Technische Universität München
TUM School of Life Sciences

Distinct genomic and functional profiles of *Staphylococcus aureus* strains from atopic dermatitis patients and healthy individuals

Zhongjie Wang

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Romana Gerner

Prüfende der Dissertation:

1. Prof. Dr. Michael Schloter
2. Prof. Dr. Claudia Traidl-Hoffmann

Die Dissertation wurde am 22.05.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 21.10.2024 angenommen.

I make no apologies for putting microorganisms on a pedestal above all other living things, for if the last blue whale is choked to death on the last panda, it would be disastrous but not the end of the world. But if we accidentally poisoned the last two species of ammonia oxidizers, that would be another matter.

It could be happening now and we wouldn't even know…

Tom Curtis

Table of Contents

Publications

List of publications:

1. **Wang, Z.**, Peng X., Hülpüsch C., Mirzaei M.K., Reiger M., Traidl-Hoffmann C., Deng L., Schloter M. 2024. Distinct prophage gene profiles of *Staphylococcus aureus* strains from atopic dermatitis patients and healthy individuals. *Microbiology Spectrum* 12:e00915-24. https://doi.org/10.1128/spectrum.00915-24
   (First author, published)

   Contributions:
   - Conceptualization and data collection
   - Bioinformatic analyses
   - Manuscript writing

2. **Wang, Z.**, Hülpüsch C., Foesel B.U., Traidl-Hoffmann C., Reiger M., Schloter M. 2024. Genomic and functional divergence of *Staphylococcus aureus* strains from atopic dermatitis patients and healthy individuals: insights from global and local scales. *Microbiology Spectrum* 12:e00571-24. https://doi.org/10.1128/spectrum.00571-24
   (First author, published)

   Contributions:
   - Conceptualization and data collection
   - Bioinformatic analyses
   - Manuscript writing

3. **Wang, Z.**, Hülpüsch C., Traidl-Hoffmann C., Reiger M. & Schloter M. 2024. Understanding the Role of *Staphylococcus aureus* in Atopic Dermatitis: Strain Diversity, Microevolution, and Prophage Influences. *Frontiers in Medicine 11:1480257.* https://doi.org/10.3389/fmed.2024.1480257
   (First author, published, Mini Review)

   Contributions:
   - Conceptualization
   - Manuscript writing

4. **Wang, Z.**, Hülpüsch, C., Schwierzeck, V., Alharbi, S.A., Reiger, M., Traidl-Hoffmann, C., Schloter M., Foesel B.U. 2022. Complete and Draft Genome Sequences of 48 *Staphylococcus aureus* Isolates Obtained from Atopic Dermatitis Patients and Healthy Controls. *Microbiol Resour Announc* 11:e00072-22. https://doi.org/10.1128/mra.00072-22
   (First author, published)

   Contributions:
   - Genomic DNA extraction
   - Genome assembly, assessment, and annotation
   - Manuscript writing

List of figures

List of tables

Summary

Atopic dermatitis (AD), a chronic inflammatory skin disease, is a growing concern globally, particularly in developed countries, affecting a significant portion of both children and adults. It is widely recognized that AD involves a complex interplay of genetic, immunological, and environmental factors, including a notable disruption in the skin microbiome. A key player in this disruption is *Staphylococcus aureus*, a versatile bacterium found in humans and animals. Known for its dual role as a commensal and an opportunistic pathogen, *S. aureus* is found in up to 30% of the human population, suggesting a considerable diversity among its strains. In the context of AD, *S. aureus* plays a pivotal role. The broken skin barrier characteristic of AD provides an ideal environment for *S. aureus* colonization, leading to increased infection risks and exacerbating the inflammation and itching associated with the condition. The bacterium's presence in AD skin, notably more abundant than in healthy individuals (HE), correlates with the severity of AD symptoms. This relationship is further complicated by the bacterium's capacity for antibiotic resistance (ABR), largely attributed to horizontal gene transfer (HGT).

Our study delved into the genetic and functional differences between *S. aureus* strains isolated from AD patients and HE. Using a global dataset of 300 publicly available strains and a local collection of 48 strains isolated in Augsburg, Germany, the study revealed AD strains had a smaller pan-genome with a higher proportion of core genes, indicating selective pressure for homogeneity in AD-specific strains. In contrast, HE strains had more accessory and unique genes. Functional analysis revealed greater functional variability in AD strains globally and significant difference between countries. Two additional AD-dominant clusters were identified, which was linked to the enrichment of transposes and ABR genes, with HGT events playing an important role. Local analysis mirrored the gene content diversity of global strains but showed distinct functional variations between AD and HE strains. These findings showed selective adaptation of AD strains and scale-dependent diversification mechanisms, enhancing understanding of *S. aureus* variances at global and local levels.

By constructing and employing a random forest classifier based on the genomes from 300 strains publicly available and 48 strains sequenced in our lab, we identified 50 key marker genes that distinguish between AD and HE strains with high accuracy as well as between human and animal strains. A notable discovery was the significant role of prophages - bacteriophages integrated into bacterial genomes - in shaping the genetic diversity and pathogenicity of *S. aureus* strains. We found that 64% of the marker genes were linked to Staphylococcal phage functionalities, with phage holin emerging as a critical differentiating factor. Further analysis highlighted that prophages significantly contribute to the distinction

between gene contents of AD and HE strains. Prophage sequences in HE strains were more diverse in gene content and functions, while those in AD strains were enriched with virulence factors, indicating their substantial role in the pathogenesis of AD-associated strains. Additionally, the study revealed that AD strains undergo stronger purifying selection, resulting in reduced gene content diversity compared to HE strains, suggesting the specific AD environment exerts selective pressures on *S. aureus* strains.

Overall, this study concludes that both local and global *S. aureus* strains adaptively evolve in response to the AD microenvironment. The presence of transposases and the variability in metal-related genes and antimicrobial resistance across different strains underscore scale-dependent diversification processes. HGT events, highlighting the genetic interchange between human-associated and environmental bacteria, further enrich our understanding of *S. aureus* evolution in the context of AD. Besides, our research underscores the critical role of prophages in the genetic diversity and pathogenicity of *S. aureus* in the context of AD. This highlights the potential of prophage profiling in developing diagnostic markers and deepening our understanding of bacterial pathogenesis in AD. Future research directions include exploring the factors and functions influencing genetic diversity in *S. aureus*, potentially leading to improved therapeutic strategies for AD.

Zusammenfassung

Die atopische Dermatitis (AD), eine chronische entzündliche Hauterkrankung, ist weltweit, insbesondere in Industrieländern, von wachsender Bedeutung und betrifft einen erheblichen Teil der Kinder und Erwachsenen. Es ist allgemein anerkannt, dass AD ein komplexes Zusammenspiel von genetischen, immunologischen und umweltbedingten Faktoren beinhaltet, einschließlich einer bemerkenswerten Störung des Hautmikrobioms. Ein Schlüsselakteur in dieser Störung ist *Staphylococcus aureus*, ein vielseitiges Bakterium, das bei Menschen und Tieren vorkommt. Bekannt für seine Doppelrolle als Kommensale und opportunistischer Pathogen, findet sich *S. aureus* bei bis zu 30 % der menschlichen Bevölkerung, was auf eine erhebliche Vielfalt seiner Stämme hindeutet. Im Kontext von AD spielt *S. aureus* eine zentrale Rolle. Die beschädigte Hautbarriere, die für AD charakteristisch ist, bietet eine ideale Umgebung für die Besiedlung durch *S. aureus*, was das Infektionsrisiko erhöht und die mit der Erkrankung verbundene Entzündung und Juckreiz verschlimmert. Die Präsenz des Bakteriums in AD-Haut, insbesondere in größerer Anzahl als bei gesunden Individuen (HE), korreliert mit der Schwere der AD-Symptome. Diese Beziehung wird weiter durch die Fähigkeit des Bakteriums zur Antibiotikaresistenz (ABR) kompliziert, die größtenteils auf horizontalen Gentransfer (HGT) zurückzuführen ist.

