
PF01575	MaoC like domain
PF01578	Cytochrome C assembly protein
PF01580	FtsK/SpoIIIE family
PF01583	Adenylylsulphate kinase
PF01584	CheW-like domain
PF01592	NifU-like N terminal domain
PF01595	Cyclin M transmembrane N-terminal domain
PF01596	O-methyltransferase
PF01597	Glycine cleavage H-protein
PF01609	Transposase DDE domain
PF01610	Transposase
PF01612	3'-5' exonuclease
PF01618	MotA/TolQ/ExbB proton channel family
PF01619	Proline dehydrogenase
PF01624	MutS domain I
PF01625	Peptide methionine sulfoxide reductase
PF01632	Ribosomal protein L35
PF01634	ATP phosphoribosyltransferase
PF01636	Phosphotransferase enzyme family
PF01638	HxlR-like helix-turn-helix
PF01641	SelR domain
PF01653	NAD-dependent DNA ligase adenylation domain
PF01656	CobQ/CobB/MinD/ParA nucleotide binding domain
PF01658	Myo-inositol-1-phosphate synthase
PF01661	Macro domain
PF01668	SmpB protein
PF01676	Metalloenzyme superfamily
PF01679	Proteolipid membrane potential modulator
PF01694	Rhomboid family
PF01695	IstB-like ATP binding protein
PF01699	Sodium/calcium exchanger protein
PF01712	Deoxynucleoside kinase
PF01713	Smr domain

Table 3.4. Continued.

PFAM	Description
PF01719	Plasmid replication protein
PF01724	Domain of unknown function DUF29
PF01725	Ham1 family
PF01726	LexA DNA binding domain
PF01734	Patatin-like phospholipase
PF01742	Clostridial neurotoxin zinc protease
PF01743	Poly A polymerase head domain
PF01751	Toprim domain
PF01755	Glycosyltransferase family 25 (LPS biosynthesis protein)
PF01757	Acyltransferase family
PF01764	Lipase (class 3)
PF01786	Alternative oxidase
PF01797	Transposase IS200 like
PF01804	Penicillin amidase
PF01807	CHC2 zinc finger
PF01808	AICARFT/IMPCHase bienzyme
PF01809	Putative membrane protein insertion efficiency factor
PF01814	Hemerythrin HHE cation binding domain
PF01817	Chorismate mutase type II
PF01818	Bacteriophage translational regulator
PF01832	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase
PF01844	HNH endonuclease
PF01848	Hok/gef family
PF01850	PIN domain
PF01867	CRISPR associated protein Cas1
PF01870	Archaeal holliday junction resolvase (hjc)
PF01872	RibD C-terminal domain
PF01881	CRISPR associated protein Cas6
PF01885	RNA 2'-phosphotransferase, Tpt1 / KptA family
PF01895	PhoU domain
PF01904	Protein of unknown function DUF72
PF01906	Putative heavy-metal-binding
PF01909	Nucleotidyltransferase domain
PF01921	tRNA synthetases class I (K)
PF01923	Cobalamin adenosyltransferase
PF01925	Sulfite exporter TauE/SafE
PF01926	50S ribosome-binding GTPase
PF01928	CYTH domain
PF01936	NYN domain
PF01939	Endonuclease NucS
PF01965	DJ-1/Pfpl family

Table 3.4. Continued.

PFAM	Description
PF01972	Serine dehydrogenase proteinase
PF01979	Amidohydrolase family
PF01981	Peptidyl-tRNA hydrolase PTH2
PF01988	VIT family
PF02011	Glycosyl hydrolase family 48
PF02012	BNR/Asp-box repeat
PF02018	Carbohydrate binding domain
PF02026	RyR domain
PF02028	BCCT, betaine/carnitine/choline family transporter
PF02040	Arsenical pump membrane protein
PF02052	Gallidermin
PF02056	Family 4 glycosyl hydrolase
PF02061	Lambda Phage CIII
PF02073	Thermophilic metalloprotease (M29)
PF02075	Crossover junction endodeoxyribonuclease RuvC
PF02082	Iron-dependent Transcriptional regulator
PF02086	D12 class N6 adenine-specific DNA methyltransferase
PF02133	Permease for cytosine/purines, uracil, thiamine, allantoin
PF02142	MGS-like domain
PF02146	Sir2 family
PF02156	Glycosyl hydrolase family 26
PF02195	ParB-like nuclease domain
PF02223	Thymidylate kinase
PF02229	Transcriptional Coactivator p15 (PC4)
PF02258	Shiga-like toxin beta subunit
PF02261	Aspartate decarboxylase
PF02264	LamB porin
PF02294	7kD DNA-binding domain
PF02303	Helix-destabilising protein
PF02305	Capsid protein (F protein)
PF02306	Major spike protein (G protein)
PF02308	MgtC family
PF02311	AraC-like ligand binding domain
PF02316	Mu DNA-binding domain
PF02329	Histidine carboxylase PI chain
PF02334	Replication terminator protein
PF02342	TerD domain
PF02347	Glycine cleavage system P-protein
PF02348	Cytidyltransferase
PF02357	Transcription termination factor nusG
PF02367	Threonylcarbamoyl adenosine biosynthesis protein TsaE

Table 3.4. Continued.

PFAM	Description
PF02368	Bacterial Ig-like domain (group 2)
PF02369	Bacterial Ig-like domain (group 1)
PF02371	Transposase IS116/IS110/IS902 family
PF02378	Phosphotransferase system, EIIC
PF02384	N-6 DNA Methylase
PF02399	Origin of replication binding protein
PF02413	Caudovirales tail fibre assembly protein, lambda gpK
PF02417	Chromate transporter
PF02427	Photosystem I reaction centre subunit IV / PsaE
PF02441	Flavoprotein
PF02452	PemK-like, MazF-like toxin of type II toxin-antitoxin system
PF02463	RecF/RecN/SMC N terminal domain
PF02464	Competence-damaged protein
PF02467	Transcription factor WhiB
PF02472	Biopolymer transport protein ExbD/TolR
PF02481	DNA recombination-mediator protein A
PF02482	Sigma 54 modulation protein / S30EA ribosomal protein
PF02486	Replication initiation factor
PF02493	MORN repeat
PF02498	BRO family, N-terminal domain
PF02502	Ribose/Galactose Isomerase
PF02507	Photosystem I reaction centre subunit III
PF02511	Thymidylate synthase complementing protein
PF02517	CPBP intramembrane metalloprotease
PF02518	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
PF02522	Aminoglycoside 3-N-acetyltransferase
PF02525	Flavodoxin-like fold
PF02527	rRNA small subunit methyltransferase G
PF02531	PsaD
PF02537	CrcB-like protein, Camphor Resistance (CrcB)
PF02540	NAD synthase
PF02543	Carbamoyltransferase N-terminus
PF02556	Preprotein translocase subunit SecB
PF02557	D-alanyl-D-alanine carboxypeptidase
PF02562	PhoH-like protein
PF02572	ATP:corrinoid adenosyltransferase BtuR/CobO/CobP
PF02585	GlcNAc-PI de-N-acetylase
PF02586	SOS response associated peptidase (SRAP)
PF02588	Uncharacterised 5xTM membrane BCR, YitT family COG1284
PF02592	Putative vitamin uptake transporter
PF02595	Glycerate kinase family

Table 3.4. Continued.

PFAM	Description
PF02599	Global regulator protein family
PF02600	Disulfide bond formation protein DsbB
PF02601	Exonuclease VII, large subunit
PF02604	Antitoxin Phd_YefM, type II toxin-antitoxin system
PF02607	B12 binding domain
PF02608	ABC transporter substrate-binding protein PnrA-like
PF02609	Exonuclease VII small subunit
PF02613	Nitrate reductase delta subunit
PF02617	ATP-dependent Clp protease adaptor protein ClpS
PF02632	BioY family
PF02635	DsrE/DsrF-like family
PF02641	Uncharacterized ACR, COG1993
PF02643	Uncharacterized ACR, COG1430
PF02645	Uncharacterised protein, DegV family COG1307
PF02650	WhiA C-terminal HTH domain
PF02652	L-lactate permease
PF02653	Branched-chain amino acid transport system / permease component
PF02656	Domain of unknown function (DUF202)
PF02661	Fic/DOC family
PF02666	Phosphatidylserine decarboxylase
PF02668	Taurine catabolism dioxygenase TauD, TfdA family
PF02672	CP12 domain
PF02675	S-adenosylmethionine decarboxylase
PF02694	Uncharacterised BCR, YnfA/UPF0060 family
PF02696	Uncharacterized ACR, YdiU/UPF0061 family
PF02700	Phosphoribosylformylglycinamide (FGAM) synthase
PF02705	K ⁺ potassium transporter
PF02706	Chain length determinant protein
PF02732	ERCC4 domain
PF02735	Ku70/Ku80 beta-barrel domain
PF02738	Molybdopterin-binding domain of aldehyde dehydrogenase
PF02739	5'-3' exonuclease, N-terminal resolvase-like domain
PF02763	Diphtheria toxin, C domain
PF02767	DNA polymerase III beta subunit, central domain
PF02768	DNA polymerase III beta subunit, C-terminal domain
PF02769	AIR synthase related protein, C-terminal domain
PF02772	S-adenosylmethionine synthetase, central domain
PF02773	S-adenosylmethionine synthetase, C-terminal domain
PF02776	Thiamine pyrophosphate enzyme, N-terminal TPP binding domain
PF02777	Iron/manganese superoxide dismutases, C-terminal domain
PF02781	Glucose-6-phosphate dehydrogenase, C-terminal domain

Table 3.4. Continued.

PFAM	Description
PF02800	Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain
PF02811	PHP domain
PF02821	Staphylokinase/Streptokinase family
PF02823	ATP synthase, Delta/Epsilon chain, beta-sandwich domain
PF02826	D-isomer specific 2-hydroxyacid dehydrogenase, NAD binding domain
PF02831	gpW
PF02837	Glycosyl hydrolases family 2, sugar binding domain
PF02861	Clp amino terminal domain, pathogenicity island component
PF02867	Ribonucleotide reductase, barrel domain
PF02868	Thermolysin metallopeptidase, alpha-helical domain
PF02876	Staphylococcal/Streptococcal toxin, beta-grasp domain
PF02878	Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain I
PF02882	Tetrahydrofolate dehydrogenase/cyclohydrolase, NAD(P)-binding domain
PF02885	Glycosyl transferase family, helical bundle domain
PF02887	Pyruvate kinase, alpha/beta domain
PF02914	Bacteriophage Mu transposase
PF02916	DNA polymerase processivity factor
PF02917	Pertussis toxin, subunit 1
PF02924	Bacteriophage lambda head decoration protein D
PF02925	Bacteriophage scaffolding protein D
PF02945	Recombination endonuclease VII
PF02954	Bacterial regulatory protein, Fis family
PF02985	HEAT repeat
PF03009	Glycerophosphoryl diester phosphodiesterase family
PF03013	Pyrimidine dimer DNA glycosylase
PF03016	Exostosin family
PF03031	NLI interacting factor-like phosphatase
PF03050	Transposase IS66 family
PF03061	Thioesterase superfamily
PF03062	MBOAT, membrane-bound O-acyltransferase family
PF03102	NeuB family
PF03104	DNA polymerase family B, exonuclease domain
PF03109	ABC1 family
PF03120	NAD-dependent DNA ligase OB-fold domain
PF03160	Calx-beta domain
PF03161	LAGLIDADG DNA endonuclease family
PF03167	Uracil DNA glycosylase superfamily
PF03175	DNA polymerase type B, organellar and viral
PF03180	NLPA lipoprotein
PF03197	Bacteriophage FRD2 protein
PF03222	Tryptophan/tyrosine permease family

Table 3.4. Continued.

PFAM	Description
PF03230	Antirestriction protein
PF03235	Protein of unknown function DUF262
PF03237	Terminase-like family
PF03245	Bacteriophage Rz lysis protein
PF03275	UDP-galactopyranose mutase
PF03279	Bacterial lipid A biosynthesis acyltransferase
PF03318	Clostridium epsilon toxin ETX/Bacillus mosquitocidal toxin MTX2
PF03330	Lytic transglycolase
PF03332	Eukaryotic phosphomannomutase
PF03349	Outer membrane protein transport protein (OMPP1/FadL/TodX)
PF03350	Uncharacterized protein family, UPF0114
PF03354	Phage Terminase
PF03358	NADPH-dependent FMN reductase
PF03364	Polyketide cyclase / dehydrase and lipid transport
PF03372	Endonuclease/Exonuclease/phosphatase family
PF03374	Phage antirepressor protein KilAC domain
PF03382	Mycoplasma protein of unknown function, DUF285
PF03400	IS1 transposase
PF03406	Phage tail fibre repeat
PF03407	Nucleotide-diphospho-sugar transferase
PF03414	Glycosyltransferase family 6
PF03420	Prohead core protein serine protease
PF03428	Replication protein C N-terminal domain
PF03431	RNA replicase, beta-chain
PF03432	Relaxase/Mobilisation nuclease domain
PF03435	Saccharopine dehydrogenase NADP binding domain
PF03446	NAD binding domain of 6-phosphogluconate dehydrogenase
PF03458	UPF0126 domain
PF03462	PCRF domain
PF03466	LysR substrate binding domain
PF03472	Autoinducer binding domain
PF03477	ATP cone domain
PF03486	HI0933-like protein
PF03496	ADP-ribosyltransferase exoenzyme
PF03498	Cytolethal distending toxin A/C domain
PF03505	Clostridium enterotoxin
PF03534	Salmonella virulence plasmid 65kDa B protein
PF03544	Gram-negative bacterial TonB protein C-terminal
PF03551	Transcriptional regulator PadR-like family
PF03567	Sulfotransferase family
PF03572	Peptidase family S41

Table 3.4. Continued.

PFAM	Description
PF03589	Antitermination protein
PF03590	Aspartate-ammonia ligase
PF03592	Terminase small subunit
PF03596	Cadmium resistance transporter
PF03614	Repressor of phase-1 flagellin
PF03616	Sodium/glutamate symporter
PF03629	Carbohydrate esterase, sialic acid-specific acetylcysteine
PF03631	Virulence factor BrkB
PF03632	Glycosyl hydrolase family 65 central catalytic domain
PF03641	Possible lysine decarboxylase
PF03658	RnfH family Ubiquitin
PF03668	P-loop ATPase protein family
PF03691	Uncharacterised protein family (UPF0167)
PF03695	Uncharacterised protein family (UPF0149)
PF03703	Bacterial PH domain
PF03710	Glutamate-ammonia ligase adenylyltransferase
PF03721	UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain
PF03738	Glutathionylspermidine synthase preATP-grasp
PF03741	Integral membrane protein TerC family
PF03747	ADP-ribosylglycohydrolase
PF03755	YicC-like family, N-terminal region
PF03781	Sulfatase-modifying factor enzyme 1
PF03783	Curli production assembly/transport component CsgG
PF03796	DnaB-like helicase C terminal domain
PF03797	Autotransporter beta-domain
PF03816	LytR_cpsA_psr family
PF03819	MazG nucleotide pyrophosphohydrolase domain
PF03837	RecT family
PF03838	Recombination protein U
PF03851	UV-endonuclease UvdE
PF03852	DNA mismatch endonuclease Vsr
PF03863	Phage maturation protein
PF03864	Phage major capsid protein E
PF03865	Haemolysin secretion/activation protein ShIB/FhaC/HecB
PF03869	Arc-like DNA binding domain
PF03883	Peroxide stress protein YaaA
PF03887	YfbU domain
PF03895	YadA-like membrane anchor domain
PF03899	ATP synthase I chain
PF03903	Phage T4 tail fibre
PF03906	Phage T7 tail fibre protein

Table 3.4. Continued.

PFAM	Description
PF03960	ArsC family
PF03965	Penicillinase repressor
PF03976	Polyphosphate kinase 2 (PPK2)
PF03994	Domain of Unknown Function (DUF350)
PF04002	RadC-like JAB domain
PF04011	LemA family
PF04014	Antidote-toxin recognition MazE, bacterial antitoxin
PF04026	SpoVG
PF04055	Radical SAM superfamily
PF04069	Substrate binding domain of ABC-type glycine betaine transport system
PF04073	Aminoacyl-tRNA editing domain
PF04076	Bacterial OB fold (BOF) protein
PF04077	DsrH like protein
PF04085	rod shape-determining protein MreC
PF04093	rod shape-determining protein MreD
PF04095	Nicotinate phosphoribosyltransferase (NAPRTase) family
PF04098	Rad52/22 family double-strand break repair protein
PF04120	Low affinity iron permease
PF04122	ell wall binding domain 2 (CWB2)
PF04127	DNA / pantothenate metabolism flavoprotein
PF04138	GtrA-like protein
PF04149	Domain of unknown function (DUF397)
PF04175	Protein of unknown function (DUF406)
PF04221	RelB antitoxin
PF04228	Putative neutral zinc metallopeptidase
PF04229	GrpB protein
PF04232	Stage V sporulation protein S (SpoVS)
PF04233	Phage Mu protein F like protein
PF04245	37-kD nucleoid-associated bacterial protein
PF04255	Protein of unknown function (DUF433)
PF04279	Intracellular septation protein A
PF04285	Protein of unknown function (DUF444)
PF04293	SpoVR like protein
PF04297	Putative helix-turn-helix protein, YlxM / p13 like
PF04307	LexA-binding, inner membrane-associated putative hydrolase
PF04308	Ribonuclease H-like
PF04313	Type I restriction enzyme R protein N terminus (HSDR_N)
PF04321	RmlD substrate binding domain
PF04326	Putative DNA-binding domain
PF04327	Cysteine protease Prp
PF04332	Protein of unknown function (DUF475)

Table 3.4. Continued.

PFAM	Description
PF04333	MlaA lipoprotein
PF04343	Protein of unknown function, DUF488
PF04352	ProQ/FINO family
PF04355	SmpA / OmlA family
PF04358	DsrC like protein
PF04365	Ribonuclease toxin, BrnT, of type II toxin-antitoxin system
PF04367	Protein of unknown function (DUF502)
PF04381	Putative exonuclease, RdgC
PF04383	KilA-N domain
PF04386	Stringent starvation protein B
PF04397	LytTr DNA-binding domain
PF04404	ERF superfamily
PF04413	3-Deoxy-D-manno-octulosonic-acid transferase (kdottransferase)
PF04439	Streptomycin adenylyltransferase
PF04446	tRNA ^{His} guanylyltransferase
PF04447	Protein of unknown function (DUF550)
PF04448	Protein of unknown function (DUF551)
PF04466	Phage terminase large subunit
PF04471	Restriction endonuclease
PF04485	Phycobilisome degradation protein nbIA
PF04492	Bacteriophage replication protein O
PF04517	Microvirus lysis protein (E), C terminus
PF04519	Polymer-forming cytoskeletal
PF04531	Bacteriophage holin
PF04542	Sigma-70 region 2
PF04545	Sigma-70, region 4
PF04546	Sigma-70, non-essential region
PF04550	Phage holin family 2
PF04551	GcpE protein
PF04556	DpnII restriction endonuclease
PF04577	Protein of unknown function (DUF563)
PF04586	Caudovirus prohead serine protease
PF04606	Ogr/Delta-like zinc finger
PF04607	Region found in RelA / SpoT proteins
PF04608	Phosphatidylglycerophosphatase A
PF04630	Phage tail tube protein
PF04647	Accessory gene regulator B
PF04657	Putative inner membrane exporter, YdcZ
PF04686	Streptomyces sporulation and cell division protein, SsgA
PF04687	Microvirus H protein (pilot protein)
PF04688	SPP1 phage holin

Table 3.4. Continued.

PFAM	Description
PF04717	Type VI secretion system/phage-baseplate injector OB domain
PF04724	Glycosyltransferase family 17
PF04726	Microvirus J protein
PF04728	Lipoprotein leucine-zipper
PF04738	Lantibiotic dehydratase, N terminus
PF04740	LXG domain of WXG superfamily
PF04748	Divergent polysaccharide deacetylase
PF04754	Putative transposase, YhgA-like
PF04761	Lactococcus bacteriophage putative transcription regulator
PF04765	Protein of unknown function (DUF616)
PF04820	Tryptophan halogenase
PF04851	Type III restriction enzyme, res subunit
PF04860	Phage portal protein
PF04865	Baseplate J-like protein
PF04883	Bacteriophage HK97-gp10, putative tail-component
PF04932	O-Antigen ligase
PF04936	Protein of unknown function (DUF658)
PF04965	Gene 25-like lysozyme
PF04966	Carbohydrate-selective porin, OprB family
PF04971	Bacteriophage P21 holin S
PF04973	Nicotinamide mononucleotide transporter
PF04976	DMSO reductase anchor subunit (DmsC)
PF04984	Phage tail sheath protein subtilisin-like domain
PF04985	Phage tail tube protein FII
PF04998	RNA polymerase Rpb1, domain 5
PF05013	N-formylglutamate amidohydrolase
PF05014	Nucleoside 2-deoxyribosyltransferase
PF05015	RelE-like toxin of type II toxin-antitoxin system HigB
PF05016	ParE toxin of type II toxin-antitoxin system, parDE
PF05025	RbsD / FucU transport protein family
PF05035	2-keto-3-deoxy-galactonokinase
PF05037	Protein of unknown function (DUF669)
PF05050	Methyltransferase FkbM domain
PF05063	MT-A70
PF05065	Phage capsid family
PF05068	Mannitol repressor
PF05069	Phage virion morphogenesis family
PF05099	Tellurite resistance protein TerB
PF05100	Phage minor tail protein L
PF05105	Bacteriophage holin family
PF05106	Phage holin family (Lysis protein S)

Table 3.4. Continued.

PFAM	Description
PF05107	CRISPR-associated protein Cas7
PF05108	Type VII secretion system ESX-1, transport TM domain B
PF05118	Aspartyl/Asparaginyl beta-hydroxylase
PF05119	Phage terminase, small subunit
PF05125	Phage major capsid protein, P2 family
PF05133	Phage portal protein, SPP1 Gp6-like
PF05135	Phage gp6-like head-tail connector protein
PF05136	Phage portal protein, lambda family
PF05144	Phage replication protein CRI
PF05147	Lanthionine synthetase C-like protein
PF05154	TM2 domain
PF05155	Phage X family
PF05170	AsmA family
PF05175	Methyltransferase small domain
PF05183	RNA dependent RNA polymerase
PF05193	Peptidase M16 inactive domain
PF05198	Translation initiation factor IF-3, N-terminal domain
PF05203	Hom_end-associated Hint
PF05226	CHASE2 domain
PF05231	MASE1
PF05257	CHAP domain
PF05263	Protein of unknown function (DUF722)
PF05265	Protein of unknown function (DUF723)
PF05268	Phage tail fibre adhesin Gp38
PF05269	Bacteriophage CII protein
PF05272	Virulence-associated protein E
PF05305	Protein of unknown function (DUF732)
PF05322	NINE Protein
PF05339	Protein of unknown function (DUF739)
PF05343	M42 glutamyl aminopeptidase
PF05345	Putative Ig domain
PF05352	Phage Connector (GP10)
PF05354	Phage Head-Tail Attachment
PF05356	Inovirus Coat protein B
PF05357	Phage Coat Protein A
PF05358	DicB protein
PF05362	Lon protease (S16) C-terminal proteolytic domain
PF05367	Phage endonuclease I
PF05368	NmrA-like family
PF05371	*Merged into existing family
PF05382	Bacteriophage peptidoglycan hydrolase

Table 3.4. Continued.