Unsere Studie untersuchte die genetischen und funktionalen Unterschiede zwischen *S. aureus*-Stämmen, die von AD-Patienten und HE isoliert wurden. Unter Verwendung eines globalen Datensatzes von 300 öffentlich verfügbaren Stämmen und einer lokalen Sammlung von 48 Stämmen, die in Augsburg, Deutschland, isoliert wurden, zeigte die Studie, dass AD-Stämme ein kleineres Pan-Genom mit einem höheren Anteil an Kerngenen aufwiesen, was auf selektiven Druck für Homogenität in AD-spezifischen Stämmen hindeutet. Im Gegensatz dazu hatten HE-Stämme mehr Zubehör- und einzigartige Gene. Die Funktionsanalyse ergab eine größere funktionale Variabilität in AD-Stämmen weltweit und signifikante Unterschiede zwischen den Ländern. Zwei zusätzliche AD-dominante Cluster wurden identifiziert, die mit der Anreicherung von Transposasen und ABR-Genen verbunden waren, wobei HGT-Ereignisse eine wichtige Rolle spielten. Die lokale Analyse spiegelte die Gendiversität der globalen Stämme wider, zeigte jedoch deutliche funktionale Variationen zwischen AD- und HE-Stämmen. Diese Befunde zeigten die selektive Anpassung von AD-Stämmen und skalenabhängige Diversifizierungsmechanismen auf und verbesserten das Verständnis von *S. aureus*-Variationen auf globaler und lokaler Ebene.

Durch den Aufbau und Einsatz eines Random-Forest-Klassifikators auf der Grundlage der Genome von 300 öffentlich verfügbaren Stämmen und 48 in unserem Labor sequenzierten

Stämmen identifizierten wir 50 Schlüsselmarker-Gene, die zwischen AD- und HE-Stämmen sowie zwischen menschlichen und tierischen Stämmen mit hoher Genauigkeit unterscheiden. Eine bemerkenswerte Entdeckung war die bedeutende Rolle von Prophagen - in bakterielle Genome integrierte Bakteriophagen - bei der Gestaltung der genetischen Diversität und Pathogenität von *S. aureus*-Stämmen. Wir fanden heraus, dass 64 % der Marker-Gene mit Staphylokokken-Phagenfunktionen verbunden waren, wobei Phagen-Holin als ein kritischer differenzierender Faktor hervortrat. Weitere Analysen zeigten, dass Prophagen erheblich zur Unterscheidung zwischen den Gengehalten von AD- und HE-Stämmen beitragen. Prophagensequenzen in HE-Stämmen waren in Gengehalt und Funktionen vielfältiger, während diejenigen in AD-Stämmen mit Virulenzfaktoren angereichert waren, was auf ihre erhebliche Rolle in der Pathogenese von AD-assoziierten Stämmen hinweist. Außerdem ergab die Studie, dass AD-Stämme einer stärkeren reinigenden Selektion unterliegen, was zu einer reduzierten Gendiversität im Vergleich zu HE-Stämmen führt und darauf hindeutet, dass die spezifische AD-Umgebung selektiven Druck auf *S. aureus*-Stämme ausübt.

Insgesamt schließt diese Studie, dass sowohl lokale als auch globale *S. aureus*-Stämme adaptiv auf das AD-Mikroumfeld reagieren. Die Präsenz von Transposasen und die Variabilität in metallbezogenen Genen und antimikrobieller Resistenz in verschiedenen Stämmen unterstreichen skalenabhängige Diversifizierungsprozesse. HGT-Ereignisse, die den genetischen Austausch zwischen menschenassoziierten und umweltbedingten Bakterien hervorheben, bereichern unser Verständnis der Evolution von *S. aureus* im Kontext von AD. Darüber hinaus unterstreicht unsere Forschung die entscheidende Rolle von Prophagen in der genetischen Diversität und Pathogenität von *S. aureus* im Kontext von AD. Dies hebt das Potenzial der Prophagen-Profilerstellung hervor, diagnostische Marker zu entwickeln und unser Verständnis der bakteriellen Pathogenese bei AD zu vertiefen. Zukünftige Forschungsrichtungen umfassen die Erforschung der Faktoren und Funktionen, die die genetische Diversität bei *S. aureus* beeinflussen, und könnten zu verbesserten therapeutischen Strategien für AD führen.

1 Introduction

1.1 Atopic dermatitis

Atopic dermatitis (also known as atopic eczema, AD) affects over 20% of children and 10% of adults in developed countries and exhibits an increase in prevalence in developing countries (Deckers *et al.*, 2012; Flohr and Mann, 2014; Weidinger and Novak, 2016; Czarnowicki *et al.*, 2017). AD is often associated with other comorbidities such as food allergy, asthma, allergic rhinitis, and mental health disorders (Weidinger and Novak, 2016). While the exact etiology of AD remains unknown, it is generally accepted that it is particularly complex and multifaceted (Shi *et al.*, 2016; Totté *et al.*, 2016; Tomczak *et al.*, 2019), with complicated interplays between genetic, immunological, skin barrier, and environmental factors involved in the process (Mu and Zhang, 2020).

1.1.1 Clinical symptoms of AD

AD is a very common skin chronic inflammation characterized by dry, itchy, inflamed, and recurrent skin lesions (Langan *et al.*, 2020). The lack of definitive criteria results in a huge difficulty in diagnosing AD, as it is highly heterogeneous regarding clinical features, severity, and course. Nevertheless, there are several common and associated features used in clinical settings, which include eczematous lesions, intense pruritus, and a chronic or relapsing disease course (Vakharia *et al.*, 2018; Reynolds *et al.*, 2020). Infants with eczema often have widespread acute lesions that mostly affect the face, cheeks, and trunk, but usually not the diaper area (Lobefaro *et al.*, 2022). These lesions can include redness, swelling, blisters, scratches, and fluid discharge. As children grow older (2 years and up), eczema becomes more localized and chronic, typically affecting the areas where the skin bends, presenting as dry skin, pale redness, and thicker skin in chronic areas (Barnetson and Rogers, 2002). Adolescents and adults might have widespread eczema or localized lesions that especially affect hands, eyelids, and flexing areas. Some adults experience eczema only on their hands or have dermatitis on their head, neck, upper body, and scalp (Lobefaro *et al.*, 2022). Two morphological variants include the follicular type, seen frequently in those of Asian and African descent, and the prurigo type, which is common in patients with a longer history of the disease, characterized by scratched bumps and hardened lumps (Smith, 2000; Torrelo, 2014; Langan *et al.*, 2020).

1.1.2 Prevalence of AD

Based on the global burden of disease (GBD) data for 2019 (https://vizhub.healthdata.org/gbd-results/), approximately 223 million individuals suffer from AD. Out of these, nearly 43 million are children aged between 1 to 4 years, underscoring the notably high occurrence in this age group. The burden of AD ranks 15th worldwide for non-fatal diseases and No. 1 for skin diseases (Laughter *et al.*, 2021; Arents *et al.*, 2022). AD can also persist into adulthood and last across the lifespan of many children (Kim *et al.*, 2016; Thyssen *et al.*, 2019). Based on the GBD data between 1990 and 2017 (Figure 1), the global prevalence rate of AD has been steadily increasing. For the age groups, it indicates that the prevalence of AD is the most pronounced in the 1-4 age group. As age progresses, the incidence diminishes swiftly until it plateaus. However, there is a resurgence in cases for individuals aged 50 and above (Laughter *et al.*, 2021).

Globally, the prevalence of AD varies in different countries. The top five most prevalent countries are all located in Europe, including Sweden, the UK, Iceland, Finland, and Denmark, with the prevalence rate fluctuating between 5300 and 7700 per 100K population. while the lowest five are developing countries including Uzbekistan, Armenia, Tajikistan, China, and Kazakhstan, with the rate ranging from 1700 to 2000 (House, 2022). Whereas, AD ranks as the foremost cause of skin-related health burdens globally and is witnessing a rising prevalence in developing regions such as Asia, Africa, Latin America, and the Middle East (Al-Afif *et al.*, 2019; Lopez Carrera *et al.*, 2019). Together, these facts make AD an important public health problem globally and an urgency to increase our understanding of it.