PFAM	Description
PF05396	Phage T7 capsid assembly protein
PF05406	WGR domain
PF05430	S-adenosyl-L-methionine-dependent methyltransferase
PF05435	Phi-29 DNA terminal protein GP3
PF05443	ROS/MUCR transcriptional regulator protein
PF05449	Putative 3TM holin, Phage_holin_3
PF05464	Phi-29-like late genes activator (early protein GP4)
PF05488	PAAR motif
PF05489	Phage Tail Protein X
PF05496	Holliday junction DNA helicase RuvB P-loop domain
PF05521	Phage head-tail joining protein
PF05532	CsbD-like
PF05534	HicB family
PF05565	Siphovirus Gp157
PF05588	Clostridium botulinum HA-17 domain
PF05593	RHS Repeat
PF05595	Domain of unknown function (DUF771)
PF05598	Transposase domain (DUF772)
PF05616	Neisseria meningitidis TspB protein
PF05635	23S rRNA-intervening sequence protein
PF05649	Peptidase family M13
PF05656	Protein of unknown function (DUF805)
PF05657	Protein of unknown function (DUF806)
PF05658	Head domain of trimeric autotransporter adhesin
PF05666	Fels-1 Prophage Protein-like
PF05690	Thiazole biosynthesis protein ThiG
PF05696	Protein of unknown function (DUF826)
PF05704	Capsular polysaccharide synthesis protein
PF05707	Zonular occludens toxin (Zot)
PF05708	Permuted papain-like amidase enzyme, YaeF/YiiX, C92 family
PF05709	Phage tail protein
PF05711	Macrocin-O-methyltransferase (TylF)
PF05717	IS66 Orf2 like protein
PF05721	Phytanoyl-CoA dioxygenase (PhyH)
PF05728	Uncharacterised protein family (UPF0227)
PF05731	TROVE domain
PF05732	Firmicute plasmid replication protein (RepL)
PF05737	Collagen binding domain
PF05766	Bacteriophage Lambda NinG protein
PF05772	NinB protein
PF05798	Bacteriophage FRD3 protein

Table 3.4. Continued.

PFAM	Description
PF05810	NinF protein
PF05816	Toxic anion resistance protein (TelA)
PF05838	Glycosyl hydrolase 108
PF05840	Bacteriophage replication gene A protein (GPA)
PF05857	TraX protein
PF05860	haemagglutination activity domain
PF05866	Endodeoxyribonuclease RusA
PF05869	DNA N-6-adenine-methyltransferase (Dam)
PF05872	Bacterial protein of unknown function (DUF853)
PF05876	Phage terminase large subunit (GpA)
PF05894	Podovirus DNA encapsidation protein (Gp16)
PF05895	Siphovirus protein of unknown function (DUF859)
PF05901	Excalibur calcium-binding domain
PF05908	Poly-gamma-glutamate hydrolase
PF05926	Phage head completion protein (GPL)
PF05929	Phage capsid scaffolding protein (GPO) serine peptidase
PF05930	Prophage CP4-57 regulatory protein (AlpA)
PF05939	Phage minor tail protein
PF05944	Phage small terminase subunit
PF05954	Phage late control gene D protein (GPD)
PF05958	tRNA (Uracil-5-)-methyltransferase
PF05962	HutD
PF05970	PIF1-like helicase
PF05973	Phage derived protein Gp49-like (DUF891)
PF05977	Transmembrane secretion effector
PF05981	CreA protein
PF05987	Bacterial protein of unknown function (DUF898)
PF05996	Ferredoxin-dependent bilin reductase
PF06006	Bacterial protein of unknown function (DUF905)
PF06013	Proteins of 100 residues with WXG
PF06019	Phage GP30.8 protein
PF06026	Ribose 5-phosphate isomerase A (phosphoriboisomerase A)
PF06035	Bacterial transglutaminase-like cysteine proteinase BTLCP
PF06048	Domain of unknown function (DUF927)
PF06056	Putative ATPase subunit of terminase (gpP-like)
PF06064	Host-nuclease inhibitor protein Gam
PF06067	Domain of unknown function (DUF932)
PF06069	PerC transcriptional activator
PF06073	Bacterial protein of unknown function (DUF934)
PF06074	Protein of unknown function (DUF935)
PF06094	Gamma-glutamyl cyclotransferase, AIG2-like

Table 3.4. Continued.

PFAM	Description
PF06107	Bacterial protein of unknown function (DUF951)
PF06114	IrrE N-terminal-like domain
PF06116	Transcriptional activator RinB
PF06120	Tail length tape measure protein
PF06125	Bacterial protein of unknown function (DUF961)
PF06133	Control of competence regulator ComK, YlbF/YmcA
PF06141	Phage minor tail protein U
PF06147	Protein of unknown function (DUF968)
PF06152	Phage minor capsid protein 2
PF06156	Initiation control protein YabA
PF06163	Bacterial protein of unknown function (DUF977)
PF06167	Glucose-regulated metallo-peptidase M90
PF06174	Protein of unknown function (DUF987)
PF06176	Lipopolysaccharide core biosynthesis protein (WaaY)
PF06177	QueT transporter
PF06183	DinI-like family
PF06194	Phage Conserved Open Reading Frame 51
PF06199	Phage tail tube protein
PF06206	CpeT/CpcT family (DUF1001)
PF06210	Protein of unknown function (DUF1003)
PF06222	Phage tail assembly chaperone
PF06223	Minor tail protein T
PF06250	YhcG PDDEXK nuclease domain
PF06252	Protein of unknown function (DUF1018)
PF06254	Putative bacterial toxin ydaT
PF06260	Protein of unknown function (DUF1024)
PF06265	Protein of unknown function (DUF1027)
PF06267	Family of unknown function (DUF1028)
PF06271	RDD family
PF06275	Protein of unknown function (DUF1031)
PF06281	Protein of unknown function (DUF1035)
PF06291	Bor protein
PF06293	Lipopolysaccharide kinase (Kdo/WaaP) family
PF06301	Bacteriophage lambda Kil protein
PF06305	Lipopolysaccharide assembly protein A domain
PF06316	Enterobacterial Ail/Lom protein
PF06319	DNA repair protein MmcB-like
PF06322	Phage NinH protein
PF06323	Phage antitermination protein Q
PF06338	ComK protein
PF06339	Ectoine synthase

Table 3.4. Continued.

PFAM	Description
PF06341	Protein of unknown function (DUF1056)
PF06356	Protein of unknown function (DUF1064)
PF06378	Protein of unknown function (DUF1071)
PF06381	Protein of unknown function (DUF1073)
PF06406	StbA protein
PF06414	Zeta toxin
PF06416	Effector protein NleG
PF06426	Serine acetyltransferase, N-terminal
PF06440	DNA polymerase III, theta subunit
PF06442	R67 dihydrofolate reductase
PF06445	Gyrl-like small molecule binding domain
PF06453	Type II heat-labile enterotoxin , B subunit (LT-IIB)
PF06486	Protein of unknown function (DUF1093)
PF06488	Phage tail tube protein
PF06508	Queuosine biosynthesis protein QueC
PF06528	Phage P2 GpE
PF06530	Phage antitermination protein Q
PF06531	Protein of unknown function (DUF1108)
PF06543	Lactococcus bacteriophage repressor
PF06559	2'-deoxycytidine 5'-triphosphate deaminase (DCD)
PF06560	Glucose-6-phosphate isomerase (GPI)
PF06563	Protein of unknown function (DUF1125)
PF06568	Domain of unknown function (DUF1127)
PF06576	Protein of unknown function (DUF1133)
PF06591	T4-like phage nuclear disruption protein (Ndd)
PF06600	Protein of unknown function (DUF1140)
PF06605	Prophage endopeptidase tail
PF06616	BsuBI/PstI restriction endonuclease domain
PF06630	Enterobacterial exodeoxyribonuclease VIII
PF06634	Protein of unknown function (DUF1156)
PF06649	Protein of unknown function (DUF1161)
PF06662	D-glucuronyl C5-epimerase C-terminus
PF06673	Lactococcus lactis bacteriophage major capsid protein
PF06688	Protein of unknown function (DUF1187)
PF06689	ClpX C4-type zinc finger
PF06693	Protein of unknown function (DUF1190)
PF06711	Protein of unknown function (DUF1198)
PF06714	Gp5 N-terminal OB domain
PF06719	AraC-type transcriptional regulator N-terminus
PF06720	Bacteriophage phi-29 early protein GP16.7
PF06725	3D domain

Table 3.4. Continued.

PFAM	Description
PF06726	Bladder cancer-related protein BC10
PF06736	Protein of unknown function (DUF1211)
PF06737	Transglycosylase-like domain
PF06739	Beta-propeller repeat
PF06745	KaiC
PF06763	Prophage minor tail protein Z (GPZ)
PF06788	Uncharacterised protein family (UPF0257)
PF06791	Prophage tail length tape measure protein
PF06794	Uncharacterised protein family (UPF0270)
PF06803	Protein of unknown function (DUF1232)
PF06805	Bacteriophage lambda tail assembly protein I
PF06806	Putative excisionase (DUF1233)
PF06810	Phage minor structural protein GP20
PF06820	Putative prophage tail fibre C-terminus
PF06824	Metal-independent alpha-mannosidase (GH125)
PF06826	Predicted Permease Membrane Region
PF06831	Formamidopyrimidine-DNA glycosylase H2TH domain
PF06841	T4-like virus tail tube protein gp19
PF06854	Bacteriophage Gp15 protein
PF06866	Protein of unknown function (DUF1256)
PF06868	Protein of unknown function (DUF1257)
PF06890	Bacteriophage Mu Gp45 spike protein
PF06891	P2 phage tail completion protein R (GpR)
PF06892	Phage regulatory protein CII (CP76)
PF06894	Bacteriophage lambda tail assembly chaperone, TAC, protein G
PF06896	Phage tail assembly chaperone proteins, TAC
PF06900	Protein of unknown function (DUF1270)
PF06908	YspA SLOG family
PF06919	Phage Gp30.7 protein
PF06926	Putative replisome organiser protein C-terminus
PF06932	Protein of unknown function (DUF1283)
PF06939	Protein of unknown function (DUF1286)
PF06941	5' nucleotidase, deoxy (Pyrimidine), cytosolic type C protein (NT5C)
PF06945	Protein of unknown function (DUF1289)
PF06946	Bacteriophage A118-like holin, Hol118
PF06950	Protein of unknown function (DUF1293)
PF06962	Putative rRNA methylase
PF06965	Na ⁺ /H ⁺ antiporter 1
PF06967	Mo-dependent nitrogenase C-terminus
PF06970	Replication initiator protein A (RepA) N-terminus
PF06992	Replication protein P

Table 3.4. Continued.

PFAM	Description
PF06995	Phage P2 GpU
PF07005	Sugar-binding N-terminal domain
PF07015	VirC1 protein
PF07022	Bacteriophage CI repressor helix-turn-helix domain
PF07026	Protein of unknown function (DUF1317)
PF07027	Protein of unknown function (DUF1318)
PF07030	Protein of unknown function (DUF1320)
PF07041	Protein of unknown function (DUF1327)
PF07043	Protein of unknown function (DUF1328)
PF07066	Lactococcus phage M3 protein
PF07067	Protein of unknown function (DUF1340)
PF07068	Major capsid protein Gp23
PF07070	SpoOM protein
PF07083	Protein of unknown function (DUF1351)
PF07087	Protein of unknown function (DUF1353)
PF07097	Protein of unknown function (DUF1359)
PF07102	Protein of unknown function (DUF1364)
PF07104	Protein of unknown function (DUF1366)
PF07105	Protein of unknown function (DUF1367)
PF07108	PipA protein
PF07116	Protein of unknown function (DUF1372)
PF07118	Protein of unknown function (DUF1374)
PF07120	Protein of unknown function (DUF1376)
PF07125	Protein of unknown function (DUF1378)
PF07129	Protein of unknown function (DUF1381)
PF07131	Protein of unknown function (DUF1382)
PF07136	Protein of unknown function (DUF1385)
PF07141	Putative bacteriophage terminase small subunit
PF07151	Protein of unknown function (DUF1391)
PF07157	DNA circularisation protein N-terminus
PF07166	Protein of unknown function (DUF1398)
PF07191	zinc-ribbons
PF07193	Protein of unknown function (DUF1408)
PF07212	Hyaluronidase protein (HylP)
PF07230	Bacteriophage T4-like capsid assembly protein (Gp20)
PF07261	Replication initiation and membrane attachment
PF07274	Protein of unknown function (DUF1440)
PF07275	Antirestriction protein (ArdA)
PF07278	Protein of unknown function (DUF1441)
PF07282	Putative transposase DNA-binding domain
PF07285	Protein of unknown function (DUF1444)

Table 3.4. Continued.

PFAM	Description
PF07319	Primosomal protein DnaI N-terminus
PF07338	Protein of unknown function (DUF1471)
PF07342	Protein of unknown function (DUF1474)
PF07352	Bacteriophage Mu Gam like protein
PF07358	Protein of unknown function (DUF1482)
PF07374	Protein of unknown function (DUF1492)
PF07377	Protein of unknown function (DUF1493)
PF07384	Protein of unknown function (DUF1497)
PF07409	Phage protein GP46
PF07410	Streptococcus thermophilus bacteriophage Gp111 protein
PF07411	Domain of unknown function (DUF1508)
PF07438	Protein of unknown function (DUF1514)
PF07439	Protein of unknown function (DUF1515)
PF07453	NUMOD1 domain
PF07455	Phage polarity suppression protein (Psu)
PF07459	CTX phage RstB protein
PF07460	NUMOD3 motif
PF07463	NUMOD4 motif
PF07471	Phage DNA packaging protein Nu1
PF07479	NAD-dependent glycerol-3-phosphate dehydrogenase C-terminus
PF07484	Phage Tail Collar Domain
PF07486	Cell Wall Hydrolase
PF07494	Two component regulator propeller
PF07498	Rho termination factor, N-terminal domain
PF07501	G5 domain
PF07505	Protein of unknown function (DUF5131)
PF07506	RepB plasmid partitioning protein
PF07508	Recombinase
PF07523	Bacterial Ig-like domain (group 3)
PF07553	Host cell surface-exposed lipoprotein
PF07564	Domain of Unknown Function (DUF1542)
PF07568	Histidine kinase
PF07591	Pretoxin HINT domain
PF07659	Nucleotide modification associated domain 1
PF07661	MORN repeat variant
PF07669	Eco57I restriction-modification methylase
PF07681	DoxX
PF07690	Major Facilitator Superfamily
PF07691	PA14 domain
PF07693	KAP family P-loop domain
PF07694	5TMR of 5TMR-LYT

Table 3.4. Continued.

PFAM	Description
PF07702	UTRA domain
PF07722	Peptidase C26
PF07724	AAA domain (Cdc48 subfamily)
PF07728	AAA domain (dynein-related subfamily)
PF07729	FCD domain
PF07730	Histidine kinase
PF07732	Multicopper oxidase
PF07733	Bacterial DNA polymerase III alpha NTPase domain
PF07739	TipAS antibiotic-recognition domain
PF07750	GcrA cell cycle regulator
PF07751	Abi-like protein
PF07755	Domain of unknown function (DUF1611_C) P-loop domain
PF07761	Protein of unknown function (DUF1617)
PF07764	Omega Transcriptional Repressor
PF07768	PVL ORF-50-like family
PF07789	Protein of unknown function (DUF1627)
PF07799	Protein of unknown function (DUF1643)
PF07825	Excisionase-like protein
PF07852	Protein of unknown function (DUF1642)
PF07859	alpha/beta hydrolase fold
PF07866	Protein of unknown function (DUF1653)
PF07867	Protein of unknown function (DUF1654)
PF07868	Protein of unknown function (DUF1655)
PF07874	Prophage protein (DUF1660)
PF07880	Bacteriophage T4 gp9/10-like protein
PF07883	Cupin domain
PF07885	Ion channel
PF07889	Protein of unknown function (DUF1664)
PF07901	Protein of unknown function (DUF1672)
PF07902	gp58-like protein
PF07921	Fibritin C-terminal region
PF07927	HicA toxin of bacterial toxin-antitoxin,
PF07929	Plasmid pRiA4b ORF-3-like protein
PF07935	ORF D-335-like protein
PF07952	Clostridium neurotoxin, Translocation domain
PF07968	Leukocidin/Hemolysin toxin family
PF07972	NrdI Flavodoxin like
PF07992	Pyridine nucleotide-disulphide oxidoreductase
PF07997	Protein of unknown function (DUF1694)
PF08000	Bacterial PH domain
PF08007	Cupin superfamily protein

Table 3.4. Continued.

PFAM	Description
PF08010	Bacteriophage protein GP30.3
PF08049	IlvB leader peptide
PF08125	Mannitol dehydrogenase C-terminal domain
PF08133	Anticodon nuclease activator family
PF08134	cIII protein family
PF08200	Bacteriophage 1.1 Protein
PF08212	Lipocalin-like domain
PF08230	CW_7 repeat
PF08237	PE-PPE domain
PF08239	Bacterial SH3 domain
PF08240	Alcohol dehydrogenase GroES-like domain
PF08241	Methyltransferase domain
PF08273	Zinc-binding domain of primase-helicase
PF08275	DNA primase catalytic core, N-terminal domain
PF08279	HTH domain
PF08281	Sigma-70, region 4
PF08282	haloacid dehalogenase-like hydrolase
PF08291	Peptidase M15
PF08298	PrkA AAA domain
PF08299	Bacterial dnaA protein helix-turn-helix
PF08310	LGFP repeat
PF08317	Spc7 kinetochore protein
PF08343	Ribonucleotide reductase N-terminal
PF08346	AntA/AntB antirepressor
PF08357	SEFIR domain
PF08378	Nuclease-related domain
PF08400	Prophage tail fibre N-terminal
PF08410	Domain of unknown function (DUF1737)
PF08423	Rad51
PF08443	RimK-like ATP-grasp domain
PF08447	PAS fold
PF08448	PAS fold
PF08450	SMP-30/Gluconolactonase/LRE-like region
PF08460	Bacterial SH3 domain
PF08534	Redoxin
PF08542	Replication factor C C-terminal domain
PF08645	Polynucleotide kinase 3 phosphatase
PF08668	HDOD domain
PF08677	GP11 baseplate wedge protein
PF08684	DNA mimic ocr
PF08706	D5 N terminal like

Table 3.4. Continued.

PFAM	Description
PF08707	Primase C terminal 2 (PriCT-2)
PF08708	Primase C terminal 1 (PriCT-1)
PF08719	NADAR domain
PF08740	BCS1 N terminal
PF08748	Phage tail assembly chaperone
PF08751	TrwC relaxase
PF08757	CotH kinase protein
PF08759	Glycosyltransferase GT-D fold
PF08761	dUTPase
PF08765	Mor transcription activator family
PF08769	Sporulation initiation factor Spo0A C terminal
PF08774	VRR-NUC domain
PF08775	ParB family
PF08787	Alginate lyase
PF08797	HIRAN domain
PF08798	CRISPR associated protein
PF08800	VirE N-terminal domain
PF08803	Putative mono-oxygenase ydhR
PF08804	gp32 DNA binding protein like
PF08809	Phage related hypothetical protein (DUF1799)
PF08812	YtxC-like family
PF08813	Phage tail tube protein, TTP
PF08816	Inhibitor of vertebrate lysozyme (Ivy)
PF08822	Protein of unknown function (DUF1804)
PF08843	Nucleotidyl transferase AbiEii toxin, Type IV TA system
PF08845	Toxin SymE, type I toxin-antitoxin system
PF08855	Domain of unknown function (DUF1825)
PF08861	Domain of unknown function DUF1828
PF08862	Domain of unknown function DUF1829
PF08863	YolD-like protein
PF08864	UPF0302 domain
PF08867	FRG domain
PF08873	Domain of unknown function (DUF1834)
PF08875	Domain of unknown function (DUF1833)
PF08890	Phage XkdN-like tail assembly chaperone protein, TAC
PF08902	Domain of unknown function (DUF1848)
PF08924	Domain of unknown function (DUF1906)
PF08931	Receptor-binding protein of phage tail base-plate Siphoviridae, head
PF08937	MTH538 TIR-like domain (DUF1863)
PF08948	Domain of unknown function (DUF1859)
PF08949	Domain of unknown function (DUF1860)

Table 3.4. Continued.

PFAM	Description
PF08951	Enterocin A Immunity
PF08960	Domain of unknown function (DUF1874)
PF08965	Domain of unknown function (DUF1870)
PF08972	Domain of unknown function (DUF1902)
PF08993	T4 gene Gp59 loader of gp41 DNA helicase
PF08994	T4 gene Gp59 loader of gp41 DNA helicase C-term
PF08998	Bacterial epsilon antitoxin
PF09003	Bacteriophage lambda integrase, Arm DNA-binding domain
PF09008	Head binding
PF09010	Anti-Sigma Factor A
PF09018	P3 major capsid protein
PF09019	EcoRII C terminal
PF09035	Excisionase from transposon Tn916
PF09048	Cro
PF09063	Phage PP7 coat protein
PF09077	Mu B transposition protein, C terminal
PF09089	Phage short tail fibre protein gp12, middle domain
PF09096	Baseplate structural protein, domain 2
PF09097	Baseplate structural protein, domain 1
PF09114	Transcription factor MotA, activation domain
PF09116	gp45 sliding clamp, C terminal
PF09124	T4 recombination endonuclease VII, dimerisation
PF09140	ATPase MipZ
PF09152	Domain of unknown function (DUF1937)
PF09158	Bacteriophage T4 MotA, C-terminal
PF09195	Restriction endonuclease BglII
PF09198	Bacteriophage T4 beta-glucosyltransferase
PF09214	Bacteriophage Prd1, adsorption protein P2
PF09215	Bacteriophage T4, Gp8
PF09225	Restriction endonuclease PvuII
PF09233	Restriction endonuclease EcoRV
PF09250	Bifunctional DNA primase/polymerase, N-terminal
PF09251	Salmonella phage P22 tail-spike
PF09293	T4 RNase H, C terminal
PF09299	Mu transposase, C-terminal
PF09300	Tectiviridae, minor capsid
PF09301	Domain of unknown function (DUF1970)
PF09306	Bacteriophage, scaffolding protein
PF09327	Domain of unknown function (DUF1983)
PF09335	SNARE associated Golgi protein
PF09343	Conserved hypothetical protein 2217 (DUF2460)

Table 3.4. Continued.