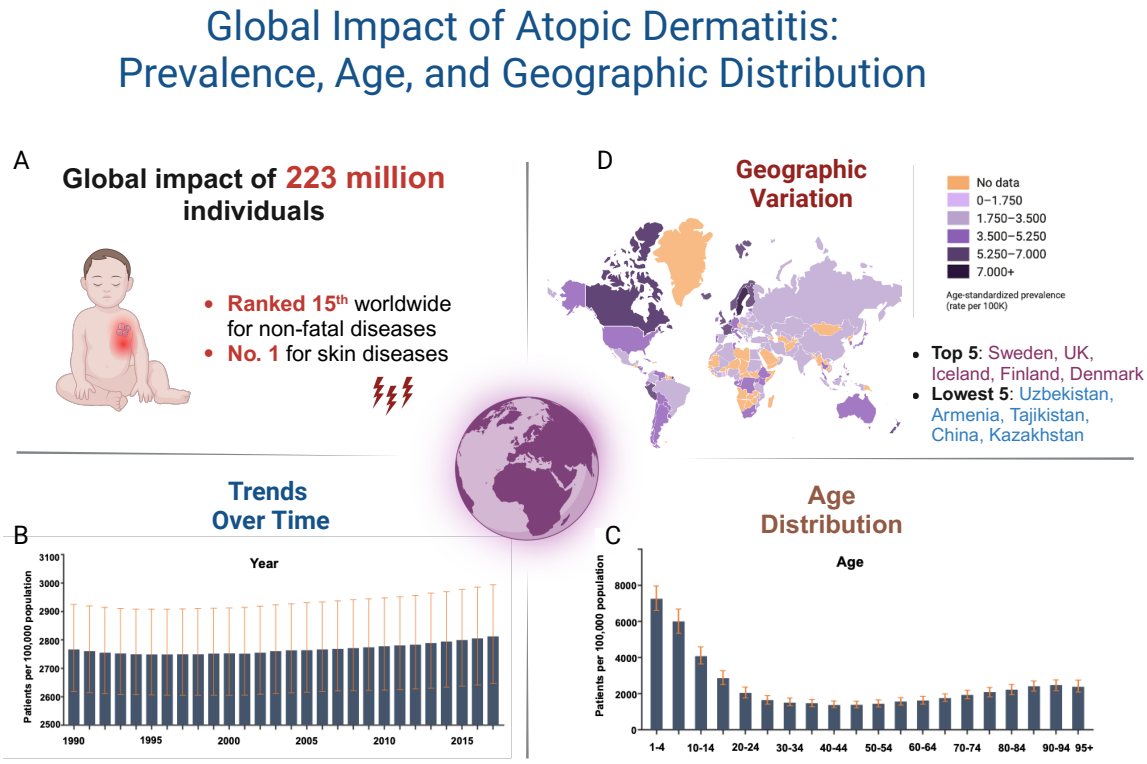

Figure 1. The global occurrence rate of AD per 100,000 individuals, showcasing A) the prevalence, B) the trends over years, C) the age distribution, and D) the geographic variation across different countries. Age-standardized prevalence of AD per country is shown for geographic variation (2017). The bars for trends and age indicate the 95% confidence intervals for these rates. This visualization is adapted based on GBD data from Laughter *et al.*, 2021, reproduced with permission.

1.1.3 Association of AD with other comorbidities

AD is closely associated with other atopic conditions, such as food allergy, asthma, and allergic rhinitis (Langan *et al.*, 2020; Lax *et al.*, 2022). While this association is evident, only about half of those with AD display an allergic constitution (Bieber *et al.*, 2022; Lax *et al.*, 2022). The combined occurrence of these diseases may have a shared genetic basis (Ferreira *et al.*, 2017). The progression of atopic diseases with age is known as the 'atopic march', traditionally believed to start with AD, then food allergies, and finally asthma and allergic rhinitis. However, this model oversimplifies the disease pathways, as they can vary widely (van Zuuren *et al.*, 2017; Ramírez-Marín and Silverberg, 2022). People with AD often experience both allergic and irritant contact dermatitis (Silverberg, 2019; Appiah *et al.*, 2022; Ramírez-Marín and Silverberg, 2022). In adults, AD has been linked to immune-mediated conditions, decreased bone health, metabolic disorders, and cardiovascular diseases, though the last two associations are debated (Davis *et al.*, 2022). In essence, AD is usually the initial indicator of atopy and is a multifaceted condition with diverse health implications.

1.1.4 Pathophysiology and mechanisms of AD

The underlying pathophysiology of AD has been subject to a lot of research. Although the precise etiology has not been fully elucidated, many factors and triggers may play a role in the development of AD. As Figure 2 depicted, these include interactions between genetic risk factors, skin barrier disruption, and microbiome alterations (Weidinger and Novak, 2016; van Zuuren *et al.*, 2017; Bieber *et al.*, 2022; Ramírez-Marín and Silverberg, 2022).

For genetic risk factors, a positive family history of AD has been identified as the most significant risk factor for its development (Apfelbacher *et al.*, 2011). AD is a highly heritable disease, with the heritability estimated to be approximately 70%-90% in twin studies (Elmose and Thomsen, 2015; Al-Shobaili *et al.*, 2016; Sliz *et al.*, 2022). While 34 genetic loci related to AD have been identified, they account for less than 20% of its heritability (Paternoster *et al.*, 2015). Most of these genetic regions are linked to immune responses and skin barrier genes, though specific functional variants remain largely unidentified (Weidinger *et al.*, 2013; Paternoster *et al.*, 2015; Martin *et al.*, 2020). The strongest known genetic risk factor for AD is null mutations in filaggrin (*FLG*), which encodes a key epidermal structural protein (Irvine *et al.*, 2011). Mutations resulting in *FLG* loss-of-function elevate the risk of AD by 3-5 times compared to healthy individuals (Irvine *et al.*, 2011; Riethmuller *et al.*, 2015) and predispose

AD patients to both asthma and peanut allergies (Rice *et al.*, 2008; Rodríguez *et al.*, 2009; Brown *et al.*, 2011).

For skin barrier disruption, AD patients display consistent epidermal barrier dysfunction in both affected and unaffected skin, evidenced by factors like increased water loss (Kelleher *et al.*, 2015), increased pH (Hülpüsch *et al.*, 2020), and modified lipid composition (Yin *et al.*, 2023). This barrier disruption is due to multiple factors, including genetic mutations like *FLG* and physical damage from scratching. Skin's type 2 immune activity downregulates barrier genes, worsening the barrier defect (Howell *et al.*, 2009; Kezic *et al.*, 2011). Stressed keratinocytes in this compromised barrier release proinflammatory signals via epidermal alarmins IL-33 and thymic stromal lymphopoietin (TSLP) that drive type 2 inflammation, further damaging the tissue and activating certain inflammatory cells (Salimi *et al.*, 2013; Saunders *et al.*, 2016).

For the dysbiosis of the skin microbiome, AD is linked to an imbalanced skin microbiome with a lower bacterial diversity (Kong *et al.*, 2012). Human skin houses diverse microorganisms that maintain a balanced environment and are necessary for skin barrier function by preventing colonization by pathogens (Rosenberg and Zilber-Rosenberg, 2018). Microorganisms influence both innate and adaptive immune responses, impacting the early development of the immune system and the onset of AD (Pothmann *et al.*, 2019). Commensal bacteria like *Staphylococcus epidermidis* and *Staphylococcus hominis* play a pivotal role in skin T-cell development, reducing inflammation, and warding off skin infections through the production of antimicrobial peptides (Nakatsuji *et al.*, 2017). In AD patients, reduced microbiome diversity is linked to disease severity and higher pathogenic bacteria levels (LEYDEN *et al.*, 1974; Kong *et al.*, 2012; Altunbulakli *et al.*, 2018; Paller *et al.*, 2019). Preliminary studies suggest that applying specific commensal organisms such as *S. epidermidis*, *S. hominis* or *Roseomonas mucosa* topically can reduce AD severity (Naik *et al.*, 2012; Zipperer *et al.*, 2016; Myles *et al.*, 2018; Paller *et al.*, 2019), emphasizing the role of commensals in countering pathogenic colonization (Zipperer *et al.*, 2016). Additionally, microbial imbalances such as *Staphylococcus aureus* and Malassezia yeasts (Glatz *et al.*, 2015) might amplify skin inflammation and further impair the barrier. In particular, subjects with AD are highly colonized by *S. aureus* (LEYDEN *et al.*, 1974; Totté *et al.*, 2016; Altunbulakli *et al.*, 2018). Nakatsuji *et al.* found that healthy skin is protected against *S. aureus* by commensal bacteria such as coagulase-negative *Staphylococcus* (CoNS) and a deficit in these commensals is linked to increased *S. aureus* colonization in AD skin. By transplanting CoNS strains onto AD skin, they noticed a reduction in *S. aureus* colonization (Nakatsuji *et al.*, 2017). Therefore, understanding the role of *S. aureus* in AD development to keep a balanced skin microbiome could pave the way for innovative treatments.