PFAM	Description
PF09355	Phage protein Gp19/Gp15/Gp42
PF09356	Phage conserved hypothetical protein BR0599
PF09373	Pseudomurein-binding repeat
PF09374	Predicted Peptidoglycan domain
PF09388	Spo0E like sporulation regulatory protein
PF09393	Phage tail tube protein
PF09397	Ftsk gamma domain
PF09411	Lipid A 3-O-deacylase (PagL)
PF09414	RNA ligase
PF09423	PhoD-like phosphatase
PF09424	Yqey-like protein
PF09467	Hypothetical protein Yopt
PF09479	Listeria-Bacteroides repeat domain (List_Bact_rpt)
PF09491	AlwI restriction endonuclease
PF09509	Protein of unknown function (Hypoth_ymh)
PF09511	RNA ligase
PF09524	Conserved phage C-terminus (Phg_2220_C)
PF09563	LlaJI restriction endonuclease
PF09568	MjaI restriction endonuclease
PF09587	Bacterial capsule synthesis protein PGA_cap
PF09588	YqaJ-like viral recombinase domain
PF09612	Bacterial protein of unknown function (HtrL_YibB)
PF09614	CRISPR-associated protein (Cas_Csy2)
PF09615	CRISPR-associated protein (Cas_Csy3)
PF09618	CRISPR-associated protein (Cas_Csy4)
PF09629	YorP protein
PF09636	XkdW protein
PF09639	YjcQ protein
PF09642	YonK protein
PF09643	YopX protein
PF09645	F-112 protein
PF09646	Gp37 protein
PF09652	Putative CRISPR-associated protein (Cas_VVA1548)
PF09669	Phage regulatory protein Rha (Phage_pRha)
PF09675	Chlamydia-phage Chp2 scaffold (Chlamy_scaf)
PF09681	N-terminal phage replisome organiser (Phage_rep_org_N)
PF09682	Bacteriophage holin of superfamily 6 (Holin_LLH)
PF09684	Phage tail protein (Tail_P2_I)
PF09693	Phage uncharacterised protein (Phage_XkdX)
PF09694	Bacterial protein of unknown function (Gcw_chp)
PF09718	Lambda phage tail tape-measure protein (Tape_meas_lam_C)

Table 3.4. Continued.

PFAM	Description
PF09723	Zinc ribbon domain
PF09825	Biotin-protein ligase, N terminal
PF09828	Chromate resistance exported protein
PF09834	Predicted membrane protein (DUF2061)
PF09848	Uncharacterized conserved protein (DUF2075)
PF09852	Predicted membrane protein (DUF2079)
PF09904	Winged helix-turn helix
PF09907	HigB_toxin, RelE-like toxic component of a toxin-antitoxin system
PF09926	Uncharacterized small protein (DUF2158)
PF09931	Uncharacterized conserved protein (DUF2163)
PF09950	Uncharacterized protein conserved in bacteria (DUF2184)
PF09954	Uncharacterized protein conserved in bacteria (DUF2188)
PF09956	Uncharacterized conserved protein (DUF2190)
PF09967	VWA-like domain (DUF2201)
PF09976	Tetratricopeptide repeat-like domain
PF09979	Uncharacterized protein conserved in bacteria (DUF2213)
PF09989	CoA enzyme activase uncharacterised domain (DUF2229)
PF09992	Phosphodiester glycosidase
PF10004	Uncharacterized protein conserved in bacteria (DUF2247)
PF10021	Uncharacterized protein conserved in bacteria (DUF2263)
PF10022	Uncharacterized protein conserved in bacteria (DUF2264)
PF10045	Uncharacterized conserved protein (DUF2280)
PF10065	Uncharacterized conserved protein (DUF2303)
PF10073	Uncharacterized protein conserved in bacteria (DUF2312)
PF10076	Uncharacterised protein conserved in bacteria (DUF2313)
PF10083	Uncharacterized protein conserved in bacteria (DUF2321)
PF10088	Uncharacterised protein conserved in bacteria (DUF2326)
PF10097	Predicted membrane protein (DUF2335)
PF10108	Predicted 3'-5' exonuclease related to the exonuclease domain of PolB
PF10109	Phage tail assembly chaperone proteins, E, or 41 or 14
PF10117	McrBC 5-methylcytosine restriction system component
PF10122	Mu-like prophage protein Com
PF10123	Mu-like prophage I protein
PF10124	Mu-like prophage major head subunit gpT
PF10127	RNA repair pathway DNA polymerase beta family
PF10130	PIN domain
PF10134	Replication initiator protein A
PF10138	vWA found in TerF C terminus
PF10145	Phage-related minor tail protein
PF10263	SprT-like family
PF10294	Lysine methyltransferase

Table 3.4. Continued.

PFAM	Description
PF10387	Protein of unknown function (DUF2442)
PF10389	*Merged into existing family
PF10465	PinA peptidase inhibitor
PF10507	Transmembrane protein 65
PF10543	ORF6N domain
PF10544	T5orf172 domain
PF10546	P63C domain
PF10547	P22_AR N-terminal domain
PF10548	P22AR C-terminal domain
PF10549	ORF11CD3 domain
PF10552	ORF6C domain
PF10554	Ash protein family
PF10592	AIPR protein
PF10614	Type VIII secretion system (T8SS), CsgF protein
PF10618	Phage tail tube protein
PF10651	BppU N-terminal domain
PF10653	Protein gp45 of Bacteriophage A118
PF10654	Protein of unknown function (DUF2481)
PF10655	Hypothetical protein of unknown function (DUF2482)
PF10656	Hypothetical protein of unknown function (DUF2483)
PF10665	Minor capsid protein
PF10666	Phage tail assembly chaperone protein Gp14 (A118)
PF10668	Phage terminase small subunit
PF10669	Protein gp23 (Bacteriophage A118)
PF10672	S-adenosylmethionine-dependent methyltransferase
PF10676	Spore germination protein gerPA/gerPF
PF10686	YspA, cpYpsA-related SLOG family
PF10688	Bacterial inner membrane protein
PF10708	Protein of unknown function (DUF2510)
PF10709	Protein of unknown function (DUF2511)
PF10711	Hypothetical protein (DUF2513)
PF10715	T4-page Endoribonuclease RegB
PF10721	Protein of unknown function (DUF2514)
PF10734	Protein of unknown function (DUF2523)
PF10743	Regulatory phage protein cox
PF10746	Phage holin T7 family, holin superfamily II
PF10758	Protein of unknown function (DUF2586)
PF10761	Protein of unknown function (DUF2590)
PF10765	Protein of unknown function (DUF2591)
PF10772	Protein of unknown function (DUF2597)
PF10779	Haemolysin XhIA

Table 3.4. Continued.

PFAM	Description
PF10784	Plasmid stability protein
PF10789	Phage RNA polymerase binding, RpbA
PF10796	Sigma-S stabilisation anti-adaptor protein
PF10800	Protein of unknown function (DUF2528)
PF10805	Protein of unknown function (DUF2730)
PF10809	Protein of unknown function (DUF2732)
PF10828	Protein of unknown function (DUF2570)
PF10834	Protein of unknown function (DUF2560)
PF10841	Protein of unknown function (DUF2644)
PF10844	Protein of unknown function (DUF2577)
PF10849	Protein of unknown function (DUF2654)
PF10854	Protein of unknown function (DUF2649)
PF10872	Protein of unknown function (DUF2740)
PF10874	Protein of unknown function (DUF2746)
PF10876	Phage tail assembly chaperone protein, TAC
PF10883	Protein of unknown function (DUF2681)
PF10886	Protein of unknown function (DUF2685)
PF10893	Protein of unknown function (DUF2724)
PF10902	WYL_2, Sm-like SH3 beta-barrel fold
PF10908	Protein of unknown function (DUF2778)
PF10910	Protein of unknown function (DUF2744)
PF10911	Protein of unknown function (DUF2717)
PF10922	Protein of unknown function (DUF2745)
PF10926	Protein of unknown function (DUF2800)
PF10928	Protein of unknown function (DUF2810)
PF10930	Protein of unknown function (DUF2737)
PF10934	Protein of unknown function (DUF2634)
PF10935	Protein of unknown function (DUF2637)
PF10948	Protein of unknown function (DUF2635)
PF10960	BhIA holin family
PF10963	Phage tail assembly chaperone
PF10965	Protein of unknown function (DUF2767)
PF10979	Protein of unknown function (DUF2786)
PF10983	Protein of unknown function (DUF2793)
PF10987	Protein of unknown function (DUF2806)
PF10991	Protein of unknown function (DUF2815)
PF11031	Bacteriophage T holin
PF11039	Protein of unknown function (DUF2824)
PF11041	Protein of unknown function (DUF2612)
PF11043	Protein of unknown function (DUF2856)
PF11053	Terminase DNA packaging enzyme

Table 3.4. Continued.

PFAM	Description
PF11056	Recombination, repair and ssDNA binding protein UvsY
PF11058	Antirestriction protein Ral
PF11087	PRD1 phage membrane DNA delivery
PF11090	Protein of unknown function (DUF2833)
PF11091	Tail-tube assembly protein
PF11110	Baseplate hub distal subunit
PF11112	Pyocin activator protein PrtN
PF11113	Head assembly gene product
PF11114	Minor capsid protein
PF11121	Protein of unknown function (DUF2639)
PF11123	DNA packaging protein
PF11126	Transcriptional regulator DsbA
PF11133	Head fiber protein
PF11134	Phage stabilisation protein
PF11149	Protein of unknown function (DUF2924)
PF11166	Protein of unknown function (DUF2951)
PF11171	Protein of unknown function (DUF2958)
PF11185	Protein of unknown function (DUF2971)
PF11192	Protein of unknown function (DUF2977)
PF11195	Protein of unknown function (DUF2829)
PF11202	Phosphoribosyl transferase (PRTase)
PF11208	Protein of unknown function (DUF2992)
PF11211	Protein of unknown function (DUF2997)
PF11213	Protein of unknown function (DUF3006)
PF11225	Protein of unknown function (DUF3024)
PF11238	Protein of unknown function (DUF3039)
PF11242	Protein of unknown function (DUF2774)
PF11243	Protein of unknown function (DUF3045)
PF11246	Base plate wedge protein 53
PF11247	Protein of unknown function (DUF2675)
PF11251	Protein of unknown function (DUF3050)
PF11284	Protein of unknown function (DUF3085)
PF11300	Protein of unknown function (DUF3102)
PF11310	Protein of unknown function (DUF3113)
PF11319	Type VI secretion system Vasil, EvfG, VC_A0118
PF11325	Domain of unknown function (DUF3127)
PF11351	Holin of 3TMs, for gene-transfer release
PF11360	Protein of unknown function (DUF3110)
PF11363	Protein of unknown function (DUF3164)
PF11367	Protein of unknown function (DUF3168)
PF11372	Domain of unknown function (DUF3173)

Table 3.4. Continued.

PFAM	Description
PF11373	Protein of unknown function (DUF3175)
PF11397	Glycosyltransferase (GlcNAc)
PF11417	Loader and inhibitor of phage G40P
PF11418	Phi29 scaffolding protein
PF11423	Regulatory protein Mnt
PF11434	Chemotaxis-inhibiting protein CHIPS
PF11436	Protein of unknown function (DUF3199)
PF11438	36-mer N-terminal peptide of the N protein (N36)
PF11440	DNA alpha-glucosyltransferase
PF11443	Domain of unknown function (DUF2828)
PF11457	Protein of unknown function (DUF3021)
PF11546	Staphylococcal complement inhibitor SCIN
PF11602	ATPase P4 of dsRNA bacteriophage phi-12
PF11611	Domain of unknown function (DUF4352)
PF11623	NAD(P)H dehydrogenase subunit S
PF11634	Nuclease inhibitor from bacteriophage T4
PF11637	ATP-dependant DNA helicase UvsW
PF11645	PD-(D/E)XK endonuclease
PF11646	Protein of unknown function DUF3258
PF11649	Virus neck protein
PF11650	P22 tail accessory factor
PF11651	P22 coat protein - gene protein 5
PF11653	Bacteriophage T7 virion assembly protein
PF11662	Protein of unknown function (DUF3263)
PF11672	zinc-finger-containing domain
PF11673	Protein of unknown function (DUF3269)
PF11681	Protein of unknown function (DUF3277)
PF11726	Inovirus Gp2
PF11739	Dicarboxylate transport
PF11742	Protein of unknown function (DUF3302)
PF11750	Protein of unknown function (DUF3307)
PF11752	Protein of unknown function (DUF3309)
PF11753	Protein of unknown function (DUF3310)
PF11774	Lsr2
PF11775	Cobalamin biosynthesis protein CobT VWA domain
PF11776	Nickel/cobalt transporter regulator
PF11809	Domain of unknown function (DUF3330)
PF11836	Phage tail tube protein, GTA-gp10
PF11860	N-acetylmuramidase
PF11863	Protein of unknown function (DUF3383)
PF11876	Protein of unknown function (DUF3396)

Table 3.4. Continued.

PFAM	Description
PF11922	Domain of unknown function (DUF3440)
PF11948	Protein of unknown function (DUF3465)
PF11950	Protein of unknown function (DUF3467)
PF11962	Peptidase_G2, IMC autoproteolytic cleavage domain
PF11974	Bacterial alpha-2-macroglobulin MG3 domain
PF11981	Domain of unknown function (DUF3482)
PF11985	Protein of unknown function (DUF3486)
PF11994	Protein of unknown function (DUF3489)
PF12002	MgsA AAA+ ATPase C terminal
PF12025	Phage protein C
PF12033	Protein of unknown function (DUF3519)
PF12083	Domain of unknown function (DUF3560)
PF12116	Stage III sporulation protein D
PF12123	N-acetylmuramoyl-L-alanine amidase
PF12138	Spherulation-specific family 4
PF12146	Serine aminopeptidase, S33
PF12167	Arm DNA-binding domain
PF12187	Viral/Archaeal nuclease
PF12193	Sulfolobus virus coat protein C terminal
PF12200	Domain of unknown function (DUF3597)
PF12206	Domain of unknown function (DUF3599)
PF12217	Catalytic beta propeller domain of bacteriophage endosialidase
PF12224	Putative amidoligase enzyme
PF12227	Protein of unknown function (DUF3603)
PF12236	Bacteriophage head to tail connecting protein
PF12283	Bacteriophage protein K
PF12294	Protein of unknown function (DUF3626)
PF12299	Protein of unknown function (DUF3627)
PF12322	T4 bacteriophage base plate protein
PF12323	Helix-turn-helix domain
PF12363	Phage tail assembly chaperone protein, TAC
PF12375	Phage protein
PF12386	Pseudomurein endo-isopeptidase Pei
PF12421	Fibronectin type III protein
PF12461	Protein of unknown function (DUF3688)
PF12472	Phage related protein
PF12566	Protein of unknown function (DUF3748)
PF12571	Phage tail-collar fibre protein
PF12587	Protein of unknown function (DUF3761)
PF12599	Protein of unknown function (DUF3768)
PF12604	Tail fibre protein gp37 C terminal

Table 3.4. Continued.

PFAM	Description
PF12631	MnmE helical domain
PF12635	Protein of unknown function (DUF3780)
PF12642	Conjugative transposon protein TcpC
PF12645	Helix-turn-helix domain
PF12648	TcpE family
PF12651	Ribbon-helix-helix domain
PF12660	Putative zinc-finger of transcription factor IIIC complex
PF12666	PrgI family protein
PF12671	Putative amidase domain
PF12677	Domain of unknown function (DUF3797)
PF12684	PDDEXK-like domain of unknown function (DUF3799)
PF12686	Protein of unknown function (DUF3800)
PF12691	Bacteriophage minor capsid protein
PF12694	Circularly permuted YpsA SLOG family
PF12697	Alpha/beta hydrolase family
PF12699	phiKZ-like phage internal head proteins
PF12705	PD-(D/E)XK nuclease superfamily
PF12706	Beta-lactamase superfamily domain
PF12708	Pectate lyase superfamily protein
PF12710	haloacid dehalogenase-like hydrolase
PF12727	PBP superfamily domain
PF12728	Helix-turn-helix domain
PF12750	Maff2 family
PF12759	InsA C-terminal domain
PF12784	PD-(D/E)XK nuclease family transposase
PF12789	Phage tail repeat like
PF12796	Ankyrin repeats (3 copies)
PF12804	MobA-like NTP transferase domain
PF12810	Glycine rich protein
PF12826	Helix-hairpin-helix motif
PF12833	Helix-turn-helix domain
PF12838	4Fe-4S dicluster domain
PF12844	Helix-turn-helix domain
PF12846	AAA-like domain
PF12849	PBP superfamily domain
PF12850	Calcineurin-like phosphoesterase superfamily domain
PF12851	Oxygenase domain of the 2OGFeDO superfamily
PF12869	tRNA_anti-like
PF12873	Domain of unknown function (DUF3825)
PF12917	HD containing hydrolase-like enzyme
PF12949	HeH/LEM domain

Table 3.4. Continued.

PFAM	Description
PF12950	TaqI-like C-terminal specificity domain
PF12957	Domain of unknown function (DUF3846)
PF12961	Domain of unknown function (DUF3850)
PF12978	Domain of Unknown Function with PDB structure (DUF3862)
PF13020	Domain of unknown function (DUF3883)
PF13022	Helix-turn-helix of insertion element transposase
PF13031	Protein of unknown function (DUF3892)
PF13064	Protein of unknown function (DUF3927)
PF13073	Protein of unknown function (DUF3937)
PF13076	Fur-regulated basic protein A
PF13078	Protein of unknown function (DUF3942)
PF13086	AAA domain
PF13088	BNR repeat-like domain
PF13102	Phage integrase SAM-like domain
PF13106	Domain of unknown function (DUF3961)
PF13108	Protein of unknown function (DUF3969)
PF13111	pPIWI_RE module N-terminal domain
PF13128	Protein of unknown function (DUF3954)
PF13132	Domain of unknown function (DUF3950)
PF13137	Protein of unknown function (DUF3983)
PF13148	Protein of unknown function (DUF3987)
PF13151	Protein of unknown function (DUF3990)
PF13152	Protein of unknown function (DUF3967)
PF13155	Toprim-like
PF13166	AAA domain
PF13173	AAA domain
PF13175	AAA ATPase domain
PF13177	DNA polymerase III, delta subunit
PF13198	Protein of unknown function (DUF4014)
PF13199	Glycosyl hydrolase family 66
PF13203	Putative metallopeptidase domain
PF13205	Bacterial Ig-like domain
PF13223	Protein of unknown function (DUF4031)
PF13227	Protein of unknown function (DUF4035)
PF13229	Right handed beta helix region
PF13238	AAA domain
PF13245	AAA domain
PF13247	4Fe-4S dicluster domain
PF13250	Domain of unknown function (DUF4041)
PF13252	Protein of unknown function (DUF4043)
PF13262	Protein of unknown function (DUF4054)

Table 3.4. Continued.

PFAM	Description
PF13264	Domain of unknown function (DUF4055)
PF13269	Protein of unknown function (DUF4060)
PF13272	Putative 2/3 transmembrane domain holin
PF13273	Protein of unknown function (DUF4064)
PF13274	Protein of unknown function (DUF4065)
PF13275	S4 domain
PF13276	HTH-like domain
PF13280	WYL domain
PF13289	SIR2-like domain
PF13290	Chitinase/beta-hexosaminidase C-terminal domain
PF13302	Acetyltransferase (GNAT) domain
PF13304	AAA domain, putative AbiEii toxin, Type IV TA system
PF13306	BspA type Leucine rich repeat region (6 copies)
PF13307	Helicase C-terminal domain
PF13310	Virulence protein RhuM family
PF13328	HD domain
PF13337	Putative ATP-dependent Lon protease
PF13347	MFS/sugar transport protein
PF13353	4Fe-4S single cluster domain
PF13354	Beta-lactamase enzyme family
PF13356	Arm DNA-binding domain
PF13358	DDE superfamily endonuclease
PF13360	PQQ-like domain
PF13361	UvrD-like helicase C-terminal domain
PF13362	Toprim domain
PF13365	Trypsin-like peptidase domain
PF13366	PD-(D/E)XK nuclease superfamily
PF13370	4Fe-4S single cluster domain of Ferredoxin I
PF13377	Periplasmic binding protein-like domain
PF13378	Enolase C-terminal domain-like
PF13382	Adenine deaminase C-terminal domain
PF13384	Homeodomain-like domain
PF13385	Concanavalin A-like lectin/glucanases superfamily
PF13392	HNH endonuclease
PF13393	Histidyl-tRNA synthetase
PF13394	4Fe-4S single cluster domain
PF13398	Peptidase M50B-like
PF13401	AAA domain
PF13403	Hint domain
PF13407	Periplasmic binding protein domain
PF13411	MerR HTH family regulatory protein

Table 3.4. Continued.

PFAM	Description
PF13412	Winged helix-turn-helix DNA-binding
PF13415	Galactose oxidase, central domain
PF13418	Galactose oxidase, central domain
PF13419	Haloacid dehalogenase-like hydrolase
PF13420	Acetyltransferase (GNAT) domain
PF13426	PAS domain
PF13427	Domain of unknown function (DUF4111)
PF13428	Tetratricopeptide repeat
PF13434	L-lysine 6-monooxygenase (NADPH-requiring)
PF13437	HlyD family secretion protein
PF13443	Cro/C1-type HTH DNA-binding domain
PF13454	FAD-NAD(P)-binding
PF13455	Meiotically up-regulated gene 113
PF13456	Reverse transcriptase-like
PF13464	Domain of unknown function (DUF4115)
PF13469	Sulfotransferase family
PF13472	GDSL-like Lipase/Acylhydrolase family
PF13475	Domain of unknown function (DUF4116)
PF13476	AAA domain
PF13478	XdhC Rossmann domain
PF13479	AAA domain
PF13481	AAA domain
PF13482	RNase_H superfamily
PF13489	Methyltransferase domain
PF13493	Domain of unknown function (DUF4118)
PF13503	Domain of unknown function (DUF4123)
PF13505	Outer membrane protein beta-barrel domain
PF13508	Acetyltransferase (GNAT) domain
PF13511	Domain of unknown function (DUF4124)
PF13518	Helix-turn-helix domain
PF13520	Amino acid permease
PF13521	AAA domain
PF13522	Glutamine amidotransferase domain
PF13524	Glycosyl transferases group 1
PF13529	Peptidase_C39 like family
PF13531	Bacterial extracellular solute-binding protein
PF13532	2OG-Fe(II) oxygenase superfamily
PF13538	UvrD-like helicase C-terminal domain
PF13539	D-alanyl-D-alanine carboxypeptidase
PF13540	Regulator of chromosome condensation (RCC1) repeat
PF13545	Crp-like helix-turn-helix domain

Table 3.4. Continued.

PFAM	Description
PF13547	GTA TIM-barrel-like domain
PF13550	Putative phage tail protein
PF13551	Winged helix-turn helix
PF13552	Protein of unknown function (DUF4127)
PF13554	Bacteriophage related domain of unknown function
PF13560	Helix-turn-helix domain
PF13561	Enoyl-(Acyl carrier protein) reductase
PF13563	2'-5' RNA ligase superfamily
PF13570	PQQ-like domain
PF13573	SprB repeat
PF13578	Methyltransferase domain
PF13580	SIS domain
PF13581	Histidine kinase-like ATPase domain
PF13589	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
PF13592	Winged helix-turn helix
PF13597	Anaerobic ribonucleoside-triphosphate reductase
PF13599	Pentapeptide repeats (9 copies)
PF13604	AAA domain
PF13609	Gram-negative porin
PF13610	DDE domain
PF13612	Transposase DDE domain
PF13614	AAA domain
PF13617	YnbE-like lipoprotein
PF13619	KTSC domain
PF13620	Carboxypeptidase regulatory-like domain
PF13635	Domain of unknown function (DUF4143)
PF13640	2OG-Fe(II) oxygenase superfamily
PF13641	Glycosyltransferase like family 2
PF13643	Domain of unknown function (DUF4145)
PF13649	Methyltransferase domain
PF13650	Aspartyl protease
PF13651	Adenine-specific methyltransferase EcoRI
PF13662	Toprim domain
PF13671	AAA domain
PF13672	Protein phosphatase 2C
PF13673	Acetyltransferase (GNAT) domain
PF13676	TIR domain
PF13683	Integrase core domain
PF13692	Glycosyl transferases group 1
PF13693	Winged helix-turn-helix DNA-binding
PF13701	Transposase DDE domain group 1

Table 3.4. Continued.

PFAM	Description
PF13702	Lysozyme-like
PF13704	Glycosyl transferase family 2
PF13707	RloB-like protein
PF13708	Domain of unknown function (DUF4942)
PF13711	Domain of unknown function (DUF4160)
PF13730	Helix-turn-helix domain
PF13737	Transposase DDE domain
PF13738	Pyridine nucleotide-disulphide oxidoreductase
PF13744	Helix-turn-helix domain
PF13749	Putative ATP-dependent DNA helicase recG C-terminal
PF13751	Transposase DDE domain
PF13759	Putative 2OG-Fe(II) oxygenase
PF13773	Domain of unknown function (DUF4170)
PF13780	Domain of unknown function (DUF4176)
PF13787	Protein of unknown function with HXXEE motif
PF13793	N-terminal domain of ribose phosphate pyrophosphokinase
PF13799	Domain of unknown function (DUF4183)
PF13808	DDE_Tnp_1-associated
PF13810	Domain of unknown function (DUF4185)
PF13814	Replication-relaxation
PF13847	Methyltransferase domain
PF13855	Leucine rich repeat
PF13856	ATP-binding sugar transporter from pro-phage
PF13859	BNR repeat-like domain
PF13876	Phage protein (N4 Gp49/phage Sf6 gene 66) family
PF13884	Chaperone of endosialidase
PF13895	Immunoglobulin domain
PF13927	Immunoglobulin domain
PF13930	DNA/RNA non-specific endonuclease
PF13935	Ead/Ea22-like protein
PF13936	Helix-turn-helix domain
PF13939	Toxin TisB, type I toxin-antitoxin system
PF13946	Domain of unknown function (DUF4214)
PF13957	Toxin YafO, type II toxin-antitoxin system
PF13958	Toxin ToxN, type III toxin-antitoxin system
PF13964	Kelch motif
PF13973	Domain of unknown function (DUF4222)
PF13974	YebO-like protein
PF13986	Domain of unknown function (DUF4224)
PF13992	YecR-like lipoprotein
PF14000	DNA packaging protein FI

Table 3.4. Continued.

PFAM	Description
PF14001	YdfZ protein
PF14011	EspG family
PF14015	Protein of unknown function (DUF4231)
PF14020	Protein of unknown function (DUF4236)
PF14022	Protein of unknown function (DUF4238)
PF14041	LppP/LprE lipoprotein
PF14058	PcfK-like protein
PF14072	DNA-sulfur modification-associated
PF14081	Domain of unknown function (DUF4262)
PF14085	Domain of unknown function (DUF4265)
PF14088	Domain of unknown function (DUF4268)
PF14090	Helix-turn-helix domain
PF14113	Type VI secretion system (T6SS), amidase effector protein 4
PF14131	Domain of unknown function (DUF4298)
PF14145	YrhK-like protein
PF14163	Super-infection exclusion protein B
PF14168	YjzC-like protein
PF14192	Domain of unknown function (DUF4314)
PF14195	Domain of unknown function (DUF4316)
PF14198	Transposon-encoded protein TnpV
PF14200	Ricin-type beta-trefoil lectin domain-like
PF14213	STAS-like domain of unknown function (DUF4325)
PF14216	Domain of unknown function (DUF4326)
PF14236	Domain of unknown function (DUF4338)
PF14239	RRXRR protein
PF14240	YHYH protein
PF14243	ATP-grasp domain, R2K clade family 3
PF14253	Bacteriophage abortive infection AbiH
PF14264	Glucosyl transferase GtrII
PF14265	Domain of unknown function (DUF4355)
PF14279	HNH endonuclease
PF14280	Domain of unknown function (DUF4365)
PF14284	PcfJ-like protein
PF14285	Domain of unknown function (DUF4367)
PF14297	Domain of unknown function (DUF4373)
PF14301	Domain of unknown function (DUF4376)
PF14311	Probable Zinc-ribbon domain
PF14312	FG-GAP repeat
PF14319	Transposase zinc-binding domain
PF14328	Domain of unknown function (DUF4385)
PF14335	Domain of unknown function (DUF4391)

Table 3.4. Continued.

PFAM	Description
PF14337	Abortive infection alpha
PF14340	Domain of unknown function (DUF4395)
PF14349	Motility related/secretion protein
PF14350	Beta protein
PF14353	CpXC protein
PF14359	Domain of unknown function (DUF4406)
PF14362	Domain of unknown function (DUF4407)
PF14367	Domain of unknown function (DUF4411)
PF14373	Superinfection immunity protein
PF14384	BrnA antitoxin of type II toxin-antitoxin system
PF14386	Domain of unknown function (DUF4417)
PF14395	Phage phiEco32-like COOH.NH2 ligase-type 2
PF14424	The BURPS668_1122 family of deaminases
PF14436	Bacterial EndoU nuclease
PF14464	Prokaryotic homologs of the JAB domain
PF14470	Bacterial PH domain
PF14471	Domain of unknown function (DUF4428)
PF14487	ssDNA thymidine ADP-ribosyltransferase, DarT
PF14489	QueF-like protein
PF14501	GHKL domain
PF14520	Helix-hairpin-helix domain
PF14525	AraC-binding-like domain
PF14528	LAGLIDADG-like domain
PF14549	DNA-binding transcriptional regulator Cro
PF14550	Putative phage serine protease XkdF
PF14553	YqbF, hypothetical protein domain
PF14568	SMI1-KNR4 cell-wall
PF14579	Helix-hairpin-helix motif
PF14594	Siphovirus ReqiPepy6 Gp37-like protein
PF14657	Arm DNA-binding domain
PF14659	Phage integrase, N-terminal SAM-like domain
PF14690	zinc-finger of transposase IS204/IS1001/IS1096/IS1165
PF14700	DNA-directed RNA polymerase N-terminal
PF14706	Transposase DNA-binding
PF14730	Domain of unknown function (DUF4468) with TBP-like fold
PF14743	DNA ligase OB-like domain
PF14748	Pyrroline-5-carboxylate reductase dimerisation
PF14791	DNA polymerase beta thumb
PF14819	Nitrile reductase, 7-cyano-7-deazaguanine-reductase N-term
PF14833	NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase
PF14849	YidC periplasmic domain

Table 3.4. Continued.

PFAM	Description
PF14859	Colicin M
PF14890	Intein splicing domain
PF14902	Domain of unknown function (DUF4494)
PF14928	Short tail fibre protein receptor-binding domain
PF14947	Winged helix-turn-helix
PF15515	MvaI/BcnI restriction endonuclease family
PF15542	Bacterial toxin 50
PF15579	Immunity protein 52
PF15601	Immunity protein 70
PF15607	Bacterial toxin 44
PF15609	Phosphoribosyl transferase
PF15610	PRTase ComF-like
PF15644	Papain fold toxin 1, glutamine deamidase
PF15645	Dermonecrotxin of the Papain-like fold
PF15648	Restriction endonuclease fold toxin 5
PF15711	Interleukin-like EMT inducer
PF15731	Antitoxin component of bacterial toxin-antitoxin system, MqsA
PF15738	Bacterial toxin of type II toxin-antitoxin system, YafQ
PF15780	Abnormal spindle-like microcephaly-assoc'd, ASPM-SPD-2-Hydin
PF15891	Nucleoside 2-deoxyribosyltransferase like
PF15892	BNR repeat-containing family member
PF15919	HicB_like antitoxin of bacterial toxin-antitoxin system
PF15933	Antitoxin to bacterial toxin RNase LS or RnIA
PF15935	RNase LS, bacterial toxin
PF15943	Putative antitoxin of bacterial toxin-antitoxin system, YdaS/YdaT
PF15944	Domain of unknown function (DUF4752)
PF15946	Domain of unknown function (DUF4754)
PF15956	Domain of unknown function (DUF4760)
PF15958	Domain of unknown function (DUF4761)
PF15962	Domain of unknown function (DUF4765)
PF15968	Membrane-anchored ion channel, Abi component
PF15969	Intracellular sensor of Lambda phage, Abi component
PF15970	HicB_like antitoxin of bacterial toxin-antitoxin system
PF16075	Domain of unknown function (DUF4815)
PF16079	Phage holin family Hol44, in holin superfamily V
PF16080	Bacteriophage holin family HP1
PF16081	Mycobacterial 2 TMS Phage Holin (M2 Hol) Family
PF16082	Bacteriophage holin family, superfamily II-like
PF16083	LydA holin phage, holin superfamily III
PF16084	LydA-holin antagonist
PF16085	Bacteriophage holin Hol, superfamily III

Table 3.4. Continued.

PFAM	Description
PF16162	Domain of unknown function (DUF4868)
PF16168	Adhesin of bacterial autotransporter system, probable stalk
PF16184	Cadherin-like
PF16190	Ubiquitin-activating enzyme E1 FCCH domain
PF16199	Radical_SAM C-terminal domain
PF16219	Domain of unknown function (DUF4879)
PF16229	Domain of unknown function (DUF4888)
PF16239	Domain of unknown function (DUF4898)
PF16243	Sm_like domain
PF16253	Domain of unknown function (DUF4909)
PF16289	PIN domain
PF16308	Domain of unknown function (DUF4950)
PF16317	Glycosyl hydrolase family 99
PF16363	GDP-mannose 4,6 dehydratase
PF16374	Cycle inhibiting factor (CIF)
PF16452	Bacteriophage CI repressor C-terminal domain
PF16459	Phage tail assembly chaperone, TAC
PF16460	Phage tail tube, TTP, lambda-like
PF16461	Lambda phage tail tube protein, TTP
PF16462	Phage tail assembly chaperone protein, TAC
PF16463	Phage tail tube protein family
PF16466	Domain of unknown function (DUF5047)
PF16468	Domain of unknown function (DUF5049)
PF16473	3' exoribonuclease, RNase T-like
PF16475	Domain of unknown function (DUF5052)
PF16510	Phage P22-like portal protein
PF16518	T3SS negative regulator, GrIR
PF16532	Sf6-type phage tail needle knob or tip of some Caudovirales
PF16542	PNKP adenylyltransferase domain, ligase domain
PF16560	Putative mobile pathogenicity island
PF16576	Barrel-sandwich domain of CusB or HlyD membrane-fusion
PF16677	DNA-packaging protein gp3
PF16684	Telomere resolvase
PF16710	N-terminal N1 domain of Vibrio phage CTXphi pIII
PF16723	Domain of unknown function (DUF5065)
PF16724	T4-like virus Myoviridae tail sheath stabiliser
PF16729	Domain of unknown function (DUF5067)
PF16754	Bacterial toxin homologue of phage lysozyme, C-term
PF16773	Lactococcus phage single-stranded DNA binding protein
PF16774	Baseplate protein
PF16775	Target recognition domain of lytic exoenzyme

Table 3.4. Continued.

PFAM	Description
PF16778	Phage tail assembly chaperone protein
PF16784	Putative HNHc nuclease
PF16786	Recombination enhancement, RecA-dependent nuclease
PF16790	Bacteriophage clamp loader A subunit
PF16792	Phage tail base-plate attachment protein of Caudovirales ORF16
PF16793	RepB DNA-primase from phage plasmid
PF16795	Archaeal phage integrase
PF16805	Phage late-transcription coactivator
PF16807	Domain of unknown function (DUF5072)
PF16814	Read-through domain
PF16822	SGNH hydrolase-like domain, acetyltransferase AlgX
PF16827	zinc-finger
PF16838	Caudoviral major tail protein N-terminus
PF16855	Small outer capsid protein
PF16857	RNA polymerase inhibitor
PF16861	Carbamoyltransferase C-terminus
PF16868	NMT1-like family
PF16898	C-terminal associated domain of TOPRIM
PF16928	DNA/protein translocase of phage P22 injectosome
PF16931	Putative phage holin
PF16938	Putative phage holin Dp-1
PF16945	Putative lactococcus lactis phage r1t holin
PF17091	Inovirus G7P protein
PF17094	Uncharacterised protein family (UPF0715)
PF17164	Domain of unknown function (DUF5122) beta-propeller
PF17191	RecG wedge domain
PF17210	SdrD B-like domain
PF17212	Tail tubular protein
PF17236	Family of unknown function (DUF5309)
PF17268	Family of unknown function (DUF5334)
PF17282	Family of unknown function (DUF5347)
PF17288	Terminase RNaseH like domain
PF17289	Terminase RNaseH-like domain
PF17318	Family of unknown function (DUF5361)
PF17338	Gene product 88
PF17345	Family of unknown function (DUF5375)
PF17356	Phage-like element PBSX protein XtrA
PF17358	Family of unknown function (DUF5384)
PF17363	Family of unknown function (DUF5388)
PF17388	Tail assembly gene products 24 & 25
PF17389	Bacterial alpha-L-rhamnosidase 6 hairpin glycosidase domain

Table 3.4. Continued.

PFAM	Description
PF17395	Family of unknown function (DUF5403)
PF17399	Domain of unknown function (DUF5405)
PF17400	Family of unknown function (DUF5406)
PF17420	Superinfection exclusion gene product 17
PF17426	Putative Gamma DNA binding protein G5P
PF17427	Phage Single-stranded DNA-binding protein
PF17429	Gene product 70
PF17441	Family of unknown function (DUF5419)
PF17448	Uncharacterized YqaH-like
PF17457	Family of unknown function (DUF5420)
PF17463	Gene product 79
PF17466	Family of unknown function
PF17468	Phage gene product 52
PF17469	Phage protein Gp68
PF17470	Phage gene product 45.2
PF17471	Gene product 63
PF17475	Clostridial binary toxin B/anthrax toxin PA domain 2
PF17481	Phage tail sheath protein beta-sandwich domain
PF17482	Phage tail sheath C-terminal domain
PF17510	Gene product 44
PF17519	Family of unknown function (DUF5444)
PF17525	Family of unknown function (DUF5447)
PF17526	Family of unknown function (DUF5448)
PF17527	Phage ALC protein
PF17528	Family of unknown function (DUF5449)
PF17537	Family of unknown function (DUF5455)
PF17545	Family of unknown function (DUF5461)
PF17548	Histone-like Protein p6
PF17549	Gene Product 17
PF17552	Family of unknown function (DUF5464)
PF17565	Family of unknown function (DUF5471)
PF17570	Family of unknown function (DUF5476)
PF17574	Inhibitor of toxin/antitoxin system (Gp4.5)
PF17578	Family of unknown function (DUF5481)
PF17583	Family of unknown function (DUF5484)
PF17585	Accessory recombination function protein
PF17587	Discriminator of mRNA degradation
PF17588	Family of unknown function (DUF5486)
PF17589	Family of unknown function (DUF5487)
PF17592	Family of unknown function (DUF5489)
PF17593	Family of unknown function (DUF5490)

Table 3.4. Continued.

PFAM	Description
PF17594	Phage Tail fiber assembly helper gene product 57
PF17595	Family of unknown function (DUF5491)
PF17597	Family of unknown function (DUF5493)
PF17599	Family of unknown function (DUF5495)
PF17600	Family of unknown function (DUF5496)
PF17602	Family of unknown function (DUF5498)
PF17603	Family of unknown function (DUF5499)
PF17613	Modifier of transcription
PF17615	Family of unknown function
PF17623	Domain of unknown function
PF17634	Gene product 67
PF17639	Family of unknown function (DUF5517)
PF17657	Bacterial DNA polymerase III alpha subunit finger domain
PF17679	gp37/Dip protein
PF17720	Family of unknown function (DUF5565)
PF17726	Dam-replacing HTH domain
PF17728	BsuBI/PstI restriction endonuclease HTH domain
PF17746	SfsA N-terminal OB domain
PF17757	UvrB interaction domain
PF17765	MmyB-like transcription regulator ligand binding domain
PF17767	Nicotinate phosphoribosyltransferase (NAPRTase) N-terminal domain
PF17768	RecJ OB domain
PF17871	AAA lid domain
PF17879	DNA ligase C-terminal domain
PF17884	Domain of unknown function (DUF5591)
PF17891	Mu-like prophage FluMu N-terminal domain
PF17936	Bacterial Ig domain
PF17937	Tetracyclin repressor-like, C-terminal domain
PF17948	DnaT DNA-binding domain
PF17957	Bacterial Ig domain
PF17961	Bacterial Ig domain
PF17963	Bacterial Ig domain
PF17969	L,D-transpeptidase C-terminal domain
PF17975	Ribonucleotide reductase alpha domain
PF17989	Actin like proteins N terminal domain
PF18013	Phage tail lysozyme
PF18043	T4 RNA ligase 2 C-terminal
PF18066	Phage ABA sandwich domain
PF18090	Centromere-binding protein HTH domain
PF18127	Domain of unknown function (DUF5598)
PF18136	DNA mitochondrial polymerase exonuclease domain

Table 3.4. Continued.

PFAM	Description
PF18143	HAD domain in Swiss Army Knife RNA repair proteins
PF18153	SMODS-associating 2TM, beta-strand rich effector domain
PF18157	MID domain of pPIWI_RE
PF18277	AbrB C-terminal domain
PF18299	ATP-grasp domain, R2K clade family 2
PF18316	S-layer protein SbsC C-terminal domain
PF18338	Lower baseplate protein N-terminal domain
PF18341	PSA endolysin C-terminal cell wall binding domain
PF18343	Dda helicase SH3 domain
PF18352	Phage protein Gp138 N-terminal domain
PF18355	Domain of unknown function (DUF5607)
PF18406	Ferredoxin-like domain in Api92-like protein
PF18430	Putative DNA-binding domain
PF18451	Contact-dependent growth inhibition CdiA C-terminal domain
PF18454	Major tropism determinant N-terminal domain
PF18607	ParA helix turn helix domain
PF18628	Viral coat protein P2 N-terminal domain
PF18667	Baseplate upper protein immunoglobulin like domain
PF18668	Tail spike TSP1/Gp66 receptor binding N-terminal domain
PF18715	Phage spike trimer
PF18723	alpha-glutamyl/putresciny l thymine pyrophosphorylase clade 1
PF18724	Alpha-glutamyl/putresciny l thymine pyrophosphorylase clade 2
PF18730	Cthe_2314-like HEPN
PF18737	MAE_28990/MAE_18760-like HEPN
PF18740	EC042_2821-lke REase
PF18747	P-loop Nucleotide Kinase2
PF18748	P-loop Nucleotide Kinase1
PF18751	P-loop Nucleotide Kinase3
PF18756	Nucleotide modification associated domain 4
PF18757	Nucleotide modification associated domain 5
PF18760	ADP-Ribosyltransferase in polyvalent proteins
PF18763	ddrB-like ParB superfamily domain
PF18766	SWI2/SNF2 ATPase
PF18780	Homing endonuclease repeat
PF18781	Phage spike trimer
PF18788	Defence against restriction A N-terminal
PF18789	Defence against restriction A C-terminal
PF18796	Large polyvalent protein-associated domain 1
PF18798	Large polyvalent protein-associated domain 3
PF18799	Large polyvalent protein-associated domain 5
PF18801	response regulator aspartate phosphatase H, N terminal

Table 3.4. Continued.

PFAM	Description
PF18809	phage-Barnase-EndoU-ColicinE5/D-RelE like nuclease1
PF18810	phage-Barnase-EndoU-ColicinE5/D-RelE like nuclease2
PF18812	phage-Barnase-EndoU-ColicinE5/D-RelE like nuclease3
PF18813	phage-Barnase-EndoU-ColicinE5/D-RelE like nuclease4
PF18814	phage-Barnase-EndoU-ColicinE5/D-RelE like nuclease5
PF18818	Zincin-like metallopeptidase
PF18823	Inorganic Pyrophosphatase
PF18824	Large polyvalent protein-associated domain 11
PF18828	Large polyvalent-protein-associated domain 15
PF18834	Large polyvalent protein associated domain 22
PF18836	Beta solenoid repeat from YDCK
PF18838	Large polyvalent protein associated domain 23
PF18847	Large polyvalent protein associated domain 29
PF18853	Large polyvalent protein associated domain 37
PF18857	Large polyvalent protein associated domain 38
PF18858	Large polyvalent protein associated domain 39
PF17552	Family of unknown function (DUF5464)
PF17565	Family of unknown function (DUF5471)
PF17570	Family of unknown function (DUF5476)
PF17574	Inhibitor of toxin/antitoxin system (Gp4.5)
PF17578	Family of unknown function (DUF5481)
PF17583	Family of unknown function (DUF5484)
PF17585	Accessory recombination function protein
PF17587	Discriminator of mRNA degradation
PF17588	Family of unknown function (DUF5486)
PF17589	Family of unknown function (DUF5487)
PF17592	Family of unknown function (DUF5489)
PF17593	Family of unknown function (DUF5490)
PF17594	Phage Tail fiber assembly helper gene product 57
PF17595	Family of unknown function (DUF5491)
PF17597	Family of unknown function (DUF5493)
PF17599	Family of unknown function (DUF5495)
PF17600	Family of unknown function (DUF5496)
PF17602	Family of unknown function (DUF5498)
PF17603	Family of unknown function (DUF5499)
PF17613	Modifier of transcription
PF17615	Family of unknown function
PF17623	Domain of unknown function
PF17634	Gene product 67
PF17639	Family of unknown function (DUF5517)
PF17657	Bacterial DNA polymerase III alpha subunit finger domain

Table 3.4. Continued.