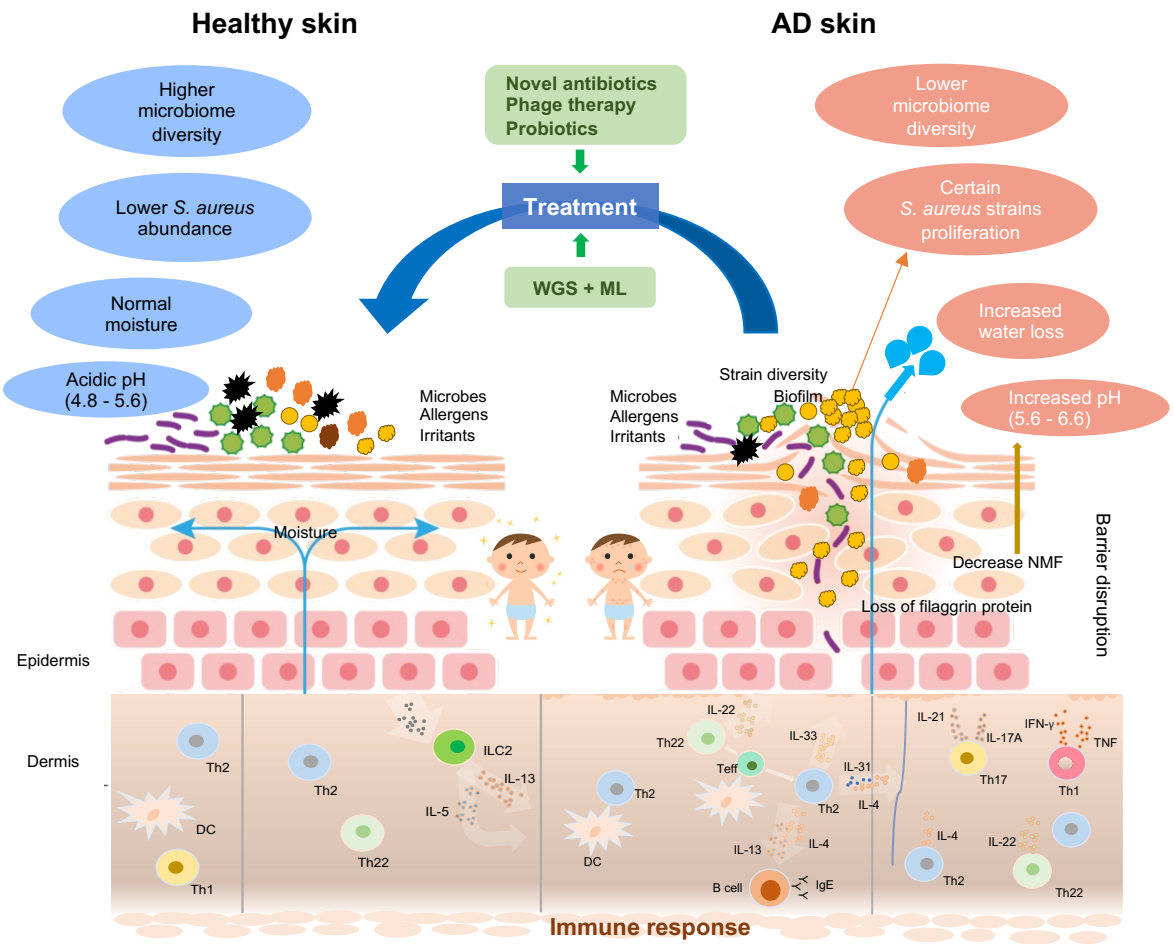

Figure 2. Schematic diagram of the normal skin physiology and principal mechanisms of AD pathogenesis with *S. aureus* colonization. Normal skin is clinically intact with balanced diverse bacteria and physiological traits. AD skin exhibits dysfunction of the epidermal barrier, alongside a reduced diverse microbiome, increased water loss on the surface, and elevated pH. In affected areas, certain *S. aureus* strains proliferate and heightened immune responses are triggered: Langerhans cells and inflammatory epidermal dendritic cells, which carry specific IgE attached to the high-affinity IgE receptor, as well as dermal dendritic cells, facilitate the introduction of allergens and antigens. Cytokines IL-4, IL-13, and IL-31 of the type-2 class directly stimulate sensory nerves, leading to itchiness. As the condition becomes more severe, there is a noticeable escalation in cytokines originating from keratinocytes and Th cells. WGS=whole genome sequencing. ML=machine learning. DC=dendritic cell. IFN=interferon. IL=interleukin. ILC=innate lymphoid cell. Teff=effector T cell. Th=T-helper cell. TNF=tumour necrosis factor.

1.2 *Staphylococcus aureus* and its association with AD

1.2.1 General characteristics of *S. aureus*

S. aureus is round-shaped and characterized by a thick peptidoglycan layer, a hallmark of Gram-positive bacteria (Silhavy *et al.*, 2010). The genome of *S. aureus* is relatively compact, with sizes ranging approximately from 2.7 to 3.0 Mb (million base pairs) encoding about 2,600 to 3,000 genes and GC content fluctuating around 32%-33% (Holden *et al.*, 2004; Obata *et al.*, 2023). The genomic content varies among strains, which contributes to their adaptability and the diversity of infections they can cause. The genome of *S. aureus* includes a notable amount of mobile genetic elements (MGEs) that account for 15-20% of the genome, such as plasmids, transposons, pathogenicity islands, and bacteriophages, which facilitate genetic exchange and contribute to the rapid acquisition of antibiotic resistance and virulence factors (Lindsay, 2010; Malachowa and Deleo, 2010).

S. aureus cells are spherical, roughly 0.5-1.5 micrometer in diameter, and typically appear in clusters resembling grapes when viewed under a microscope (Gnanamani *et al.*, 2017). *S. aureus* can grow in a wide range of temperatures, from 7°C to 48°C, with an optimal temperature around 37°C (Kadariya *et al.*, 2014). It can also tolerate high salt concentrations (up to 15% NaCl), which is why it can grow in foods like ham (Kadariya *et al.*, 2014). It is facultatively anaerobic, meaning it can grow in the presence or absence of oxygen, although aerobic conditions are generally more favorable for rapid growth (Belay and Rasooly, 2002). *S. aureus* exhibits versatility in substrate utilization and adaptability to host nutritional landscape, which supports its growth in varied environments and host tissues (Teoh *et al.*, 2021).

S. aureus has a long evolutionary history as a multi-host commensal and opportunistic pathogen in both human and animal hosts (O'Gara, 2017; Richardson *et al.*, 2018). *S. aureus* is a common member of human flora and colonizes approximately 20-30% of humans persistently in the nose (Wertheim *et al.*, 2005) and frequently in other body niches such as the skin, throat, axillae, and intestine (Williams, 1963; Lowy, 1998). *S. aureus* has also been isolated from the skin and mucous membranes of animals like livestock (cattle, sheep, goats, etc.) and pets like dogs and cats (Haag *et al.*, 2019; Howden *et al.*, 2023). The ability of *S. aureus* to survive on surfaces and in various environmental conditions contributes to its spread in both community and healthcare settings (Howden *et al.*, 2023). *S. aureus* is normally a commensal bacterium of the human body, with certain strains even suggested to harbor beneficial immunomodulatory properties (Kumar *et al.*, 2015; Yang *et al.*, 2018). Nevertheless,

it is often recognized as one of the notoriously foremost causes of many clinical manifestations, ranging from mild skin infections to serious invasive conditions, including impetigo, bacteraemia, osteomyelitis, septic arthritis, pneumonia, and endocarditis (Lowy, 1998; Tong *et al.*, 2015).

Acting as an opportunistic pathogen, *S. aureus* plays an important role in inflammatory skin infections like AD (Breuer *et al.*, 2002; De Benedetto *et al.*, 2009; Kim *et al.*, 2011; Park *et al.*, 2013; Shi *et al.*, 2018). Despite substantial evidence highlighting its significant role in disease progression and symptom severity, the mechanisms by which *S. aureus* contributes to AD, including its interactions with the overall skin microbiome and its role in skin inflammation, have not yet been fully understood. Therefore, understanding the general features of *S. aureus*, from its occurrence to its genomic plasticity and pathogenic potential, is crucial for developing effective prevention and treatment strategies against the diverse array of infections it can cause.

1.2.2 Prevalence of *S. aureus* in AD patients

Healthy skin harbors diverse commensal bacteria, but skin affected by severe AD is largely colonized by *S. aureus* (Kong *et al.*, 2012; Byrd *et al.*, 2017; Altunbulakli *et al.*, 2018), as Figure 3 depicted. Deep metagenomic sequencing has revealed a decrease in microbial diversity during an AD flare, with a shift toward the dominance of *S. aureus* which can be reverted to normal abundance during treatment and recovery, reflecting a shift of pathogenic degree of *S. aureus* (Kong *et al.*, 2012; Byrd *et al.*, 2017). An amplicon analysis also showed that AD is usually characterized by reduced microbiome diversity, accompanied by *S. aureus* predominance (Altunbulakli *et al.*, 2018). Another amplicon study further highlighted that dysbiosis in the skin microbiome of AD patients is predominantly driven by the evenness, rather than the richness, of the skin microbiome, attributed to the overgrowth of *S. aureus* (Rauer *et al.*, 2023). The exact connection between *S. aureus* proliferation and AD onset is uncertain. However, early life colonization by non-*S. aureus* staphylococcus bacteria may reduce AD risk, while early *S. aureus* colonization could increase it (Kim *et al.*, 2019).

S. aureus is prevalent on the skin of AD patients, with carriage rates between 30%-100%, compared to a 20% prevalence in healthy individuals (HE) (Higaki *et al.*, 1999; Guzik *et al.*, 2005; Park *et al.*, 2013; Tauber *et al.*, 2016; Totté *et al.*, 2016). A meta-analysis found that 70% of AD patients carried *S. aureus* on lesional skin, versus 39% on non-lesional skin and 62% for the nose (Totté *et al.*, 2016), which demonstrates the importance of colonization with *S. aureus* in AD.

Both the nose and skin are significant reservoirs of *S. aureus* in AD patients, with frequent recolonization between these areas potentially increasing AD severity (Chiu *et al.*, 2009; Alsterholm *et al.*, 2017). Research indicates that 65-77.3% of AD patients have *S. aureus* in both their noses and skin (Breuer *et al.*, 2002; Tomi *et al.*, 2005; Na *et al.*, 2012), markedly higher than the 10.2% of HE (Matsui *et al.*, 2000). This disparity is more pronounced in children with AD, where *S. aureus* colonization affects 57%-100% of this group, higher than the healthy group (Lo *et al.*, 2010; Pascolini *et al.*, 2011; Shi *et al.*, 2016). Notably, over 40% of children suffer from *S. aureus* colonization both on the lesional skin and in the nose (Pascolini *et al.*, 2011). In adults with AD, about 54%-100% of this group were colonized by *S. aureus* strains (Breuer *et al.*, 2002; Kim *et al.*, 2009; Clausen *et al.*, 2017, 2019). *S. aureus* most frequently colonize lesional skin (56-96.2%), followed by the nose (46.1-64.1%) and non-lesional skin (28-39%) (Matsui *et al.*, 2000; Park *et al.*, 2016; Totté *et al.*, 2016; Alsterholm *et al.*, 2017; Clausen *et al.*, 2017). Furthermore, about 55% of AD patients are consistent carriers of *S. aureus* and they exhibit more severe symptoms than intermittent carriers or non-carriers (Alsterholm *et al.*, 2017). More importantly, *S. aureus* colonizing the anterior nares could be a potential source for recolonization of other skin sites (K. Patel, H. Wyatt, E. M. Kubiak, S, 2001; Breuer *et al.*, 2002).

Higher density, fraction, and total CFU/cm^2 (colony forming units) were revealed on the lesional skin of AD patients than in the healthy control group (Monti *et al.*, 1996; Abeck and Mempel, 1998). The density of *S. aureus* colonization on both lesional and non-lesional skin positively correlates with AD disease severity (LEYDEN *et al.*, 1974; Tauber *et al.*, 2016), as evidenced by both RT-PCR studies and earlier culture-based studies (Higaki *et al.*, 1999; Guzik *et al.*, 2005). Besides, a recent study reported that both lesional and non-lesional skin showed significantly higher total bacterial load, increased *S. aureus* relative abundance and cell numbers compared to healthy controls, with severe AD patients possessing even higher *S. aureus* cell numbers and total bacterial load, underscoring both higher relative and absolute abundances of *S. aureus* in AD (De Tomassi *et al.*, 2023).

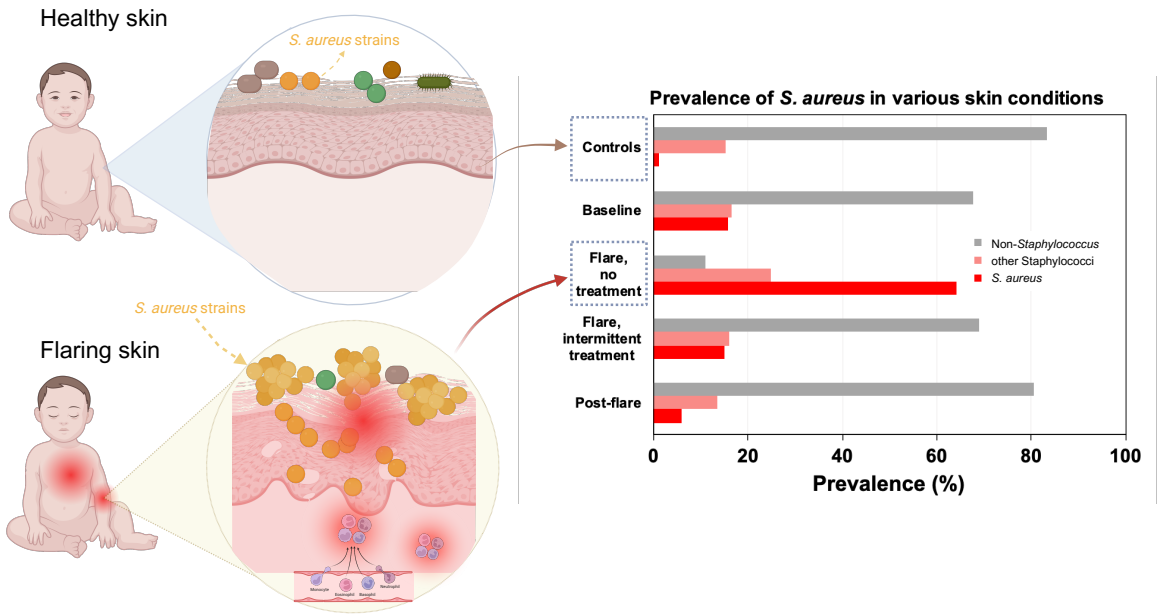

Figure 3. Schematic of the prevalence of *S. aureus* strains in different skin conditions. Mean relative abundance of *S. aureus* (red), other *Staphylococci* (pink), and non-*Staphylococcus* in the skin for controls and AD disease states: baseline, flare (no-treatment and intermittent-treatment), and post-flare. The barplot illustration of abundance is adapted based on the data provided by Kong *et al*., 2012, with permission, wherein the mean proportion of all individuals for each disease state is recalculated and utilized.

1.2.3 Pathogenicity of *S. aureus* in AD patients

S. aureus has evolved various strategies to anchor itself in the challenging environment of healthy, intact human skin. However, the skin of individuals with AD is more receptive to *S. aureus* colonization as the skin barrier is broken. Colonization by *S. aureus* may trigger a vicious cycle, intensifying both *S. aureus* infections and AD flare-ups through the induction of TSLP and Th2/Th17-type inflammatory responses (Vu *et al.*, 2010; Byrd *et al.*, 2017). Several factors contribute to the colonization of AD skin by *S. aureus*, which includes the intensity of *S. aureus*-corneocyte binding, reduced antimicrobial peptides, diminished levels of filaggrin and its degradation products, heightened Th2/Th17 cytokines, microbial imbalance, and modified lipid profiles (Kim *et al.*, 2019).

To enhance its colonization of human nasal mucosa and skin, *S. aureus* deploys various factors facilitating surface binding and survival. These factors include microbial surface components recognizing adhesive matrix molecules like fibronectin-binding proteins (Fnbp A and Fnbp B), fibrinogen-binding proteins (ClfA and ClfB), iron-regulated surface determinant A (IsdA), and wall teichoic acid (Cho, Strickland, Tomkinson, *et al.*, 2001; Weidenmaier *et al.*, 2004; Clarke *et al.*, 2006; Burian *et al.*, 2010). *S. aureus* adheres more efficiently to biopsies from non-lesional skin of AD patients than those from psoriasis patients or healthy individuals (Cho, Strickland, Boguniewicz, *et al.*, 2001). This bacterial adherence predominantly occurs in the stratum corneum. Changes in the stratum corneum's composition and corneocyte morphology likely could explain why *S. aureus* preferentially adheres to AD skin. Fibronectin in AD skin may be present in the stratum corneum, serving as a ligand for *S. aureus* binding (Cho, Strickland, Boguniewicz, *et al.*, 2001). The fibronectin-binding proteins (Fnbp A and Fnbp B) on the surface of *S. aureus* enable it to adhere strongly to AD skin (Foster *et al.*, 2014). Elevated levels of Th2 cytokines, such as interleukin (IL)-4 in AD promote *S. aureus* binding to fibronectin and fibrinogen (Cho, Strickland, Tomkinson, *et al.*, 2001). *S. aureus* has other ligands exposed in the stratum corneum, such as the cell-wall-anchored (CWA) protein ClfB, which binds to specific envelope proteins loricrin and cytokeratin (Mulcahy *et al.*, 2012) and has shown to promote bacterial adherence to corneocytes from AD patients (Fleury *et al.*, 2017).