PFAM	Description
PF17679	gp37/Dip protein
PF17720	Family of unknown function (DUF5565)
PF17726	Dam-replacing HTH domain
PF17728	BsuBI/PstI restriction endonuclease HTH domain
PF17746	SfsA N-terminal OB domain
PF17757	UvrB interaction domain
PF17765	MmyB-like transcription regulator ligand binding domain
PF17767	Nicotinate phosphoribosyltransferase (NAPRTase) N-terminal domain
PF17768	RecJ OB domain
PF17871	AAA lid domain
PF17879	DNA ligase C-terminal domain
PF17884	Domain of unknown function (DUF5591)
PF17891	Mu-like prophage FluMu N-terminal domain
PF17936	Bacterial Ig domain
PF17937	Tetracyclin repressor-like, C-terminal domain
PF17948	DnaT DNA-binding domain
PF17957	Bacterial Ig domain
PF17961	Bacterial Ig domain
PF17963	Bacterial Ig domain
PF17969	L,D-transpeptidase C-terminal domain
PF17975	Ribonucleotide reductase alpha domain
PF17989	Actin like proteins N terminal domain
PF18013	Phage tail lysozyme
PF18043	T4 RNA ligase 2 C-terminal
PF18066	Phage ABA sandwich domain
PF18090	Centromere-binding protein HTH domain
PF18127	Domain of unknown function (DUF5598)
PF18136	DNA mitochondrial polymerase exonuclease domain
PF18143	HAD domain in Swiss Army Knife RNA repair proteins
PF18153	SMODS-associating 2TM, beta-strand rich effector domain
PF18157	MID domain of pPIWI_RE
PF18277	AbrB C-terminal domain
PF18299	ATP-grasp domain, R2K clade family 2
PF18316	S-layer protein SbsC C-terminal domain
PF18338	Lower baseplate protein N-terminal domain
PF18341	PSA endolysin C-terminal cell wall binding domain
PF18343	Dda helicase SH3 domain
PF18352	Phage protein Gp138 N-terminal domain
PF18355	Domain of unknown function (DUF5607)
PF18406	Ferredoxin-like domain in Api92-like protein
PF18430	Putative DNA-binding domain

Table 3.4. Continued.

PFAM	Description
PF18451	Contact-dependent growth inhibition CdiA C-terminal domain
PF18454	Major tropism determinant N-terminal domain
PF18607	ParA helix turn helix domain
PF18628	Viral coat protein P2 N-terminal domain
PF18667	Baseplate upper protein immunoglobulin like domain
PF18668	Tail spike TSP1/Gp66 receptor binding N-terminal domain
PF18715	Phage spike trimer
PF18723	alpha-glutamyl/putresciny l thymine pyrophosphorylase clade 1
PF18724	Alpha-glutamyl/putresciny l thymine pyrophosphorylase clade 2
PF18730	Cthe_2314-like HEPN
PF18737	MAE_28990/MAE_18760-like HEPN
PF18740	EC042_2821-lke REase
PF18747	P-loop Nucleotide Kinase2
PF18748	P-loop Nucleotide Kinase1
PF18751	P-loop Nucleotide Kinase3
PF18756	Nucleotide modification associated domain 4
PF18757	Nucleotide modification associated domain 5
PF18760	ADP-Ribosyltransferase in polyvalent proteins
PF18763	ddrB-like ParB superfamily domain
PF18766	SWI2/SNF2 ATPase
PF18780	Homing endonuclease repeat
PF18781	Phage spike trimer
PF18788	Defence against restriction A N-terminal
PF18789	Defence against restriction A C-terminal
PF18796	Large polyvalent protein-associated domain 1
PF18798	Large polyvalent protein-associated domain 3
PF18799	Large polyvalent protein-associated domain 5
PF18801	response regulator aspartate phosphatase H, N terminal
PF18809	phage-Bamase-EndoU-ColicinE5/D-RelE like nuclease1
PF18810	phage-Bamase-EndoU-ColicinE5/D-RelE like nuclease2
PF18812	phage-Bamase-EndoU-ColicinE5/D-RelE like nuclease3
PF18813	phage-Bamase-EndoU-ColicinE5/D-RelE like nuclease4
PF18814	phage-Bamase-EndoU-ColicinE5/D-RelE like nuclease5
PF18818	Zincin-like metallopeptidase
PF18823	Inorganic Pyrophosphatase
PF18824	Large polyvalent protein-associated domain 11
PF18828	Large polyvalent-protein-associated domain 15
PF18834	Large polyvalent protein associated domain 22
PF18836	Beta solenoid repeat from YDCK
PF18838	Large polyvalent protein associated domain 23
PF18847	Large polyvalent protein associated domain 29

Table 3.4. Continued.

PFAM	Description
PF18853	Large polyvalent protein associated domain 37
PF18857	Large polyvalent protein associated domain 38
PF18858	Large polyvalent protein associated domain 39

Table 3.5. CTn or Phage Associated TIGRFAMs.

TIGRFAM	Description	
CTn/ICE TIGRFAMs		
TIGR00180	ParB-like partition proteins	
TIGR00621	single-strand binding protein	
TIGR00929	VirB4_CagE	type IV secretion/conjugal transfer ATPase, VirB4 family
TIGR01690	ICE_RAQPRD	integrative conjugative element protein, RAQPRD family
TIGR01764		excise: DNA binding domain; excisionase family
TIGR02524	dot_icm_DotB	Dot/Icm secretion system ATPase DotB
TIGR02525	plasmid_TraJ	plasmid transfer ATPase TraJ
TIGR02527	dot_icm_IcmQ	Dot/Icm secretion system protein IcmQ
TIGR02686	relax_trwC	conjugative relaxase domain
TIGR02738	TrbB	type-F conjugative transfer system pilin assembly thiol-disulfide isomerase
TIGR02739	TraF	type-F conjugative transfer system pilin assembly protein TraF
TIGR02740	TraF-like	TraF-like protein
TIGR02741	TraQ	type-F conjugative transfer system pilin chaperone TraQ
TIGR02742	TrbC_Ftype	type-F conjugative transfer system pilin assembly protein TrbC
TIGR02743	TraW	type-F conjugative transfer system protein TraW
TIGR02744	TrbI_Ftype	type-F conjugative transfer system protein TrbI
TIGR02746	TraC-F-type	type-IV secretion system protein TraC
TIGR02747	TraV	type IV conjugative transfer system protein TraV
TIGR02750	TraN_Ftype	type-F conjugative transfer system mating-pair stabilization protein TraN
TIGR02755	TraX_Ftype	type-F conjugative transfer system pilin acetylase TraX
TIGR02756	TraK_Ftype	type-F conjugative transfer system secretin TraK
TIGR02758	TraA_TIGR	type IV conjugative transfer system pilin TraA
TIGR02759	TraD_Ftype	type IV conjugative transfer system coupling protein TraD
TIGR02760	TraI_TIGR	conjugative transfer relaxase protein TraI
TIGR02761	TraE_TIGR	type IV conjugative transfer system protein TraE
TIGR02762	TraL_TIGR	type IV conjugative transfer system protein TraL
TIGR02767	TraG-Ti	Ti-type conjugative transfer system protein TraG
TIGR02768	TraA_Ti	Ti-type conjugative transfer relaxase TraA
TIGR02771	TraF_Ti	conjugative transfer signal peptidase TraF
TIGR02775	TrbG_Ti	P-type conjugative transfer protein TrbG
TIGR02780	TrbJ_Ti	P-type conjugative transfer protein TrbJ
TIGR02781	VirB9	P-type conjugative transfer protein VirB9
TIGR02782	TrbB_P	P-type conjugative transfer ATPase TrbB
TIGR02783	TrbL_P	P-type conjugative transfer protein TrbL
TIGR03344	VI_effect_Hcp1	type VI secretion system effector, Hcp1 family
TIGR03345	VI_ClpV1	type VI secretion ATPase, ClpV1 family
TIGR03346	chaperone_ClpB	ATP-dependent chaperone protein ClpB

Table 3.5. Continued.

TIGRFAM	Description	
TIGR03347	VI_chp_1	type VI secretion protein, VC_A0111 family
TIGR03348	VI_lcmF	type VI secretion protein lcmF
TIGR03349	IV_VI_DotU	type IV/VI secretion system protein, DotU family
TIGR03350	type_VI_ompA	type VI secretion system peptidoglycan-associated domain
TIGR03351	PhnX-like	phosphonate-like hydrolase
TIGR03352	VI_chp_3	type VI secretion lipoprotein, VC_A0113 family
TIGR03353	VI_chp_4	type VI secretion protein, VC_A0114 family
TIGR03354	VI_FHA	type VI secretion system FHA domain protein
TIGR03355	VI_chp_2	type VI secretion protein, EvpB/VC_A0108 family
TIGR03356	BGL	beta-galactosidase
TIGR03357	VI_zyyme	type VI secretion system lysozyme-like protein
TIGR03358	VI_chp_5	type VI secretion protein, VC_A0107 family
TIGR03359	VI_chp_6	type VI secretion protein, TIGR03359 family
TIGR03360	VI_minor_1	type VI secretion-associated protein, VC_A0118 family
TIGR03361	VI_Rhs_Vgr	type VI secretion system Vgr family protein
TIGR03362	VI_chp_7	type VI secretion-associated protein, VC_A0119 family
TIGR03363	VI_chp_8	type VI secretion-associated protein, ImpA family
TIGR03373	VI_minor_4	type VI secretion-associated protein, BMA_A0400 family
TIGR03734	PRTRC_parB	
TIGR03735	PRTRC_A	
TIGR03736	PRTRC_ThiF	
TIGR03737	PRTRC_B	
TIGR03738	PRTRC_C	
TIGR03739	PRTRC_D	
TIGR03741	PRTRC_E	
TIGR03742	PRTRC_F	
TIGR03743	SXT_TraD	conjugative coupling factor TraD, SXT/TOL subfamily
TIGR03744	traC_PFL_4706	conjugative transfer ATPase, PFL_4706 family
TIGR03745	conj_TIGR03745	integrating conjugative element membrane protein, PFL_4702 family
TIGR03746	conj_TIGR03746	integrating conjugative element protein, PFL_4703 family
TIGR03747	conj_TIGR03747	integrating conjugative element membrane protein, PFL_4697 family
TIGR03748	conj_PiIL	integrating conjugative element protein PiIL, PFGI-1 class
TIGR03749	conj_TIGR03749	integrating conjugative element protein, PFL_4704 family
TIGR03750	conj_TIGR03750	conjugative transfer region protein, TIGR03750 family
TIGR03751	conj_TIGR03751	conjugative transfer region lipoprotein, TIGR03751 family
TIGR03752	conj_TIGR03752	integrating conjugative element protein, PFL_4705 family
TIGR03753	blh_monoox	beta-carotene 15,15'-monooxygenase, Brp/Blh family
TIGR03754	conj_TOL_TraD	conjugative coupling factor TraD, PFGI-1 class

Table 3.5. Continued.

TIGRFAM	Description	
TIGR03755	conj_TIGR03755	integrating conjugative element protein, PFL_4711 family
TIGR03756	conj_TIGR03756	integrating conjugative element protein, PFL_4710 family
TIGR03757	conj_TIGR03757	integrating conjugative element protein, PFL_4709 family
TIGR03758	conj_TIGR03758	integrating conjugative element protein, PFL_4701 family
TIGR03759	conj_TIGR03759	integrating conjugative element protein, PFL_4693 family
TIGR03760	ICE_TraI_Pfluor	integrating conjugative element relaxase, PFGI-1 class
TIGR03761	ICE_PFL4669	integrating conjugative element protein, PFL_4669 family
TIGR03764	ICE_PFGI_1_parB	integrating conjugative element, PFGI_1 class, ParB family protein
TIGR03765	ICE_PFL_4695	integrating conjugative element protein, PFL_4695 family
TIGR03779	Bac_Flav_CT_M	
TIGR03780	Bac_Flav_CT_N	
TIGR03781	Bac_Flav_CT_K	
TIGR03782	Bac_Flav_CT_J	
TIGR03783	Bac_Flav_CT_G	
TIGR04346	DotA_TraY	conjugal transfer/type IV secretion protein DotA/TraY
TIGR04359	TrbK_RP4	entry exclusion lipoprotein TrbK
TIGR04360	other_trbK	conjugal transfer region protein TrbK
TIGR04361	TrbK_Ti	entry exclusion protein TrbK
Phage TIGRFAMs		
TIGR00616	rect	recombinase, phage RecT family
TIGR01537	portal_HK97	phage portal protein, HK97 family
TIGR01538	portal_SPP1	phage portal protein, SPP1 family
TIGR01539	portal_lambda	phage portal protein, lambda family
TIGR01540	portal_PBSX	phage portal protein, PBSX family
TIGR01541	tape_meas_lam_C	phage tail tape measure protein, lambda family
TIGR01542	A118_put_portal	phage portal protein, putative, A118 family
TIGR01543	proheadase_HK97	phage prohead protease, HK97 family
TIGR01547	phage_term_2	phage terminase, large subunit, PBSX family
TIGR01551	major_capsid_P2	phage major capsid protein, P2 family
TIGR01554	major_cap_HK97	phage major capsid protein, HK97 family
TIGR01555	phge_rel_HI1409	phage-associated protein, HI1409 family
TIGR01558	sm_term_P27	putative phage terminase, small subunit, P27 family
TIGR01560	put_DNA_pack	uncharacterized phage protein (possible DNA packaging)
TIGR01563	gp16_SPP1	putative phage head-tail adaptor
TIGR01593	holin_tox_secr	toxin secretion/phage lysis holin
TIGR01594	holin_lambda	phage holin, lambda family
TIGR01598	holin_phiLC3	holin, phage phi LC3 family
TIGR01600	phage_tail_L	phage minor tail protein L

Table 3.5. Continued.

TIGRFAM	Description	
TIGR01603	maj_tail_phi13	phage major tail protein, phi13 family
TIGR01610	phage_O_Nterm	phage replication protein O, N-terminal domain
TIGR01611	tail_tube	phage major tail tube protein
TIGR01613	primase_Cterm	phage/plasmid primase, P4 family, C-terminal domain
TIGR01618	phage_P_loop	phage nucleotide-binding protein
TIGR01629	rep_II_X	phage/plasmid replication protein, gene II/X family
TIGR01630	psiM2_ORF9	phage uncharacterized protein, C-terminal domain
TIGR01633	phi3626_gp14_N	putative phage tail component, N-terminal domain
TIGR01634	tail_P2_I	phage tail protein I
TIGR01635	tail_comp_S	phage virion morphogenesis protein
TIGR01636	phage_rinA	phage transcriptional regulator, RinA family
TIGR01637	phage_arpU	phage transcriptional regulator, ArpU family
TIGR01641	phageSPP1_gp7	phage head morphogenesis protein, SPP1 gp7 family
TIGR01644	phage_P2_V	phage baseplate assembly protein V
TIGR01665	put_anti_recept	phage minor structural protein, N-terminal region
TIGR01669	phage_XkdX	phage uncharacterized protein, XkdX family
TIGR01671	phage_TIGR01671	phage conserved hypothetical protein TIGR01671
TIGR01673	holin_LLH	phage holin, LL-H family
TIGR01674	phage_lambda_G	phage minor tail protein G
TIGR01712	phage_N6A_met	phage N-6-adenine-methyltransferase
TIGR01714	phage_rep_org_N	phage replisome organizer, putative, N-terminal region
TIGR01715	phage_lam_T	phage tail assembly protein T
TIGR01725	phge_HK97_gp10	phage protein, HK97 gp10 family
TIGR01760	tape_meas_TP901	phage tail tape measure protein, TP901 family, core region
TIGR01913	bet_lambda	phage recombination protein Bet
TIGR02126	phgtail_TP901_1	phage major tail protein, TP901-1 family
TIGR02215	phage_chp_gp8	phage conserved hypothetical protein, phiE125 gp8 family
TIGR02216	phage_TIGR02216	phage conserved hypothetical protein
TIGR02218	phg_TIGR02218	phage conserved hypothetical protein BR0599
TIGR02219	phage_NlpC_fam	putative phage cell wall peptidase, NlpC/P60 family
TIGR02220	phg_TIGR02220	phage conserved hypothetical protein, C-terminal domain
TIGR02241	TIGR02241	conserved hypothetical phage tail region protein
TIGR02242	tail_TIGR02242	phage tail protein domain
TIGR02419	C4_traR_proteo	phage/conjugal plasmid C-4 type zinc finger protein, TraR family
TIGR02642	phage_xxxx	uncharacterized phage protein
TIGR02681	phage_pRha	phage regulatory protein, Rha family
TIGR02974	phageshock_pspF	psp operon transcriptional activator
TIGR02975	phageshock_pspG	phage shock protein G

Table 3.5. Continued.

TIGRFAM	Description	
TIGR02976	phageshock_pspB	phage shock protein B
TIGR02977	phageshock_pspA	phage shock protein A
TIGR02978	phageshock_pspC	phage shock protein C
TIGR02979	phageshock_pspD	phage shock protein PspD
TIGR02981	phageshock_pspE	phage shock operon rhodanese PspE
TIGR03033	phage_rel_nuc	putative phage-type endonuclease
TIGR03299	LGT_TIGR03299	phage/plasmid-like protein TIGR03299
TIGR03495	phage_LysB	phage lysis regulatory protein, LysB family
TIGR03600	phage_DnaB	phage replicative helicase, DnaB family
TIGR04111	BcepMu_gp16	phage-associated protein, BcepMu gp16 family

Table 3.6. Bacterial Strains Used in this Study.

Shortname in text	Relevant Phenotype		
	CTn	B ₁₂ Transport	Resistance
Parental and Recipient Strains:			
<i>B. thetaiotaomicron</i>	—	+	—
<i>B. thetaiotaomicron</i> Erm ^R	—	+	Erm ^R
<i>B. ovatus</i>	BoCTn	+	—
<i>B. thetaiotaomicron</i> 7330	BtCTn	+	—
<i>B. thetaiotaomicron</i> ΔΔΔ	—	-	—
<i>B. thetaiotaomicron</i> ΔΔΔ Erm ^R	—	-	Erm ^R
<i>B. thetaiotaomicron</i> ΔΔΔ Tet ^R	—	-	Tet ^R
Transconjugants from Parental Strains:			
<i>B. thetaiotaomicron</i> ΔΔΔ Erm ^R + BoCTn	BoCTn	+	Erm ^R
<i>B. thetaiotaomicron</i> ΔΔΔ Erm ^R + BtCTn	BtCTn	+	Erm ^R
Transconjugants from Transconjugant Strains:			
<i>B. thetaiotaomicron</i> ΔΔΔ Tet ^R + BoCTn	BoCTn	+	Tet ^R
<i>B. thetaiotaomicron</i> ΔΔΔ Tet ^R + BtCTn	BtCTn	+	Tet ^R
Unmarked transconjugant strains:			
<i>B. thetaiotaomicron</i> ΔΔΔ + BoCTn	BoCTn	+	—
<i>B. thetaiotaomicron</i> ΔΔΔ + BtCTn	BtCTn	+	—
Barcoded Strains for competitions:			
<i>B. thetaiotaomicron</i> Δ <i>locus1</i> Δ <i>locus3</i> BC	—	+, <i>locus2</i> only	Tet ^R
<i>B. thetaiotaomicron</i> Δ <i>locus1</i> Δ <i>locus2</i> BC	—	+, <i>locus3</i> only	Tet ^R
<i>B. thetaiotaomicron</i> Δ <i>BT1953</i> Δ <i>BT2094</i> BC	—	+, <i>locus1</i> <i>btuB</i>	Tet ^R
<i>B. thetaiotaomicron</i> ΔΔΔ BC			
<i>B. thetaiotaomicron</i> ΔΔΔ + BoCTn BC	BoCTn	+	Tet ^R
<i>B. thetaiotaomicron</i> ΔΔΔ + BtCTn BC	BtCTn	+	Tet ^R
Strains carrying tetracycline resistance marked CTns:			
<i>B. thetaiotaomicron</i> ΔΔΔ + BoCTn_Tet ^R	BoCTn	+	Tet ^R
<i>B. thetaiotaomicron</i> ΔΔΔ + BtCTn_Tet ^R	BtCTn	+	Tet ^R
Transconjugants isolated from the mouse gut:			
Mouse-isolated BoCTn TC-1	BoCTn	+	Erm ^R , Tet ^R
Mouse-isolated BoCTn TC-2	BoCTn	+	Erm ^R , Tet ^R
Mouse-isolated BtCTn TC-1	BtCTn	+	Erm ^R , Tet ^R
Mouse-isolated BtCTn TC-2	BtCTn	+	Erm ^R , Tet ^R

Table 3.7. Transfer Frequencies from Conjugations with Predicted Corrinoid Transport-Encoding CTns.

Conjugation Pairs		Transfer Frequency					Colonies that regained growth	
		Replicate 1	Replicate 2	Replicate 3	Replicate 4	Averages:	In MM + B ₁₂	
<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 TN Library	x	<i>B. ovatus</i> ATCC 8483	1.17E-07	1.31E-08	5.20E-08	4.08E-05	6.08E-08	47/48
		<i>B. thetaiotaomicron</i> 7330	2.13E-06	2.61E-07	2.26E-06	1.08E-04	1.55E-06	47/48
<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 att1::pNB _{U2} -bla- ermGp	x	<i>B. ovatus</i> ATCC 8483	1.38E-07	1.24E-07			1.31E-07	16/16
		<i>B. thetaiotaomicron</i> 7330	6.53E-06	1.15E-05			9.01E-06	16/16
<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 att1::pNB _{U2} -bla-tetQ	x	<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 TN + BoCTn transconjugate #1	1.28E-05	6.25E-04			1.68E-04	
		<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 TN + BoCTn transconjugate #2	1.70E-05	1.66E-05				
		<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 TN + BIC _{Tn} transconjugate #1	5.21E-06	7.78E-06			3.99E-06	
		<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 TN + BIC _{Tn} transconjugate #2	7.20E-07	2.26E-06				
<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 pXchange_tdk	x	<i>B. ovatus</i> ATCC 8483	2.56E-07	4.58E-07			3.57E-07	
		<i>B. thetaiotaomicron</i> 7330	1.78E-05	1.03E-05			1.40E-05	

Table 3.7. Continued.