Additionally, *S. aureus* expresses factors like CWA protein IsdA and aureolysin to counter host cationic antimicrobial responses, rendering it resistant to bactericidal fatty acids and antimicrobial peptides (Clarke *et al.*, 2006, 2007). Other CWA proteins promote adhesion to nasal squamous epithelial cells, which may also contribute to *S. aureus* colonization on AD skin due to the similarities between these cells and skin corneocytes (Corrigan *et al.*, 2009).

Corneocyte morphology in AD patients is significantly altered, particularly in patients with low natural moisturizing factor (NMF), with projections emerging from the surface of corneocytes (Riethmuller *et al.*, 2015). These projections are coated with corneodesmosin, typically found at cell-cell junctions. The altered distribution of corneodesmosin in AD corneocytes and the structural changes due to low NMF suggest that stratum corneum of AD skin might present unique ligands for *S. aureus* adherence (Riethmuller *et al.*, 2015).

Moreover, Virulence factors (VF) of *S. aureus* contribute to its evasion in AD. *S. aureus* produces various secreted and wall-anchored VFs that play roles in the pathogenesis of both superficial and invasive infections (Berube and Wardenburg, 2013; Spaulding *et al.*, 2013; Foster *et al.*, 2014). Some of these VFs have been linked to the pathogenesis of AD based on studies on keratinocytes, immune cells, and murine allergy models, as well as their presence in AD skin samples. When *S. aureus* grows in vivo (Sen *et al.*, 2016), it integrates short-chain unbranched fatty acids into its cytoplasmic membrane, which increases membrane fluidity, potentially impacting the expression of VFs (Sun *et al.*, 2012) and the bacterium's ability to withstand host defenses, like resistance to oxidative stress via staphyloxanthin (Sen *et al.*, 2016). The integration of skin fatty acids into lipoproteins also enhances their proinflammatory properties (Nguyen *et al.*, 2015). Superantigens, such as toxic shock syndrome toxin-1 (TSST1) and the staphylococcal enterotoxin serotypes SEA, SEB, SEC, SED, SEE, or SEG, can shift the immune response toward Th2 cytokines which result in excessive T cell cytokine production (Spaulding *et al.*, 2013) and trigger mast cell degranulation directly (Ono *et al.*, 2012), favoring *S. aureus* colonization in AD (Laouini *et al.*, 2003). Cysteine protease staphopain and metalloprotease aureolysin in *S. aureus* can cleave the antimicrobial peptide LL-37 (Sieprawska-Lupa *et al.*, 2004; Sonesson *et al.*, 2017). Staphopain grows in a biofilm on the skin surface, along with host proteases induced by *S. aureus*, which may protect *S. aureus* from antimicrobial peptides in AD skin. The V8 serine protease affects the skin barrier integrity (Hirasawa *et al.*, 2010) and other *S. aureus* proteases may enhance bacterial penetration of the epidermal barrier and disrupt the barrier function (Nakatsuji *et al.*, 2016). The staphylococcal serine protease-like (Spl) proteins trigger a Th2-driven immune response, with recombinant SplD eliciting allergic lung inflammation with antigen-specific IgE and elevated Th2 cytokines (Stentzel *et al.*, 2017). These suggest potential links between these bacterial VFs and immunological responses in AD patients.

A wide spectrum of antimicrobial resistance (AMR) is another important mechanism for *S. aureus* pathogenesis in AD. AMR is a global concern, impacting our ability to treat infectious diseases and posing serious threats to human health (Prestinaci *et al.*, 2015). Various topical and systemic antimicrobial agents are used routinely to target *S. aureus* in the treatment of

AD. Over the past few decades, *S. aureus* has evolved resistance to many of the commonly used antibiotics (Vestergaard *et al.*, 2019). The application of antimicrobials has had a direct impact on the rise and dissemination of prominent AMR strains, making the treatment even more challenging (Chambers and DeLeo, 2009; Harkins *et al.*, 2019; Saheb Kashaf *et al.*, 2023). There is a growing body of evidence suggesting that *S. aureus* strains with AMRs might be more virulent, leading to exacerbated symptoms in AD patients. Some studies have proposed that resistance mechanisms, such as the presence of the *mecA* gene in MRSA, might be linked with increased toxin production (Koosha *et al.*, 2016), which can make MRSA-colonized AD lesions more inflamed and harder to treat. Resistant strains often have a heightened ability to form biofilms that are complex bacterial communities encased in a matrix, making them more resilient to host immune defenses and antibiotic treatment (Piechota *et al.*, 2018; Ou *et al.*, 2020; Leshem *et al.*, 2022). Biofilms can enhance *S. aureus* persistence on AD skin and intensify the inflammatory response (Domenico *et al.*, 2019). Besides, the application of antibiotics can cause dysbiosis of the skin microbiota, reducing beneficial microbes and allowing resistant *S. aureus* to thrive (Zhang *et al.*, 2015). A dysbiotic skin microbiome can further impair barrier function and immune responses in AD (Kobayashi *et al.*, 2015; Nakatsuji *et al.*, 2017).

AMR profiles of *S. aureus* from different geographical locations are also different. An analysis from Ireland revealed that antimicrobial resistance profiles of *S. aureus* in children with AD (n=50) differ from those in healthy and nonatopic individuals (n=49), with a notable prevalence of fusidic acid resistance and distinct genetic mechanisms in AD cases, along with variations in antiseptic resistance determinants (C. P. Harkins *et al.*, 2018). This result suggests that strong selective pressure drives the emergence and maintenance of specific resistances in AD. A study from Brazil involving 91 AD patients and 100 *S. aureus* isolates observed that methicillin-susceptible *S. aureus* exhibited low resistance to mupirocin (1.1%) and fusidic acid (5.9%), but high resistance to neomycin (42.6%) and bacitracin (100%) (Bessa *et al.*, 2016). Fusidic acid resistance was linked to more severe AD, as indicated by higher EASI scores (median 17.8 vs. 5.7, p=.009) (Bessa *et al.*, 2016). These findings emphasize the importance of restricting the differential use of antimicrobial agents to minimize resistance issues in different regions.

Overall, the association between AMR in *S. aureus* and its augmented pathogenicity in AD underscores the need for judicious antibiotic use. Over-reliance on antibiotics can further propagate resistance and potentially worsen AD outcomes. Alternatives to traditional antibiotics, such as phage therapy, antimicrobial peptides, and probiotic interventions, are being explored to tackle resistant *S. aureus* in AD.

1.2.4 Horizontal gene transfer (HGT) of *S. aureus*

HGT enables organisms to supplement their inherited genetic repertoire, frequently as an adaptation to environmental alterations, thereby markedly enhancing genetic diversity and expediting genomic innovation and evolutionary processes through the incorporation of novel genes into extant genomes (Jain, 2003; Arnold *et al.*, 2022). Antibiotic resistance genes (ARGs) encoded on MGEs are rapidly spread through HGT, posing a significant challenge in clinical settings by obstructing antibiotic treatments and facilitated by environmental biofilms that act as critical vectors for their dissemination (Michaelis and Grohmann, 2023). Occurring across diverse species, HGT effectively circumvents the more gradual mechanism of de novo gene development. *S. aureus* hosts numerous virulence and resistance genes within mobile genetic elements, highlighting the significant impact of HGT on its evolution (Lindsay, 2010; Cafini *et al.*, 2017). For example, the widespread antibiotic resistance observed in Methicillin-resistant *S. aureus* (MRSA) strains is largely attributed to HGT, with transduction—the process of transferring DNA between bacteria via bacteriophages—being the main mechanism for the spread of methicillin resistance in *S. aureus* (Lee *et al.*, 2018; Turner *et al.*, 2019). Studies have demonstrated that MRSA strains evolve particularly in virulence and antibiotic resistance genes (McCarthy and Lindsay, 2010; McCarthy *et al.*, 2014). Thus, investigating the HGT of ARGs in *S. aureus* is essential for understanding its adaptability and resistance mechanisms.

1.2.5 Strain diversity of *S. aureus* in AD

Diverse and versatile bacteria live on the human skin. Studies have emphasized the significance of understanding the skin microbiota at the most detailed strain-level differences (Byrd *et al.*, 2018; Caldwell *et al.*, 2022). The genetic diversity makes it challenging to discern strain-level functionality (Oh *et al.*, 2014; Zhou *et al.*, 2020; Conwill *et al.*, 2022). To truly understand the skin microbiome's role, it is crucial to study microbial strain diversity in the context of the disease background.