Conjugation Pairs		Transfer Frequency					Colonies that regained growth in MM + B ₁₂
		Replicate 1	Replicate 2	Replicate 3	Replicate 4	Averages:	
<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 att1::pNBU2-bla- ermGb	x	<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BoCTn	5.38E-06			N/A	
		<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BICTn	4.59E-06			N/A	
<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 att1::pNBU2-bla- ermGb	x	<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BoCTn, tetQ	1.36E-06	5.87E-06		3.62E-06	
		<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BICTn, tetQ	3.68E-07	7.30E-06		3.84E-06	
<i>B. uniformis</i> ATCC 8492 attN::pNBU2_erm	x	<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BoCTn, tetQ	1.40E-06	1.50E-07		7.75E-07	
		<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BICTn, tetQ	8.33E-07	6.00E-06		3.42E-06	
<i>B. vulgatus</i> ATCC 8482 attN::pNBU2_erm	x	<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BoCTn, tetQ	3.33E-09	N.D.		N/A	
		<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BICTn, tetQ	1.64E-09	N.D.		N/A	
<i>B. fragilis</i> NCTC 9343 attN::pNBU2_erm	x	<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BoCTn, tetQ	N.D.	N.D.		N.D.	
		<i>B. thetaioaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BICTn, tetQ	N.D.	N.D.		N.D.	

Table 3.7. Continued.

Conjugation Pairs		Transfer Frequency					Colonies that regained growth in MM + B ₁₂	
		Replicate 1	Replicate 2	Replicate 3	Replicate 4	Averages:		
<i>B. caccae</i> ATCC 43185 attN::pNB _{U2_erm}	x	<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BoCTn, tetQ	N.D.				N.D.	
		<i>B. thetaiotaomicron</i> VP1-5482 Δtdk Δlocus1Δlocus2Δlocus3 + BtCTn, tetQ	N.D.				N.D.	

N.D. = No colonies detected

Table 3.8. DNA sequencing data generated for *Bacteroidetes* transconjugants used in this study.

Sample	Sequencer	Chemistry	Read length (bp)	Read pairs	Quality filtered read pairs	% reads mapping	Locus 1 Δ9,520 bp	Locus 2 Δ12,398 bp	Locus 3 Δ12,430 bp	BoCtn	BICtn
<i>B. thetaiotaomicron</i> ΔΔΔ + BoCtn str. 1181	HiSeq2500	HiSeq SBS v4	2×160	3,352,841	2,187,144	98.80	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ + BoCtn str. 1184	HiSeq2500	HiSeq SBS v4	2×160	2,884,819	1,963,242	98.30	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ + BoCtn str. 1762	HiSeq2500	HiSeq SBS v4	2×160	2,555,977	1,717,852	98.80	✓	✓	✓	<i>attB</i> 2	-
<i>B. thetaiotaomicron</i> ΔΔΔ + BoCtn str. 1763	HiSeq2500	HiSeq SBS v4	2×160	3,482,847	2,310,638	98.90	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ EmrR + BoCtn str. 1764	HiSeq2500	HiSeq SBS v4	2×160	3,757,938	2,406,035	98.40	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ EmrR + BoCtn str. 1765	HiSeq2500	HiSeq SBS v4	2×160	3,213,507	2,135,275	98.90	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ EmrR + BoCtn str. 1766	HiSeq2500	HiSeq SBS v4	2×160	2,278,082	1,504,216	98.70	✓	✓	✓	<i>attB</i> 4	-
<i>B. thetaiotaomicron</i> ΔΔΔ EmrR + BoCtn str. 1767	HiSeq2500	HiSeq SBS v4	2×160	1,929,610	1,327,049	98.90	✓	✓	✓	<i>attB</i> 3	-
<i>B. thetaiotaomicron</i> ΔΔΔ TeIR + BoCtn str. 1768	HiSeq2500	HiSeq SBS v4	2×160	3,518,376	2,263,374	98.60	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ TeIR + BoCtn str. 1769	HiSeq2500	HiSeq SBS v4	2×160	3,736,694	2,255,586	98.50	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ TeIR + BoCtn str. 1770	HiSeq2500	HiSeq SBS v4	2×160	2,009,470	1,425,932	98.60	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ + BICtn str. 1182	HiSeq2500	HiSeq SBS v4	2×160	2,637,255	1,854,213	97.20	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ + BICtn str. 1185	HiSeq2500	HiSeq SBS v4	2×160	3,148,739	1,942,612	97.10	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ + BICtn str. 1186	HiSeq2500	HiSeq SBS v4	2×160	3,650,649	2,270,219	96.70	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ + BICtn str. 1771	HiSeq2500	HiSeq SBS v4	2×160	2,685,944	1,678,590	97.20	✓	✓	✓	<i>attB</i> 1	Δ12,623 bp
<i>B. thetaiotaomicron</i> ΔΔΔ + BICtn str. 1772	HiSeq2500	HiSeq SBS v4	2×160	3,424,031	2,281,446	97.60	✓	✓	✓	<i>attB</i> 1	Δ12,623 bp
<i>B. thetaiotaomicron</i> ΔΔΔ EmrR + BICtn str. 1773	HiSeq2500	HiSeq SBS v4	2×160	2,729,998	1,782,972	97.90	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ EmrR + BICtn str. 1774	HiSeq2500	HiSeq SBS v4	2×160	2,732,204	1,794,684	98.10	✓	✓	✓	<i>attB</i> 1	Δ12,623 bp
<i>B. thetaiotaomicron</i> ΔΔΔ EmrR + BICtn str. 1775	HiSeq2500	HiSeq SBS v4	2×160	2,528,220	1,709,134	97.40	✓	✓	✓	<i>attB</i> 1	Δ12,623 bp
<i>B. thetaiotaomicron</i> ΔΔΔ TeIR + BICtn str. 1776	HiSeq2500	HiSeq SBS v4	2×160	2,410,614	1,744,718	97.90	✓	✓	✓	<i>attB</i> 1	Δ12,623 bp
<i>B. thetaiotaomicron</i> ΔΔΔ TeIR + BICtn str. 1777	HiSeq2500	HiSeq SBS v4	2×160	3,277,714	2,221,034	98.30	✓	✓	✓	<i>attB</i> 1	-
<i>B. thetaiotaomicron</i> ΔΔΔ TeIR + BICtn str. 1778	HiSeq2500	HiSeq SBS v4	2×160	3,383,702	2,184,694	97.10	✓	✓	✓	<i>attB</i> 2	-

attB 1 = BT1410/BT14141, *attB* 2 = BT2505/BT2506, *attB* 3 = BT4227, *attB* 4 = BT4088, Δ12,623 bp corresponds to loss of restriction modification integron Btne7332251-66

Table 3.9. BICtIn Element Protein Domains and Families.

Start	Stop	Strand	Gene	Synonym	Gene Category	P FAM	P FAM_description	TIGRFAM	TIGRFAM_description
2722543	2722677	-	-	Bthe7332230	Non-element				
2722819	2724174	+	-	Bthe7332231	Non-element				
2724174	2726948	+	-	Bthe7332232	Non-element				
2727211	2727891	-	-	Bthe7332233	Non-element				
2727118	2727173	-	-	Bthe733_rep1A	Non-element				
2727888	2729015	-	-	Bthe7332234	Non-element				
2730344	2730361			Bthe733_atl	atl				
2730706	2731719	+	-	Bthe7332235	integrase	PF00589, PF13102	Phage integrase family , Phage integrase SAM-like domain		
2731740	2731856	+	-	Bthe7332236					
2731919	2732623	+	-	Bthe7332237					
2732815	2733198	+	-	Bthe7332238		PF12728	Helix-turn-helix domain	TIGR01764	excise: DNA binding domain
2733212	2734276	+	-	Bthe7332239		PF13481	AAA domain		
2734380	2735267	+	trap	Bthe7332240		PF13155	Toprin-like		
2735512	2735958	+	-	Bthe7332241					
2735981	2736325	-	-	Bthe7332242					
2736341	2736631	+	traE1	Bthe7332243	Not present in BICtIn	PF13572	Domain of unknown function (DUF4134)		
2737173	2738984	+	-	Bthe7332244	Not present in BICtIn	PF00078, PF01348	Reverse transcriptase (RNA-dependent DNA polymerase , Type II intron maturase	TIGR04416	group II RT mat: group II intron reverse transcrip
2739081	2739203	+	traE2	Bthe7332245		PF13572	Domain of unknown function (DUF4134)		
2739219	2739515	+	-	Bthe7332246					
2739672	2742335	+	-	Bthe7332247					
2742422	2744893	+	-	Bthe7332248					
2744890	2745615	+	-	Bthe7332249					
2745655	2746401	+	-	Bthe7332250		PF16464	Domain of unknown function (DUF5045)		

Table 3.9. Continued.

Start	Stop	Strand	Gene	Synonym	Gene Category	P FAM	P FAM_description	TIGRFAM	TIGRFAM_description
2746414	2746641	+	-	Bthe7332251					
2747232	2747495	+	-	Bthe7332252	Not present in BoCTn	PF12728	Helix-turn-helix domain	TIGR01764	excise: DNA binding domain
2747501	2747848	+	-	Bthe7332253	Not present in BoCTn				
2747898	2748167	+	-	Bthe7332254	Not present in BoCTn				
2748967	2749374	+	-	Bthe7332255	Not present in BoCTn				
2749390	2750601	+	ORF1	Bthe7332256	Not present in BoCTn and Mobilization gene	PF03432	Relaxase/Mobilisation nuclease domain		
2750626	2751171	+	-	Bthe7332257	Not present in BoCTn				
2751181	2751918	-	-	Bthe7332258	Not present in BoCTn	PF08378, PF01396	Nuclease-related domain , Topoisomerase DNA binding C4 zinc finger		
2751928	2752371	-	-	Bthe7332259	Not present in BoCTn				
2752427	2753146	-	-	Bthe7332260	Not present in BoCTn				
2753184	2753798	-	-	Bthe7332261	Not present in BoCTn				
2753834	2754721	-	-	Bthe7332262	Not present in BoCTn	PF06616	BsubI/PstI restriction endonuclease C-terminus		
2754754	2756208	-	-	Bthe7332263	Not present in BoCTn	PF07669	Eco67I restriction-modification methylase		
2756210	2756422	-	-	Bthe7332264	Not present in BoCTn	PF13443	Cro/C1-type HTH DNA-binding domain		
2756559	2757737	-	-	Bthe7332265	Not present in BoCTn	PF00589	Phage integrase family		
2757734	2758915	-	-	Bthe7332266	Not present in BoCTn	PF00589	Phage integrase family		
2759184	2760170	+	-	Bthe7332267					
2760212	2760607	+	-	Bthe7332268					

Table 3.9. Continued.

Start	Stop	Strand	Gene	Synonym	Gene Category	P FAM	P FAM_description	TIGRFAM	TIGRFAM_description
2760611	2761225	+	traK	Bthe7332269	Transfer element			TIGR03781	Bac_Flav_CT_K: Bacteroides conjugative transposon
2761238	2761618	+	-	Bthe7332270					
2761629	2762111	+	-	Bthe7332271		PF05869	DNA N-6-adenine-methyltransferase (Dam)		
2762119	2763270	+	traM	Bthe7332272	Transfer element	PF12508	Conjugative transposon	TIGR03779	Bac_Flav_CT_M: Bacteroides conjugative transposon
2763282	2764085	+	-	Bthe7332273		PF01555	DNA methylase		
2764114	2764965	+	traN	Bthe7332274	Transfer element	PF13595	Domain of unknown function (DUF4138)	TIGR03780	Bac_Flav_CT_N: Bacteroides conjugative transposon
2764962	2765486	+	-	Bthe7332275					
2765513	2766169	+	-	Bthe7332276					
2766183	2766917	+	-	Bthe7332277		PF03118	Bacterial RNA polymerase		
2766921	2769107	+	ORF2	Bthe7332278	Transfer element	PF02534	Type IV secretory system Conjugative DNA transfer		
2769215	2770435	-	-	Bthe7332279		PF06250	Protein of unknown function (DUF1016)		
2770578	2770772	+	-	Bthe7332280					
2770862	2770981	+	-	Bthe7332281					
2771071	2771445	-	-	Bthe7332282					
2771479	2771751	-	-	Bthe7332283					
2771759	2772124	-	-	Bthe7332284					
2772479	2774212	-	-	Bthe7332285		PF02195	ParB-like nuclease domain	TIGR00180	parB_part: ParB/RepB/SpoJ family partition protei
2774637	2774891	-	-	Bthe7332286					
2774899	2775654	-	traA	Bthe7332287		PF13614, PF01656, PF06564, PF10609	AAA domain , CobQ/CobB/Mind/Para nucleotide binding domain , Cellulose biosynthesis protein BcsQ , NUBPL iron-transfer P-loop NTPase		

Table 3.9. Continued.

Start	Stop	Strand	Gene	Synonym	Gene Category	P FAM	P FAM_description	TIGRFAM	TIGRFAM_description
2800438	2800602	+	-	Bthe7332311					
2800626	2800814	-	-	Bthe7332312					
2800824	2801294	+	-	Bthe7332313					
2801300	2802988	+	-	Bthe7332314		PF08401	Domain of unknown function (DUF1738)		
2803065	2803466	-	-	Bthe7332315	Not present in BoCTn				
2803488	2804723	-	-	Bthe7332316	Not present in BoCTn				
2804720	2805493	-	-	Bthe7332317	Not present in BoCTn	PF17131, PF07044, PF07044	Outer membrane lipoprotein-sorting protein , Protein of unknown function (DUF1329) , Protein of unknown function (DUF1329)		
2805490	2807859	-	-	Bthe7332318	Not present in BoCTn	PF03176, PF03176, PF02460	MMPL family , MMPL family , Patched family		
2807881	2808504	-	-	Bthe7332319	Not present in BoCTn	PF08241, PF13649, PF13847, PF13489, PF08242, PF01209, PF05401, PF02353	Methyltransferase domain , Methyltransferase domain , Methyltransferase domain , Methyltransferase domain , ubiE/COQ5 methyltransferase family , Nodulation protein S (NodS) , Mycolic acid cyclopropane synthetase		
2808658	2809266	-	-	Bthe7332320	Not present in BoCTn	PF00440	Bacterial regulatory proteins		
2809594	2810139	-	-	Bthe7332321					
2810360	2810500	+	-	Bthe7332322					
2810488	2810970	-	-	Bthe7332323					
2811203	2811319	+	-	Bthe7332324					
2811371	2811484	-	-	Bthe7332325					
2811484	2811676	+		RF00174					
2812010	2812801	+	-	Bthe7332326		PF14717	Domain of unknown function (DUF4465)		

Table 3.9. Continued.

Start	Stop	Strand	Gene	Synonym	Gene Category	PFAM	PFAM_description	TIGRFAM	TIGRFAM_description
2812803	2814704	+	-	BtHe7332327		PF16820, PF16820, PF00801	PKD-like domain , PKD-like domain , PKD domain		
2814713	2816761	+	-	BtHe7332328	Not present in BoCTn	PF16341	Domain of unknown function (DUF4971)		
2816846	2817925	+	btuG	BtHe7332329		PF05096	Glutamine cyclotransferase		
2818074	2820029	+	btuB	BtHe7332330	BtuB	PF07715	TonB-dependent Receptor Plug Domain		
2820034	2821581	+	btuF	BtHe7332331	Btu operon	PF01497	Periplasmic binding protein		
2821552	2822154	+	btuC	BtHe7332333	Btu operon	PF01032	FecCD transport family		
2822151	2822909	+	btuD	BtHe7332334	Btu operon	PF00005, PF13304	ABC transporter , AAA domain		
2822906	2823502	+	-	BtHe7332335		PF10050	Predicted metal-binding protein (DUF2284)		
2823603	2825003	-	-	BtHe7332336		PF00691	OmpA family		
2825058	2826515	-	-	BtHe7332337	Mobilization gene	PF01076	Plasmid recombination enzyme		
2826734	2827579	-	-	BtHe7332338					
2827569	2828159	-	-	BtHe7332339					
2828150	2828527	-	-	BtHe7332340					
2828548	2829018	-	-	BtHe7332341					
2829022	2829729	-	-	BtHe7332342					
2829757	2829987	+	-	BtHe7332343					
2830110	2830292	-	-	BtHe7332344					
2830310	2830921	+	-	BtHe7332345		PF00239	Resolvase		
2831197	2831478	+	-	BtHe7332346		PF00216	Bacterial DNA-binding protein		
2831566	2831946	-	-	BtHe7332347		PF00436	Single-strand binding protein family		
2832160	2832303	+	-	BtHe7332348					
2832254	2832565	+	-	BtHe7332349					
2832627	2832884	-	-	BtHe7332350					
2832901	2833227	-	-	BtHe7332351					
2833239	2833415	-	-	BtHe7332352					

Table 3.9. Continued.

Start	Stop	Strand	Gene	Synonym	Gene Category	P FAM	P FAM_description	TIGRFAM	TIGRFAM_description
2776014	2780987	-	-	Bthe7332288		PF00271, PF04851	Helicase conserved C-terminal domain , Type III restriction enzyme		
2781576	2782208	-	-	Bthe7332289					
2782266	2782424	-	-	Bthe7332290					
2782484	2782618	-	-	Bthe7332291					
2782947	2783693	+	-	Bthe7332292					
2783690	2784763	+	-	Bthe7332293		PF08843	Nucleotidyl transferase AbiE11 toxin		
2784760	2785494	+	-	Bthe7332294		PF01902	Dipthamide synthase	TIGR00290	MJ0570_dom: MJ0570-related uncharacterized domain
2785528	2786046	+	-	Bthe7332295		PF02572	ATP:corninoid adenosyltransferase BtuR/CobO/CobP	TIGR00708	cobA: cob(II)ynic acid a
2786195	2786800	-	-	Bthe7332296		PF01551	Peptidase family M23		
2786809	2786985	+	-	Bthe7332297					
2787068	2787826	+	ORF3	Bthe7332298		PF13351, PF13101	Protein of unknown function (DUF4099) , Protein of unknown function (DUF3945)		
2787847	2788071	+	-	Bthe7332299					
2788088	2789665	+	-	Bthe7332300					
2789669	2791057	+	trap	Bthe7332301		PF13155	Toprin-like		
2791054	2791551	+	-	Bthe7332302		PF03703	Bacterial PH domain		
2791574	2792221	+	-	Bthe7332303					
2792191	2792919	+	-	Bthe7332304					
2792934	2793599	+	-	Bthe7332305					
2793624	2794454	+	-	Bthe7332306		PF01551	Peptidase family M23		
2794478	2794759	+	-	Bthe7332307					
2795032	2796555	+	-	Bthe7332308					
2796757	2798844	+	-	Bthe7332309		PF01624	MutS domain I		
2798857	2800413	+	-	Bthe7332310		PF01832, PF01551	Mannosyl-glycoprotein endo-beta-N- acetylglucosamin, Peptidase family M23		

Table 3.9. Continued.

Start	Stop	Strand	Gene	Synonym	Gene Category	P FAM	P FAM_description	TIGRFAM	TIGRFAM_description
2833434	2833727	-	-	Bthe7332353					
2833814	2833927	+	-	Bthe7332354					
2834397	2834528	+	-	Bthe7332355					
2834524	2834541				Bthe733_atR	att			
2834812	2835111	-	-	Bthe7332356	Non-element				
2835214	2836305	-	-	Bthe7332357	Non-element				
2836346	2836747	-	-	Bthe7332358	Non-element				
2836744	2836923	-	-	Bthe7332359	Non-element				
2837142	2838314	-	-	Bthe7332360	Non-element				

Table 3.10. BoCtIn Element Protein Domains and Families.

Start	Stop	Strand	Gene	Synonym	Gene Category	PFAM	PFAM description	TIGRFAM	TIGRFAM description
5364957	5365601	+	-	BACOV/A04670	Non-element				
5365603	5366811	+	-	BACOV/A04671	Non-element				
5366924	5367079	+	-	BACOV/A04672	Non-element				
5367092	5368108	+	-	BACOV/A04673	Non-element				
5368115	5369386	+	-	BACOV/A04674	Non-element				
5369288	5369305			BACOVA_attR	att				
5369526	5369645	-	-	BACOV/A04675					
5369749	5369859	-	-	BACOV/A04676					
5369902	5370030	-	-	BACOV/A04677					
5370102	5370395	+	-	BACOV/A04678					
5370414	5370590	+	-	BACOV/A04679					
5370602	5370928	+	-	BACOV/A04680					
5370945	5371202	+	-	BACOV/A04681					
5371264	5371578	-	-	BACOV/A04682					
5371526	5371729	-	-	BACOV/A04683					
5371883	5372263	+	-	BACOV/A04684		PF00436	Single-strand binding protein family	TIGR00621	ssb: single-stranded DNA-binding protein
5372351	5372632	-	-	BACOV/A04685		PF00216	Bacterial DNA-binding protein		
5372700	5372804	-	-	BACOV/A04686					
5372908	5373519	-	-	BACOV/A04687		PF00239	Resolvase: N terminal domain		
5373537	5373719	+	-	BACOV/A04688					
5373758	5374807	+	-	BACOV/A04689					
5374811	5375281	+	-	BACOV/A04690					
5375278	5375679	+	-	BACOV/A04691					
5375670	5376260	+	-	BACOV/A04692					
5376244	5377095	+	-	BACOV/A04693					
5377184	5377339	+	-	BACOV/A04694					
5377314	5378771	+	-	BACOV/A04695	Mobilization gene	PF01076	Plasmid recombination enzyme		
5378826	5380226	+	-	BACOV/A04696		PF00691	OmpA family		
5380254	5380922	-	-	BACOV/A04697		PF10050	Predicted metal-binding protein (DUF2284)		
5380919	5381122	-	-	BACOV/A04698	Not present in BtCtIn				
5381134	5381871	-	-	BACOV/A04699	Not present in BtCtIn				

Table 3.10. Continued.