Many studies have highlighted the great strain-level diversity within *S. aureus*, defined with clonal complexes (CC) or sequence types (ST), and consolidated the link between specific strains and AD pathogenesis (Clausen *et al.*, 2017; Fleury *et al.*, 2017), as particular strains of *S. aureus* are found more frequently on AD-affected skin than others (Byrd *et al.*, 2017; Altunbulakli *et al.*, 2018; Catriona P. Harkins *et al.*, 2018). Metagenomic sequencing of *S. aureus* from AD patients demonstrated that more severe patients are usually colonized with a single clade of *S. aureus* during disease flare which can be reverted to normal abundance

during treatment and recovery (Kong *et al.*, 2012; Byrd *et al.*, 2017) and the strains of *S. aureus* in AD patients differ from those in unaffected carriers (Simpson *et al.*, 2018). Several studies have identified that CC1 isolates are over-presented on AD-inflamed skin while CC30 isolates occur most frequently on healthy controls (Melles *et al.*, 2004; Yeung *et al.*, 2011; Rojo *et al.*, 2014; Fleury *et al.*, 2017; Catriona P. Harkins *et al.*, 2018). A large number of VFs secreted by *S. aureus* have been proven to play roles in the pathogenesis of skin infections via supporting cell wall anchoring or proinflammation that often show strain-specific patterns (Spaulding *et al.*, 2013; Riethmuller *et al.*, 2015), such as superantigens (Spaulding *et al.*, 2013) and phenol-soluble modulins (Peschel and Otto, 2013). It was observed that *S. aureus* colonizing AD skin harbors strain-specific cell wall surface proteins, which affect skin adhesion and induce altered immune responses (Iwamoto *et al.*, 2017, 2019; Moriwaki *et al.*, 2019). The surging evidence of considerable strain diversity within *S. aureus* species potentially explains the variance between pathogenic versus non-pathogenic strains, which could be caused by tens to hundreds of genes resulting in various phenotypic or physiological consequences, such as AMRs and VFs (Qin *et al.*, 2016; Nakatsuji *et al.*, 2017). However, a recent genomic characterization revealed that *S. aureus* strains cannot be differentiated based on whole-genome in terms of the health status of the subject the strains were isolated, including healthy skin, AD non-lesional skin, and AD lesional skin (Obata *et al.*, 2023).

AD creates a unique microenvironment on the skin, potentially impacting the microbial landscape and, in particular, influencing the prevalence and behavior of *S. aureus*. This specific environment might drive adaptive selection, leading to a more homogenous set of *S. aureus* strains. These strains are believed to be genetically and, therefore, functionally distinct from those isolated from HE. Such differences could be the basis for the observed variation in pathogenicity and host response between individuals with AD and healthy controls. However, the presence of systemic genomic variations between AD and HE strains, or the possibility of their roles shifting from commensal to pathogenic depending on skin conditions, remains to be fully understood.

Moreover, gene acquisition and mutations or polymorphisms in specific genes have been identified as playing a crucial role in certain AD-associated strains. Prior research has shown differences between *S. aureus* strains isolated from AD and HE, suggesting that strains associated with AD are better suited to thrive on inflamed skin, thereby offering them a selective advantage in the context of AD (Byrd *et al.*, 2017; Catriona P. Harkins *et al.*, 2018; Key *et al.*, 2023; Saheb Kashaf *et al.*, 2023). A metagenomic study from the US revealed that strains of *S. aureus* from patients with severe AD elicited stronger inflammatory and immune responses in vitro compared to strains from individuals with mild AD or healthy skin (Byrd *et*

al., 2017). A study from the UK showed over 80% of strains from an AD patient had acquired a plasmid carrying the beta-lactamase gene (*blaZ*) conferring penicillin resistance, a feature absent in the HE group (Catriona P. Harkins *et al.*, 2018). A cohort study from Japan detected dysfunctional mutations of the accessory gene regulatory (*Agr*) quorum-sensing system predominantly in the HE group, highlighting a connection between a functional *Agr* system and AD (Nakamura *et al.*, 2020). Recent insight has shown that adaptive capsule loss via *capD* mutations, crucial for initiating the *S. aureus* capsular polysaccharide synthesis pathway, was more common in AD worldwide (Key *et al.*, 2023), which provided a competitive advantage of improved adherence to AD skin. Comparative genomics of *S. aureus* in Japan discovered that only $ST97_A$, a subclade of ST97 detected in AD lesional skin, possessed the complete tryptophan (*trp*) operon, enabling bacterial survival without exogenous *trp* on AD skin, where the *trp* level was significantly reduced (Obata *et al.*, 2023). While these studies highlight the genetic variations between AD and HE strains, with only Key *et al*. emphasizing the need for a global analysis (Key *et al.*, 2023), the identified variations to date are mostly localized, suggesting sporadic adaptations rather than widespread differences. Consequently, it is imperative to adopt a global perspective to investigate whether consistent genomic differences exist between strains from AD and HE groups.

To understand the association of *S. aureus* strains with AD, various genotyping methods have been employed. Pulsed Field Gel Electrophoresis (PFGE) has been used to genotype *S. aureus* strains (He *et al.*, 2014), assigning each strain to a specific pulsotype (e.g., A, B, C). In one study, pulsotype B (48%) was the most frequent AD-related isolate type (Lo *et al.*, 2010), while another study indicated additional 28 pulsotypes could be assigned to AD patients (Lomholt *et al.*, 2005). Multi-locus sequence typing (MLST) method indicated that ST188 (19.4%) and ST1 (13.9%) were the most AD-related (Kim *et al.*, 2009), while a recent extensive analysis revealed differential ST predomination in AD in various geographical regions on a global scale (Saheb Kashaf *et al.*, 2023). A comprehensive genomic characterization of *S. aureus* isolated from the Japanese cohort revealed that ST diversity was pronouncedly lost in severe AD skin compared to healthy skin, with ST188 as the most frequently detected ST in lesional AD skin and ST30 only detected in healthy skin (Obata *et al.*, 2023). Based on CC typing, Yeung *et al.* showed that CC45 (21%), CC5 (14%), and CC15 (14%) were among the most common CCs in AD patients (Yeung *et al.*, 2011). Rojo *et al.*, demonstrated CC5 (31.2%), CC15 (18.7%), and CC30 (18.7%) as the most frequent CCs in AD, whereas CC30 (48.3%) was most prevalent in the control group (Rojo *et al.*, 2014). Harkins *et al.* proved CC1 (20%) as the most prevalent in AD whereas CC30 (33%) and CC45 (22%) were predominant in the noses of the healthy cohort (Catriona P. Harkins *et al.*, 2018). Key *et al.* identified CC30 (33%) as the largest fraction in AD (Key *et al.*, 2023). Although these methods link specific *S. aureus*

strains to the AD condition, the highest prevalence of any clone or type associated with AD identified did not exceed 50%, which underscores their limited effectiveness in distinguishing AD-related *S. aureus* strains. Consequently, there is a critical need for more powerful techniques to differentiate *S. aureus* strains linked to AD from those in healthy skin.

Phylogenetic analysis revealed the evolutionary trajectory of *S. aureus* strains in humans and animals, showing coevolution with humans and multiple human-to-animal transmission events (Weinert *et al.*, 2012; Shepheard *et al.*, 2013). These transmissions resulted in distinct lineages adapting to new animal hosts (Weinert *et al.*, 2012), with animal-to-human transmissions being rarer (Spoor *et al.*, 2013). Moreover, recent decades have seen the emergence of MRSA strains exhibiting low host specificity, particularly CC130 and CC398 (Price *et al.*, 2012). Therefore, investigating the association between human-derived *S. aureus* strains (from both AD and healthy individuals) and animal-origin strains, along with their potential for zoonotic transmission via host-switching, offers critical insights into the evolutionary patterns and adaptability of this pathogen.

Overall, extensive research utilizing different methods have documented genetic variations in *S. aureus* strains that are closely associated with AD, indicating these strains harbor unique traits that enable them to thrive on AD skin and contribute to AD pathogenesis. These findings suggest that these variations could influence AD at multiple levels, including clonal differences, gene variations, or gene polymorphisms. Despite these insights, the intrinsic nature of these strains—whether inherently pathogenic or evolved from commensal strains—remains elusive. Hence, deepening our understanding of *S. aureus*'s genetic diversity is essential to elucidate its role in AD pathogenesis. As more evidence emerges, we inch closer to fully understanding the complex relationship between *S. aureus* and AD.