Start	Stop	Strand	Gene	Synonym	Gene Category	PFAM	PFAM description	TIGRFAM	TIGRFAM description
5381875	5382375	-	-	BACOV/A04700	Not present in BICIn	PF12821	Threonine/Serine exporter, Thre		
5382372	5383139	-	-	BACOV/A04701	Not present in BICIn	PF06738	Putative threonine/serine exporter		
5383154	5383909	-	btuD	BACOV/A04702	Btu operon	PF00005	ABC transporter		
5383906	5384886	-	btuC	BACOV/A04703	Btu operon	PF01032	FecCD transport family		
5384887	5386026	-	btuF	BACOV/A04704	Btu operon	PF01497	Periplasmic binding protein		
5386031	5388181	-	btuB	BACOV/A04705	BtuB	PF07715	TonB-dependent Receptor Plug Domain		
5388135	5389214	-	btuG	BACOV/A04706		PF05096	Glutamine cyclotransferase		
5389300	5391399	-	-	BACOV/A04707	Not present in BICIn				
5391423	5393168	-	-	BACOV/A04708		PF16820	PKDlike domain		
5393244	5394212	-	-	BACOV/A04709		PF14717	Domain of unknown function (DUF4465)		
5394243	5394356	-	-	BACOV/A04710					
		-		RF00174					
5394619	5394738	+	-	BACOV/A04711					
5394788	5394919	-	-	BACOV/A04712					
5394984	5395466	+	-	BACOV/A04713					
5395447	5395608	+	-	BACOV/A04714					
5395625	5395768	-	-	BACOV/A04715					
5395816	5396358	+	-	BACOV/A04716					
5396355	5396612	+	-	BACOV/A04717					
5396624	5396998	+	-	BACOV/A04718	Not present in BICIn				
5397059	5398747	-	-	BACOV/A04719		PF08401	Domain of unknown function (DUF1738)		
5398753	5399409	-	-	BACOV/A04720					
5399444	5399620	-	-	BACOV/A04721					
5399633	5401189	-	-	BACOV/A04722		PF01832, PF01551	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminyl , Peptidase family M23		
5401202	5403463	-	-	BACOV/A04723		PF01624	MutS domain I		

Table 3.10. Continued

Start	Stop	Strand	Gene	Synonym	Gene Category	PFAM	PFAM description	TIGRFAM	TIGRFAM description
5403491	5405290	-	-	BACOV/A04724					
5405299	5405580	-	-	BACOV/A04725					
5405604	5406434	-	-	BACOV/A04726	PF01551	Peptidase family M23			
5406459	5407124	-	-	BACOV/A04727					
5407139	5407822	-	-	BACOV/A04728					
5407837	5408526	-	-	BACOV/A04729					
5408507	5409004	-	-	BACOV/A04730	PF03703	Bacterial PH domain			
5409001	5410389	-	trap	BACOV/A04731	PF13155	Toprim-like			
5410393	5411970	-	-	BACOV/A04732					
5411987	5412211	-	-	BACOV/A04733					
5412232	5412990	-	ORF3	BACOV/A04734	PF13351, PF13101 (DUF4099) , Protein of unknown function (DUF3945)	Protein of unknown function (DUF4099) , Protein of unknown function (DUF3945)			
5413073	5413183	-	-	BACOV/A04735					
5413258	5413863	+	-	BACOV/A04736	PF01551	Peptidase family M23			
5414012	5414530	-	-	BACOV/A04737	PF02572	ATP:cofnoid adenosyltransferase BtuR/CobO/CobP	TI GR007/08	cobA: cob(l)yrinic acid a-c-di amide adenosyltransf	
5414564	5415298	-	-	BACOV/A04738	PF01902	Diphthamide synthase	TI GR00290	MJ0570_ dom.: MJ0570-related uncharacterized domain	
5415295	5416368	-	-	BACOV/A04739	PF08843	Nucleotidyl transferase AbiEII toxin; Type IV TA			
5416365	5417087	-	-	BACOV/A04741					
5417062	5417184	+	-	BACOV/A04740					
5417306	5417419	-	-	BACOV/A04742					
5417461	5417565	-	-	BACOV/A04743					
5417568	5417792	+	-	BACOV/A04744					
5417850	5418482	+	-	BACOV/A04745					
5418780	5418926	-	-	BACOV/A04746					
5419071	5423084	+	-	BACOV/A04747	PF02384, PF04851	N-6 DNA Methylase , Type III restriction enzyme; res subunit			
5423381	5423503	+	-	BACOV/A04748	Not present in BICIn				

Table 3. 10. Continued

Start	Stop	Strand	Gene	Synonym	Gene Category	PFAM	PFAM description	TIGRFAM	TIGRFAM description
5423483	5423599	-	-	BACOV/A04749	Not present in BICIn				
5423741	5425408	+	-	BACOV/A04750	Not present in BICIn	PF00078, PF13655, PF08388	Reverse transcriptase (RNA- dependent DNA polymera , Nterminal domain of reverse transcriptase , Group II intron; maturase-s-specific domain	TIGR04416	group_II_RT_mat group II intron reverse transcrip
5425392	5426507	+	-	BACOV/A04751					
5425450	5425526	+	-	BACOV/A04752					
5426867	5427622	+	traA	BACOV/A04753	PF13614	AAA domain			
5427630	5427884	+	-	BACOV/A04754					
5428003	5428299	-	-	BACOV/A04755					
5428309	5430042	+	-	BACOV/A04756	PF02195	ParB-like nuclease domain		TIGR00180	parB_part: ParB/RepB/Spo0J family partition protei
5430222	5430497	-	-	BACOV/A04758					
5430397	5430762	+	-	BACOV/A04757					
5430770	5431042	+	-	BACOV/A04759					
5431076	5431450	+	-	BACOV/A04760					
5431540	5431752	-	-	BACOV/A04761					
5431749	5432036	-	-	BACOV/A04762					
5432092	5433306	+	-	BACOV/A04763	PF06250	Protein of unknown function (DUF1016)			
5433414	5435600	-	ORF2	BACOV/A04764	PF02534	Type IV secretion system Conjugative DNA transfer			
5435604	5436350	-	-	BACOV/A04765	PF03118	Bacterial RNA polymerase; alpha chain C terminal			
5436352	5437008	-	-	BACOV/A04766					
5437035	5437559	-	-	BACOV/A04767					
5437556	5438407	-	tran	BACOV/A04768	PF13595	Domain of unknown function (DUF4138)		TIGR03780	Bac_Flav_CT_N: Bacteroides conjugative transposon
5438428	5439231	-	-	BACOV/A04769	PF01555	DNA methylase			
5439243	5440418	-	tram	BACOV/A04770	PF12508	Conjugative transposon; Tram		TIGR03779	Bac_Flav_CT_M: Bacteroides conjugative transposon
5440402	5440908	-	-	BACOV/A04771	PF05869	DNA N6-adenine-methyltransferase (Dam)			
5440895	5441275	-	-	BACOV/A04772					

Table 3. 10. Continued

Start	Stop	Strand	Gene	Synonym	Gene Category	PFAM	PFAM description	TIGRFAM	TIGRFAM description
5441288	5441902	-	traK	BACOV/A04773	Transfer element			TIGR03781	Bac_Flav_CT_K: Bacteroides conjugative transposon
5441906	5442301	-	-	BACOV/A04774					
5442343	5443476	-	-	BACOV/A04775					
5443489	5444256	-	-	BACOV/A04776		PF16464	Domain of unknown function (DUF5045)		
5444275	5445000	-	-	BACOV/A04777					
5444997	5447543	-	-	BACOV/A04778					
5447555	5450218	-	-	BACOV/A04779					
5450345	5450683	+	-	BACOV/A04780					
5450687	5451064	-	traE2	BACOV/A04781		PF13572	Domain of unknown function (DUF4134)		
5451119	5451445	-	traE1	BACOV/A04782		PF13572	Domain of unknown function (DUF4134)		
5451447	5451986	-	-	BACOV/A04783					
5452138	5453025	-	traP	BACOV/A04784		PF13155	Toprim-like		
5453129	5454229	-	-	BACOV/A04785		PF13481	AAA domain		
5454207	5454602	-	-	BACOV/A04786		PF12728	Helix-turn-helix domain	TIGR01764	excise: DNA binding domain; excisionase family
5454782	5455486	-	-	BACOV/A04787					
545549	5455674	-	-	BACOV/A04788					
5455686	5456831	-	-	BACOV/A04789	Integrase	PF00589, PF13102, PF17293	Phage integrase family , Phage integrase SAM-like domain , Aam DNA-binding domain		
5456984	5457163	-	-	BACOV/A04790					
5457044	5457061			BACOVA_atl	at				
5457173	5459872	+	-	BACOV/A04791	Non-element				
5460017	5461282	-	-	BACOV/A04792	Non-element				
5461272	5463575	-	-	BACOV/A04793	Non-element				
5463556	5464830	-	-	BACOV/A04794	Non-element				
5465031	5466089	-	-	BACOV/A04795	Non-element				

Table 3.11. DNA sequencing data generated for Bacteroidetes species used in this study.

Sample	Sequencer	Chemistry	Read length (bp)	Read pairs	Quality filtered read pairs	Scaffolded length	No. scaffolds (≥300 bp)	Max. scaffold length	N ₅₀	% reads mapping
<i>Bacteroides thetaiotaomicron</i> VPI-0633-1	HiSeq5000	TruSeq SBS	2×150	4,930,308	2,917,663	5,790,881	59	546,872	266,759	96.86
<i>Bacteroides thetaiotaomicron</i> VPI-0940-1	HiSeq5000	TruSeq SBS	2×150	5,114,161	3,195,967	6,077,038	81	594,359	179,042	97.08
<i>Bacteroides thetaiotaomicron</i> VPI-J19-343	HiSeq5000	TruSeq SBS	2×150	6,254,917	3,548,491	6,100,520	67	548,715	256,527	96.70
<i>Bacteroides thetaiotaomicron</i> VPI-C11-15	HiSeq2500	HiSeq RS v2	2×100	5,773,495	4,094,789	6,447,831	85	411,259	187,570	92.39
<i>Bacteroides uniformis</i> VPI-61	HiSeq5000	TruSeq SBS	2×150	3,829,029	2,361,428	4,635,717	42	548,204	207,883	97.05
<i>Bacteroides uniformis</i> VPI-909	HiSeq5000	TruSeq SBS	2×150	3,688,548	2,069,267	4,593,792	38	661,053	244,286	96.83
<i>Bacteroides uniformis</i> VPI-C20-25	HiSeq5000	TruSeq SBS	2×150	5,235,435	3,386,621	4,281,266	30	653,661	413,230	97.42
<i>Bacteroides uniformis</i> VPI-C7-17	HiSeq5000	TruSeq SBS	2×150	5,295,039	3,157,326	4,341,168	31	1,098,084	550,304	97.08
<i>Bacteroides uniformis</i> VPI-R5-33	HiSeq5000	TruSeq SBS	2×150	2,410,072	1,434,781	4,259,548	29	526,344	322,465	89.82
<i>Bacteroides nordi</i> WAL7936*	HiSeq5000	TruSeq SBS	2×150	2,688,636	1,549,589	5,452,512	47	491,859	284,651	96.73
<i>Bacteroides dorei</i> VPI-227 ^o	HiSeq5000	TruSeq SBS	2×150	4,002,356	2,343,944	5,356,240	67	569,746	218,716	96.91
<i>Bacteroides dorei</i> VPI-3776A ^o	HiSeq5000	TruSeq SBS	2×150	2,432,797	1,456,979	5,233,865	56	485,258	219,454	96.72
<i>Bacteroides dorei</i> VPI-6598B	HiSeq2500	HiSeq RS v2	2×100	5,005,785	4,178,843	5,915,454	94	502,624	146,581	94.98

* Originally deposited in the VPI strain collection as *B. uniformis*

^o Originally deposited in VPI strain collection as *B. vulgatus*

Table 3.12. Bacterial Strains Used in this Study.

Strain name	Short name in text	Notes	Source and/or reference
<i>E. coli</i> S17-1Apir		F- <i>recA</i> 1 <i>pro</i> -82 <i>endA</i> 1 <i>thiE</i> 1 <i>hsdR</i> 17 <i>creC</i> 510 <i>RP4-2</i> (<i>Tc</i> :: <i>Mu</i> -1, <i>Km</i> :: <i>Tn7</i>)Apir, Amp ^S	Simon et al. 1983
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i>	<i>B. thetaiotaomicron</i>	wildtype deletion background, Δ <i>B</i> T2275, <i>FUGRR</i> , Em ^S , Tet ^S	Koropatkina et al. 2008
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> att1::p <i>NBU2-bla-emG</i> b	<i>B. thetaiotaomicron</i> Em ^R		
<i>B. ovatus</i> ATCC 8483	<i>B. ovatus</i>	wildtype strain Em ^S , Tet ^S	
<i>B. thetaiotaomicron</i> 7330	<i>B. thetaiotaomicron</i> 7330	wildtype strain Em ^S , Tet ^S	
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2 Δ <i>locus</i> 3	<i>B. thetaiotaomicron</i> ΔΔΔ		Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2 Δ <i>locus</i> 3 att1::p <i>NBU2-bla-emG</i> b	<i>B. thetaiotaomicron</i> ΔΔΔ Em ^R		This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2 Δ <i>locus</i> 3 att2::p <i>NBU2-bla-teIQ</i>	<i>B. thetaiotaomicron</i> ΔΔΔ Tet ^R		This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>B</i> T1963 Δ <i>B</i> T2094		<i>btuB</i> from Locus 1 and <i>btuFCDG</i> from Locus 2 and 3	Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 3		Locus 2 <i>btuB</i> <i>FCDG</i> only	Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2		Locus 3 <i>btuB</i> <i>FCDG</i> only	Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2 Δ <i>locus</i> 3 att1::p <i>NBU2-bla-emG</i> b + <i>B</i> oC <i>Tn</i>	<i>B. thetaiotaomicron</i> ΔΔΔ Em ^R + <i>B</i> oC <i>Tn</i>	Transconjugant from parental strain	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2 Δ <i>locus</i> 3 att1::p <i>NBU2-bla-emG</i> b + <i>B</i> oC <i>Tn</i>	<i>B. thetaiotaomicron</i> ΔΔΔ Em ^R + <i>B</i> oC <i>Tn</i>	Transconjugant from parental strain	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2 Δ <i>locus</i> 3 att2::p <i>NBU2-bla-teIQ</i> + <i>B</i> oC <i>Tn</i>	<i>B. thetaiotaomicron</i> ΔΔΔ Tet ^R + <i>B</i> oC <i>Tn</i>	Transconjugant from transconjugant strain	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ <i>tdk</i> Δ <i>locus</i> 1 Δ <i>locus</i> 2 Δ <i>locus</i> 3 att2::p <i>NBU2-bla-teIQ</i> + <i>B</i> oC <i>Tn</i>	<i>B. thetaiotaomicron</i> ΔΔΔ Tet ^R + <i>B</i> oC <i>Tn</i>	Transconjugant from transconjugant strain	This study

Table 3.12. Continued.

Strain name	Short name in text	Notes	Source and/or reference
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 Δ locus3 + BoCTn	<i>B. thetaiotaomicron</i> $\Delta\Delta\Delta$ + BoCTn	Unmarked transconjugant strain	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 Δ locus3 + BtCTn	<i>B. thetaiotaomicron</i> $\Delta\Delta\Delta$ + BtCTn	Unmarked transconjugant strain	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ B71953 Δ B72094 attN::pNBU2_tet _{BC09}	<i>B. thetaiotaomicron</i> Δ B71953 Δ B72094 BC	Barcoded strain for competitions	Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus3 att2::pNBU2_tet _{BC01}	<i>B. thetaiotaomicron</i> Δ locus1 Δ locus3 BC	Barcoded strain for competitions	Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 att2::pNBU2_tet _{BC08}	<i>B. thetaiotaomicron</i> Δ locus1 Δ locus2 BC	Barcoded strain for competitions	Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 Δ locus3 att1::pNBU2_tet _{BC14}	<i>B. thetaiotaomicron</i> $\Delta\Delta\Delta$ BC	Barcoded strain for competitions	Degnan et al. 2014
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 Δ locus3 + BoCTn att2::pNBU2_tet _{BC18}	<i>B. thetaiotaomicron</i> $\Delta\Delta\Delta$ + BoCTn BC	Barcoded strain for competitions	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 Δ locus3 + BtCTn att2::pNBU2_tet _{BC06}	<i>B. thetaiotaomicron</i> $\Delta\Delta\Delta$ + BtCTn BC	Barcoded strain for competitions	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 Δ locus3 + BoCTn_tetQ	<i>B. thetaiotaomicron</i> $\Delta\Delta\Delta$ + BoCTn_Tet ^R	TetQ in the CTn near the cornoid transport locus	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ cdk Δ locus1 Δ locus2 Δ locus3 + BtCTn_tetQ	<i>B. thetaiotaomicron</i> $\Delta\Delta\Delta$ + BtCTn_Tet ^R	TetQ in the CTn near the cornoid transport locus	This study
<i>B. uniformis</i> ATCC 8492 attN::pNBU2_erm			This study
<i>B. vulgatus</i> ATCC 8482 attN::pNBU2_erm			This study
<i>B. fragilis</i> NCTC 9343 attN::pNBU2_erm			This study
<i>B. caecae</i> ATCC 43185 attN::pNBU2_erm			This study
<i>B. uniformis</i> ATCC 8492 attN::pNBU2_erm + BoCTn_tetQ		TetQ in the CTn near the cornoid transport locus	This study
<i>B. uniformis</i> ATCC 8492 attN::pNBU2_erm + BtCTn_tetQ		TetQ in the CTn near the cornoid transport locus	This study
<i>B. vulgatus</i> ATCC 8482 attN::pNBU2_erm + BoCTn_tetQ		TetQ in the CTn near the cornoid transport locus	This study
<i>B. vulgatus</i> ATCC 8482 attN::pNBU2_erm + BtCTn_tetQ		TetQ in the CTn near the cornoid transport locus	This study

Table 3.12. Continued.

Strain name	Short name in text	Notes	Source and/or reference
<i>B. thetaiotaomicron</i> VPI-5482 Δ tdk Δ locus1 Δ locus2 Δ locus3 att1::pNBU2-b1a-emGcb + BoCTn_tetQ isolate 1	Mouse-isolated BoCTn TC-1	Transconjugant isolated from the mouse gut	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ tdk Δ locus1 Δ locus2 Δ locus3 att1::pNBU2-b1a-emGcb + BoCTn_tetQ isolate 2	Mouse-isolated BoCTn TC-2	Transconjugant isolated from the mouse gut	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ tdk Δ locus1 Δ locus2 Δ locus3 att1::pNBU2-b1a-emGcb + BiCTn_tetQ isolate 1	Mouse-isolated BiCTn TC-1	Transconjugant isolated from the mouse gut	This study
<i>B. thetaiotaomicron</i> VPI-5482 Δ tdk Δ locus1 Δ locus2 Δ locus3 att1::pNBU2-b1a-emGcb + BiCTn_tetQ isolate 2	Mouse-isolated BiCTn TC-2	Transconjugant isolated from the mouse gut	This study

AmpR, ampicillin resistant; EmrR, erythromycin resistant; FudRR, FudR resistant; FudRS, FudR sensitive

Table 3.13. Plasmids Used in this Study.

Plasmid name	Notes	Source and/or reference
pXchange_tdk	Construct carrying cloned tdk (BT2275) for counter-selection, AmpR, ErmR, FUDRS	Koropatkin <i>et al.</i> 2008
pXchange_BT4111/BT4112	Construct with intergenic region between BT4111 and BT4112	This work
pXchange_tdk_BocTn_TetQ	TetQ gene inserted into SOE product of BACOVA4684-BACOVA4685 intergenic region	This work
pXchange_tdk_BtCTn_TetQ	TetQ gene inserted into SOE product of Bthe2346-Bthe2347 intergenic region	This work
pNBU2-bla-ermGb	Integrates cloned fragments to NBU2 att1 and/or att2 sites, AmpR ErmR	Koropatkin <i>et al.</i> 2008
pNBU2-bla-tetQ	Integrates cloned fragments to NBU2 att1 and/or att2 sites, AmpR TetR	Martens <i>et al.</i> 2008
pNBU2_tet_BC01	Barcoded version of pNBU2-bla-tetQ	Martens <i>et al.</i> 2008
pNBU2_tet_BC04	Barcoded version of pNBU2-bla-tetQ	Degnan <i>et al.</i> 2014
pNBU2_tet_BC06	Barcoded version of pNBU2-bla-tetQ	Degnan <i>et al.</i> 2014
pNBU2_tet_BC08	Barcoded version of pNBU2-bla-tetQ	Degnan <i>et al.</i> 2014
pNBU2_tet_BC09	Barcoded version of pNBU2-bla-tetQ	Costliow and Degnan 2017
pNBU2_tet_BC14	Barcoded version of pNBU2-bla-tetQ	Martens <i>et al.</i> 2008

Table 3.14. Primers Used in this Study.

Primer name	Sequence (5' - 3')*	Fragment Length	Source and/or reference
Primers for general amplification of <i>kluB</i> Family N alleles			
BT1953_general_F1	CTGGTATATCAATTCCAACCG	755	This work
BT1953_general_R1	ATGATCCAGTCGTGATG		This work
Primers for detection of BoCTn-B12 from <i>B. ovaus</i> ATCC 8483			
BACOV/A04707_F1	AACGATGCCAGATTTTCATCC	953	This work
BACOV/A04707_R1	CCGGTTTGACACATTTACC		This work
Primers for detection of BkCTn-B12 from <i>B. thetaiotaomicron</i> 7330			
Bthe7332318_R1	ATGCACGCTGTCTATGTCG	866	This work
Bthe7332318_F1	CCACCTGTTTCGTTCTTCG		This work
Primers for cloning of <i>teq</i> gene			
teq_pNBU2_F_BamHI	GCCGGATCCACGCGTTATCTCCTTAAC	2079	Campbell <i>et al.</i> , 2020
teq_pNBU2_R_SpeI	GGCACTAGTTTATTTTGTATGATGACATTTGTTTGGAAATG		Campbell <i>et al.</i> , 2020
Primers for insertion of <i>teq</i> into BoCTn-B12 through allelic exchange			
RF_BACOV/A4685_NotI	ATCGCGCCGACGACACCGCGGAAAGTATGGCC	1017	This work
LF_BACOV/A4685_SOE	ACTCATTTACGGGTGGTTGGACCATGGATTCATTCACCTAGTCTGTACTTTTATCTCTACTTGTTC		This work
RF_BACOV/A4684_SOE	GAAACAAGTAGAGTATAAAGTACAGACTAGTGAATGATGATTCATGGTGCAACACACCGTAATGAGT		This work
LF_BACOV/A4684_Sall	ACTGTGACGTCCCTATTCCTTTCTGTGCGGACGGG	929	This work

Table 3.14. Continued.

Primer name	Sequence (5' - 3')*	Fragment Length	Source and/or reference
PNBU2_tet_BC qPCR primers			
UNIV-R	CACAAATATGACCAACAAGAAATCC		Martens <i>et al.</i> 2008
BC01-F	ATGTGCCAATTGTCACTTCTCA		Martens <i>et al.</i> 2008
BC04-F	CTCCATPAAAGCGCATACCGACTA		Degnan <i>et al.</i> 2014
BC06-F	GATTACGGCGTATAGATTGTTGT		Degnan <i>et al.</i> 2014
BC08-F	CTTCCTGCCCGGCTPAAAAACAGCA		Degnan <i>et al.</i> 2014
BC09-F	TCAATTCGGGGACTGGGCTTAGA		Costilow and Degnan 2017
BC14-F	GGCAGCGCAATTCTCATCTAATG		Martens <i>et al.</i> 2008
Excision qPCR primers^Δ			
B_the_3F	GAGGGTGTCTATTTCCGAAG	98	Goodman <i>et al.</i> 2009
B_the_3R	GTTCCCTGATCCAGTGTGTTGG		Goodman <i>et al.</i> 2009
attB_F3	TTGAAACGGAACAAACACACG		This work
attB_R3	AGACCGGTTACTTATTAATCTGCG		This work
attC_F1	GTAAGAGAAAGAAAGTAGTGC		This work
attC_R1	ATTGGGTGTGTTTGCAATTTTG		This work

*Restriction sites underlined

^ΔattB = attB_F3/attB_R3 213bp; attC = attC_F1/attC_R2 200bp; attL = attC_R1/attB_R3 80bp; attR = attC_F1/attB_F3 333bp

3.8 **References**

1. Allen RH, Stabler SP. 2008. Identification and quantitation of cobalamin and cobalamin analogues in human feces. *American Journal of Clinical Nutrition*. 87(5):1324–35.
2. Alvarez-Martinez CE, Christie PJ. 2009. Biological diversity of prokaryotic type IV secretion systems. *MMBR*. 73(4):775–808.
3. Bacic M, Parker AC, Stagg J, Whitley HP, Wells WG, et al. 2005. Genetic and structural analysis of the *Bacteroides* conjugative transposon CTn341. *Journal of Bacteriology*. 187(8):2858–69.
4. Bertram J, Strätz M, Dürre P. 1991. Natural transfer of conjugative transposon Tn916 between gram-positive and gram-negative bacteria. *Journal of Bacteriology*. 173(2):443–48.
5. Bookout AL, Cummins CL, Kramer MF, Pesola JM, Mangelsdorf DJ. 2006. High-throughput real-time quantitative reverse transcription PCR. *Current Protocols in Molecular Biology*, pp. 1–28.
6. Buckwold SL, Shoemaker NB, Sears CL, Franco AA. 2007. Identification and characterization of conjugative transposons CTn86 and CTn9343 in *Bacteroides fragilis* strains. *Applied and Environmental Microbiology*. 73(1):53–63.
7. Buffie CG, Pamer EG. 2013. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat Rev Immunol*. 13(11):790–801.
8. Burrus V, Pavlovic G, Decaris B, Guédon G. 2002. Conjugative transposons: The tip of the iceberg. *Molecular Microbiology*. 46(3):601–10.
9. Campbell DE, Ly LK, Ridlon JM, Hsiao A, Whitaker RJ, Degnan PH. 2020. Infection with *Bacteroides* phage BV01 alters the host transcriptome and bile acid metabolism in a common human gut microbe. *Cell Reports*. 32(11):108142.
10. Carpena N, Manning KA, Dokland T, Marina A, Penadés JR. 2016. Convergent evolution of pathogenicity islands in helper cos phage interference. *Phil. Trans. R. Soc. B*. 371(1707):20150505.
11. Casjens S, Winn-Stapley DA, Gilcrease EB, Morona R, Kühlewein C, et al. 2004. The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *Journal of Molecular Biology*. 339(2):379–94.

12. Chimento DP, Mohanty AK, Kadner RJ, Wiener MC. 2003. Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nature structural biology*. 10(5):394–401.
13. Clark AJ, Inwood W, Cloutier T, Dhillon TS. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *Journal of Molecular Biology*. 311(4):657–79.
14. Costliow ZA, Degnan PH. 2017. Thiamine acquisition strategies impact metabolism and competition in the gut microbe *Bacteroides thetaiotaomicron*. *msystems*. 2(5):00116-17.
15. Coyne MJ, Roelofs KG, Comstock LE. 2016. Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC Genomics*. 17(1):58.
16. Coyne MJ, Zitomersky NL, McGuire M, Zitomersky L, Earl AM, Comstock E. 2014. Evidence of extensive DNA transfer between Bacteroidales species within the human gut. *mBio*. 5(3):e01305-14.
17. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. pp. 1394–1403.
18. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. In *Engineering and Analyzing Multicellular Systems*, ed L Sun, W Shou. 1151:165–88. New York, NY: Springer New York.
19. Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL. 2014. Human gut microbes use multiple transporters to distinguish vitamin B₁₂ analogs and compete in the gut. *Cell Host & Microbe*. 15(1):47–57.
20. Degnan PH, Moran NA. 2008. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Molecular Ecology*. 17(3):916–29.
21. Degnan PH, Taga ME, Goodman AL. 2014. Vitamin B₁₂ as a modulator of gut microbial ecology. *Cell Metabolism*. 20(5):769–78.
22. Di Girolamo PM, Bradbeer C. 1971. Transport of vitamin B₁₂ in *Escherichia coli*. *Journal of bacteriology*. 106(3):745–50.

23. Duerkop BA, Clements CV, Rollins D, Rodrigues JLM, Hooper LV. 2012. A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proceedings of the National Academy of Sciences*. 109(43):17621–26.
24. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32(5):1792–97.
25. Enright AJ. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*. 30(7):1575–84.
26. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, et al. 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host and Microbe*. 6(3):279–89.
27. Gupta A, Vlamakis H, Shoemaker N, Salyers AA. 2003. A new *Bacteroides* conjugative transposon that carries an *ermB* Gene. *Applied and Environmental Microbiology*. 69(11):6455–63.
28. Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*. 464(7290):908–12.
29. Hehemann J-H, Kelly AG, Pudlo NA, Martens EC, Boraston AB. 2012. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences*. 109(48):19786–91.
30. Holdeman LV, Moore WEC, Virginia Polytechnic Institute and State University., Anaerobe Laboratory. 1977. Anaerobe laboratory manual. Blacksburg, VA: V.P.I. Anaerobe Laboratory.
31. Hong SN, Rhee PL. 2014. Unraveling the ties between irritable bowel syndrome and intestinal microbiota. *World Journal of Gastroenterology*. 20(10):2470–81.
32. Husain F, Veeranagouda Y, Boente R, Tang K, Mulato G, Wexler HM. 2014. The Ellis Island Effect: A novel mobile element in a multi-drug resistant *Bacteroides fragilis* clinical isolate includes a mosaic of resistance genes from Gram-positive bacteria. *Mobile Genetic Elements*. 4(4):e29801.
33. Jacobson A, Lam L, Rajendram M, Tamburini F, Honeycutt J, et al. 2018. A gut commensal-produced metabolite mediates colonization resistance to *Salmonella* infection. *Cell Host & Microbe*. 24(2):296-307.e7.
34. John GK, Mullin GE. 2016. The gut microbiome and obesity. *Current oncology reports*. 18(7):45.

35. Johnson CM, Grossman AD. 2016. The composition of the cell envelope affects conjugation in *Bacillus subtilis*. *J. Bacteriol.* 198(8):1241–49.
36. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *Journal of Molecular Biology.* 299(1):27–51.
37. Kamada N, Kim Y-G, Sham HP, Vallance BA, Puente JL, et al. 2012. Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science.* 336(6086):1325–29.
38. Kenley JS, Leighton M, Bradbeer C. 1978. Transport of vitamin B₁₂ in *Escherichia coli*: corrinoid specificity of the outer membrane receptor. *The Journal of Biological Chemistry.* 253(5):1341–46.
39. Koropatkin NM, Martens EC, Gordon JI, Smith TJ. 2008. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure.* 16(7):1105–15.
40. Litvak Y, Bäumlér AJ. 2019. The founder hypothesis: A basis for microbiota resistance, diversity in taxa carriage, and colonization resistance against pathogens. *PLoS Pathog.* 15(2):e1007563.
41. Martens EC, Chiang HC, Gordon JI. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host and Microbe.* 4(5):447–57.
42. Meehan CJ, Beiko RG. 2012. Lateral gene transfer of an ABC transporter complex between major constituents of the human gut microbiome. *BMC Microbiol.* 12:248.
43. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees.
44. O'Brien A, Newland J, Miller S, Holmes R, Smith H, Formal S. 1984. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science.* 226(4675):694–96.
45. Quesada-Gómez C. 2011. *Bacteroides* mobilizable and conjugative genetic elements: antibiotic resistance among clinical isolates. *Revista española de quimioterapia : publicación oficial de la Sociedad Española de Quimioterapia.* 24(4):184–90.
46. Rankin DJ, Rocha EPC, Brown SP. 2011. What traits are carried on mobile genetic elements, and why? *Heredity.* 106(1):1–10.

47. Rice P, Longden, L, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*. 16(6):276–77.
48. Roth J, Lawrence J, Bobik T. 1996. Cobalamin (Coenzyme B₁₂): synthesis and biological significance. *Annual Review of Microbiology*. 50(1):137–81.
49. Rowland I, Gibson G, Heinken A, Scott K, Swann J, et al. 2018. Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr*. 57(1):1–24.
50. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, et al. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research*. 37(6):e45–e45.
51. Salyers A, Gupta A, Wang Y. 2004. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology*. 12(9):412–16.
52. Salyers a a, Shoemaker NB. 1994. Broad-host-range gene transfer plasmids and conjugative transposons. *FEMS Microbiology Ecology*. 15(1–2):15–22.
53. Scott KP. 2002. The role of conjugative transposons in spreading antibiotic resistance between bacteria that inhabit the gastrointestinal tract. *Cellular and Molecular Life Sciences (CMLS)*. 59(12):2071–82.
54. Sequeira RP, McDonald JAK, Marchesi JR, Clarke TB. 2020. Commensal *Bacteroidetes* protect against *Klebsiella pneumoniae* colonization and transmission through IL-36 signaling. *Nat Microbiol*. 5(2):304–13.
55. Shannon P. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*. 13(11):2498–2504.
56. Shkoporov AN, Khokhlova EV, Kulagina EV, Smeianov VV, Kuchmiy AA, et al. 2013. Analysis of a novel 8.9kb cryptic plasmid from *Bacteroides uniformis*, its long-term stability and spread within human microbiota. *Plasmid*. 69(2):146–59.
57. Shoemaker NB, Vlamakis H, Hayes K, Salyers AA. 2001. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Appl. Environ. Microbiol*. 67(2):561–68.
58. Sommer F, Bäckhed F. 2013. The gut microbiota--masters of host development and physiology. *Nature reviews. Microbiology*. 11(4):227–38.

59. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22(21):2688–90.
60. Stecher B, Denzler R, Maier L, Bernet F, Sanders MJ, et al. 2012. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences*. 109(4):1269–74.
61. The Human Microbiome Project. 2012. Structure, function and diversity of the healthy human microbiome. *Nature*. 486(7402):207–14.
62. Top EM, Springael D. 2003. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Current Opinion in Biotechnology*. 14(3):262–69.
63. Tribble GD, Parker AC, Smith CJ. 1999. Genetic structure and transcriptional analysis of a mobilizable, antibiotic resistance transposon from *Bacteroides*. *Plasmid*. 42(1):1–12.
64. Trieu-Cuot P, Carlier C, Courvalin P. 1988. Conjugative plasmid transfer from *Enterococcus faecalis* to *Escherichia coli*. *Journal of Bacteriology*. 170(9):4388–91.
65. Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML. 2000. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *BioTechniques*. 29(1):52–54.
66. van Dongen S. A Cluster Algorithm for Graphs. Centrum voor Wiskunde en Informatica.
67. Varel VH, Bryant MP. 1974. Nutritional features of *Bacteroides fragilis* subsp. *fragilis*. *Applied microbiology*. 28(2):251–57.
68. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, et al. 2019. Interplay between the human gut microbiome and host metabolism. *Nat Commun*. 10(1):4505.
69. Waldor MK, Mekalanos JJ. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*. 272(5270):1910–14.
70. Wang GR, Shoemaker NB, Jeters RT, Salyers AA. 2011. CTn12256, a chimeric *Bacteroides* conjugative transposon that consists of two independently active mobile elements. *Plasmid*. 66(2):93–105.
71. Wang J, Shoemaker NB, Wang GR, Salyers AA. 2000. Characterization of a *Bacteroides* mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene. *Journal of Bacteriology*. 182(12):3559–71.

72. Wang Y, Shoemaker NB, Salyers AA. 2004. Regulation of a *Bacteroides* operon that controls excision and transfer of the conjugative transposon CTnDOT. *Journal of Bacteriology*. 186(9):2548–57.
73. Welch RA, Jones KR, Macrina FL. 1979. Transferable lincosamide-macrolide resistance in *Bacteroides*. *Plasmid*. 2(2):261–68.
74. Wesslund NA, Wang GR, Song B, Shoemaker NB, Salyers AA. 2007. Integration and excision of a newly discovered *Bacteroides* conjugative transposon, CTnBST. *Journal of Bacteriology*. 189(3):1072–82.
75. Wexler AG, Schofield WB, Degnan PH, Folta-Stogniew E, Barry NA, Goodman AL. 2018. Human gut *Bacteroides* capture vitamin B₁₂ via cell surface-exposed lipoproteins. *eLife*. 7:1–20.
76. Whittle G, Shoemaker NB, Salyers AA. 2002. The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cellular and Molecular Life Sciences*. 59(12):2044–54.
77. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, et al. 2009. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proceedings of the National Academy of Sciences*. 106(10):3698–3703.
78. Yi S, Seth EC, Men YJ, Stabler SP, Allen RH, et al. 2012. Versatility in corrinoid salvaging and remodeling pathways supports corrinoid-dependent metabolism in *Dehalococcoides mccartyi*. *Applied and Environmental Microbiology*. 78(21):7745–52.
79. Zhang Y, Rodionov DA, Gelfand MS, Gladyshev VN. 2009. Comparative genomic analyses of nickel, cobalt and vitamin B₁₂ utilization. *BMC genomics*. 10:78.
80. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, et al. 2019. Adaptive evolution within gut microbiomes of healthy people. *Cell Host & Microbe*. 25(5):656-667.e8.

CHAPTER 4:

Conclusions and future directions

4.1 Conclusions and future directions

The metal utilization of some metalloproteins is predicted to have evolved over time (1, 2, 6). As most metalloprotein metal loading is dictated by exogenous concentrations of metals, it follows that metal availability might drive the neofunctionalization of metalloproteins (2, 3, 9). However, this has been difficult to directly evaluate due to the lack of a system with known evolutionary starting and ending points where relevant metal limitation could be enacted. The studies presented herein with the staphylococcal superoxide dismutases (SODs) represent the first time a study of metalloprotein neofunctionalization has been performed in the context of a native organism experiencing physiologically relevant stresses.

Staphylococcus aureus encodes two SODs, Mn-dependent SodA and cambialistic (maintains activity with either Fe or Mn) SodM. SodM is more similar to SodA in *S. aureus* than to any other SOD and is predicted to have arisen by first a duplication of SodA and subsequent neofunctionalization to SodM (2, 10). The staphylococcal SODs therefore represent a promising tool to enable studies into how metal limitation could lead to metalloenzyme neofunctionalization as they are very similar, evolutionarily related, and can be studied within their native organism (*S. aureus*) experiencing physiologically relevant stressors (metal limitation and oxidative stress). I demonstrated that even a single amino acid

swapping mutation in SodA can lead to significant increases in the ability of *S. aureus* to resist metal limitation and oxidative stress both in culture and during infection (Figures 2.2, 2.9, 2.19).

Interestingly, these studies also revealed that apparent increases in Fe-dependent activity are not the only predictor of success, as some SodA amino acid swapped protein variants with increased *in vitro* Fe-dependent activity did not lead to an increase in the ability of *S. aureus* to resist these host defenses (Table 2.1, Figure 2.9). While the studies presented herein assessed SOD activity of SodA variants in *S. aureus* that had previously been assessed *in vitro*, these studies did not uncover any statistically significant differences in SOD activity in these variants as compared to wild-type SodA (Figure 2.16). However, the current results are underpowered. Additionally, these studies focused on the effect of swapping mutations in three amino acids previously identified to play a role in dictating metal specificity in these enzymes; however, there are other amino acids that differ between these SODs that warrant further investigation (2). There was an incomplete conversion from SodA to SodM metal specificity and perhaps these other differing residues might push the SodA variant metal specificity even closer to that of SodM. Also, as *S. aureus* was able to make the switch from SodA to SodM, perhaps these other differing residues helped enable that change by stabilizing any detrimental mutations, such as Leu160Phe. Finally, the SodA protein variant that consistently led to a benefit for *S. aureus* (Gly159Leu) still maintains a high degree of activity with Mn (Table 2.1).

Therefore, further investigations into the balance between gain of Fe-dependent activity and loss of Mn-dependent activity would aid in the understanding of the pressures on the neofunctionalization from Mn dependency to cambialism.

When the quantity of peptide containing AA160 was assessed in these studies, the level of SodA protein with the Leu160Phe mutation was found to be the same whether or not WT SodA was present (Figure 2.18). This indicates that the detrimental nature of Leu160Phe mutations is not due to destabilization and subsequent loss of the protein. Perhaps instead Leu160Phe mutations lead to an inability to properly acquire the correct metal leading to presence of the protein but loss of SOD activity. These results also support that even though Leu160Phe mutations were tolerable and led to increased Fe-dependent activity in combination with Gly159Leu *in vitro*, the biological path is much more constrained. Broadly, the path from one metal dependency to another is likely more permissible biochemically, but the host organism's biology likely places significant restrictions evolutionarily. Further investigations in this area are required to fully address these questions.

Metal loading and selectivity of most metalloproteins is thought to be driven by metal availability (4, 5, 8, 11). This work on the staphylococcal SODs showed that when swapping mutations Gly159Leu or Phe19Ile-Gly159Leu were present, these SodA protein variants demonstrated increased loading with Fe even when wild-type SodA was loading with Mn (Figure 2.16). This suggests that rather than being based largely on the metal content of the environment, there

might be certain residues that help dictate metal selectivity in these SODs.

Future experiments should be performed to elucidate this further and investigate the effect of this in other metalloproteins.

The investigations contained within this work focus on the predicted evolution from SodA to SodM, or from Mn-dependency to cambialism, with a focus on how metal availability could act as a driver for these changes. The Fe-dependent activity of SodM is important for the pathogenesis of *S. aureus* and this work shows that small incremental changes in metal dependency that include an increase in Fe-dependent activity are beneficial. However, if Fe-dependent activity is crucial to this pathogen, why did SodM stop at cambialism, why not evolve further to a fully Fe-dependent enzyme? This leads to the hypothesis that cambialism itself, not just the ability to use Fe, must be beneficial to *S. aureus*. Future experiments evaluating the ability of *S. aureus* to resist Mn limitation and oxidative stress while encoding a fully Fe-dependent SOD, a different cambialistic SOD, or SodM would further understanding of the evolution and importance of SodM and whether Fe-dependency is the most important factor or cambialism itself enables *S. aureus* to better adapt to the changing conditions of the host during infection.

When *S. aureus* is grown in culture, SodA is important when Mn is available and SodM is important when Mn is limited (7). *In vivo*, only SodM is crucial for infection (7). This leads to the question of why Mn-dependent SodA is maintained in *S. aureus*, why not evolve SodA itself to be cambialistic? The work presented here begins to suggest that biological factors may be a main constraint

in this evolution. For example, detrimental mutations in SodA may be especially problematic as they cannot be stabilized by the presence of a second SOD or some SodA variants cannot properly acquire Fe when they need to. One way to begin to address this question, was to construct a *S. aureus* strain with SodA replaced with SodM. However, preliminary results indicate that this chromosomal integration is not biologically favored. Future experiments to address this question might be focused on attempting to integrate the SodA protein variants utilized in the *sodA*² strains that have increased Fe-dependent activity or another cambialistic enzyme. Additionally, more attempts at alternatives to chromosomal integration should be investigated.

Together, the works provided herein establish metal availability as a driver of metalloenzyme evolution and suggest that metal loading in SODs might also be driven by the protein itself, not just metal availability. While these works focused on the staphylococcal SODs, it is likely that these concepts extend beyond SODs and beyond *S. aureus*, encouraging future experiments on other enzymes and other organisms.

4.2 References

1. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM. 2008. Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem*. 13(8):1205–18.
2. Barwinska-Sendra A, Garcia YM, Sendra KM, Baslé A, Mackenzie ES, et al. 2020. An evolutionary path to altered cofactor specificity in a metalloenzyme. *Nat Commun*. 11(1):2738.
3. Beyer WF, Fridovich I. 1991. *In vivo* competition between iron and manganese for occupancy of the active site region of the manganese-superoxide dismutase of *Escherichia coli*. *Journal of Biological Chemistry*. 266(1):303–8.
4. Capdevila DA, Edmonds KA, Giedroc DP. 2017. Metallochaperones and metalloregulation in bacteria. *Essays In Biochemistry*. 61(2):177–200.
5. Culotta VC, Yang M, O'Halloran TV. 2006. Activation of superoxide dismutases: Putting the metal to the pedal. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 1763(7):747–58.
6. Dupont CL, Butcher A, Valas RE, Bourne PE, Caetano-Anolles G. 2010. History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proceedings of the National Academy of Sciences*. 107(23):10567–72.
7. Garcia YM, Barwinska-Sendra A, Tarrant E, Skaar EP, Waldron KJ, Kehl-Fie TE. 2017. A superoxide dismutase capable of functioning with iron or manganese promotes the resistance of *Staphylococcus aureus* to calprotectin and nutritional immunity. *PLOS Pathogens*. 13(1):e1006125.
8. Imlay JA. 2015. Common mechanisms of bacterial metal homeostasis. In *Trace Metals and Infectious Diseases*, ed JO Nriagu, EP Skaar. Cambridge (MA): MIT Press.
9. Martin ME, Byers BR, Olson MOJ, Salin ML, Arceneaux JE, Tolbert C. 1986. A *Streptococcus mutans* superoxide dismutase that is active with either manganese or iron as a cofactor. *Journal of Biological Chemistry*. 261(20):9361–67.
10. Valderas MW, Hart ME. 2001. Identification and characterization of a second superoxide dismutase gene (*sodM*) from *Staphylococcus aureus*. *Journal of Bacteriology*. 183(11):3399–3407.
11. Whittaker JW. 2010. Metal uptake by manganese superoxide dismutase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 1804(2):298–307.