1.3 Role of prophages in *S. aureus*

Bacteriophages, often simply referred to as phages, are viruses that infect and replicate within bacteria. These viruses specifically target bacteria, injecting their genetic material and hijacking the bacterial machinery to replicate. Yet, the relationship between phages and bacteria is more nuanced than a simple predator-prey dynamic. Over millennia, this interaction has profoundly influenced the evolution and ecology of bacterial communities in myriad ways. As the most abundant organisms on earth, phages play a pivotal role in continuously shaping the diversity, richness, abundance, and function of microbial communities (Clokie *et al.*, 2011; Koskella and Brockhurst, 2014; Naureen *et al.*, 2020; Chevallereau *et al.*, 2022). For bacteria, phages are not just threats but also agents of change. As depicted in Figure 4, they can introduce new genes into bacterial populations, bestowing them with novel abilities—from enhanced metabolic functions to resistance against antibiotics (Howard-Varona *et al.*, 2020). Furthermore, by controlling bacterial numbers, phages help maintain the balance of microbial ecosystems, be it in the soil (Graham *et al.*, 2019), ocean (Suttle, 2007), or even the human (Hoyles *et al.*, 2014; Santiago-Rodriguez *et al.*, 2015). Their ability to specifically target certain bacteria also holds promise for therapeutic applications (Durr and Leipzig, 2023), especially as traditional antibiotics face challenges of resistance. Understanding the multifaceted roles of phages in bacteria is thus vital for both fundamental microbiology and its applications in medicine and biotechnology.

Due to the significance of *S. aureus* in human health and disease, the relationship between phages and *S. aureus* is intricate. Understanding the dynamic interplay between phages and *S. aureus* offers a unique lens to view the diversity and, more importantly, the pathogenicity of *S. aureus* (Gummalla *et al.*, 2023). *S. aureus* is known for its ability to cause a range of diseases, from minor skin infections to life-threatening conditions like septicemia (Lowy, 1998; Foster, 2005; Knox *et al.*, 2015; Lee *et al.*, 2018; Turner *et al.*, 2019). The extent to which phages enhance or mitigate *S. aureus* functional and pathogenic potential is both biologically and clinically important, which might have direct implications for therapeutic interventions, especially in an age where antibiotic resistance is a mounting challenge.

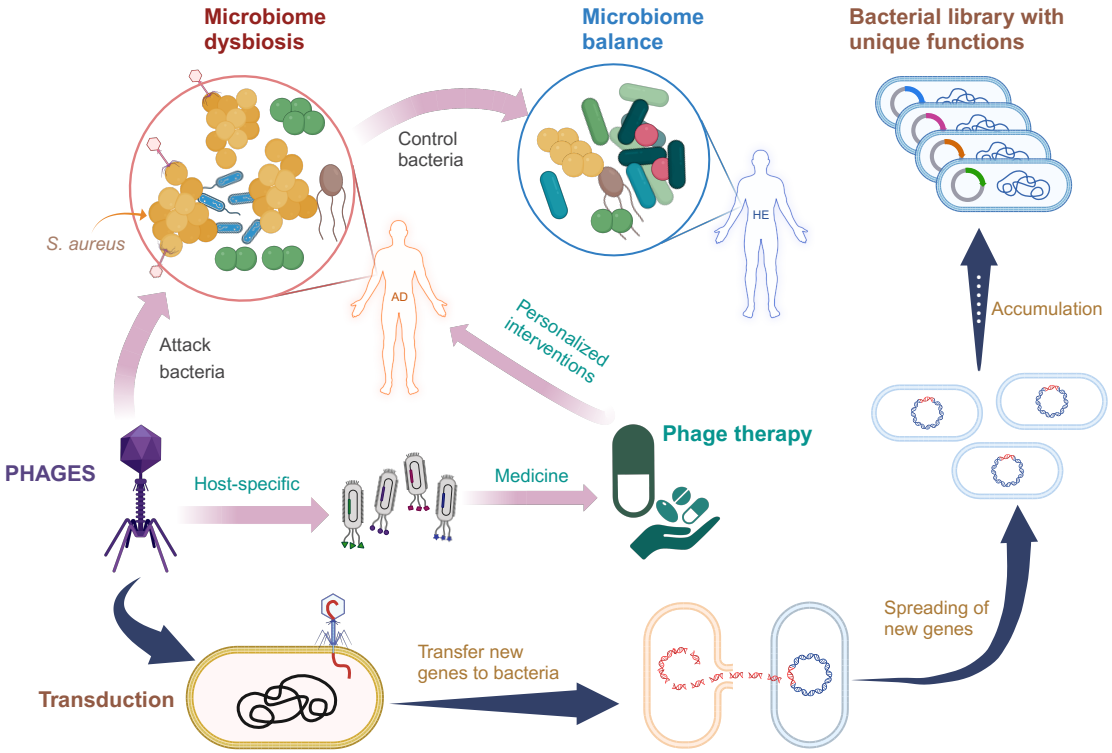

Figure 4. Role of phages in bacterial regulation and medical applications. Phages exhibit host-specific characteristics, allowing them to specifically target and attack their hosts. This action results in the control of bacterial numbers within the community. Additionally, this specificity can be harnessed to develop phage therapy for personalized medical interventions, with the potential for correcting microbiome dysbiosis and promoting microbiome balance. Moreover, this diagram outlines how prophages attack bacteria, facilitating the transfer of new genes which contributes to the creation of a bacterial library with unique functions. Transduction begins when transducing phages invade target bacteria. A subsequent infection cycle leads to the DNA from the transducing phage particle being transmitted to, and incorporated by, a recipient strain. This process transfers new genes to the bacteria, creating a transductant. Following bacterial reproduction, the new genes received from phages are spread. The accumulation of this process eventually leads to a bacterial library with strains possessing distinctive metabolic functions, enhancing their utility in medical and biotechnological applications. AD, atopic dermatitis; HE, healthy individuals.

1.3.1 Phage-mediated control of bacteria numbers

Phages possess a remarkable potential for predation and play a crucial role in regulating bacterial population dynamics via different strategies, profoundly influencing bacterial abundance and diversity (Voigt *et al.*, 2021). One of the most pronounced impact of phages on bacteria is a significant reduction in their numbers (Breitbart and Rohwer, 2005). Phages are estimated to be 20-30% more abundant than bacteria, and to sustain their populations, they are responsible for an astounding 10^{24} infections per second (Fuhrman and Noble, 1995). These infections result in the elimination of approximately 4-50% of the total bacterial population in natural environments at a given moment (Fuhrman and Noble, 1995). For example, viral lysis eliminates about 20-40% of the prokaryote population in surface waters daily, making it as significant as grazing in terms of causing microbial mortality (Suttle, 1994, 2007).

Beyond their lytic actions, phages also curtail bacterial numbers through chronic infections (lysogenic), which slow bacterial growth rates, thus benefiting faster-growing bacterial species and altering the microbial community richness (Bikard and Marraffini, 2012; Parsons *et al.*, 2012). Additionally, lysogenic bacteria can confer resistance against lytic phages, providing a competitive advantage. However, the evolution of bacterial resistance to phages ensures that an ecological balance is maintained between them (Labrie *et al.*, 2010; Parsons *et al.*, 2012).

Phages control bacterial populations by alternating between lysogenic cycles, where they integrate into the host DNA and replicate passively, and lytic cycles, where they actively reproduce and destroy the host cell (Voigt *et al.*, 2021), in which holins, encoded in bacteriophage genomes, act as a clock function determining the timing of the end of the infection cycle (Wang *et al.*, 2000; Saier and Reddy, 2015). Holins create pores in bacterial cytoplasmic membranes to facilitate the release of endolysins that break down the cell wall and trigger bacterial cell death, thereby promoting virion release (Young and Bläsi, 1995). In the context of the influence of phages on *S. aureus* in AD, it has been proved an *S. aureus* phage can effectively reduce the *S. aureus* abundance and ameliorate the disease exacerbation (Shimamori *et al.*, 2020).

1.3.2 Phage-mediated horizontal gene transfer

S. aureus is a notable human pathogen known for its adaptability. The genetic variability in bacterial populations is driven by mutation rates and HGT, the latter of which is predominantly facilitated by mobile genetic elements (MGEs) like temperate bacteriophages (Rankin *et al.*,

2011). Unlike the very conservative core genome, the variety of MGEs in *S. aureus* shows significant variability. Around 15-20% of the genome consists of MGEs, which encompass bacteriophages, pathogenicity islands, plasmids, transposons, integrative conjugative elements (ICEs), and staphylococcal chromosome cassettes (SCCs) (Lindsay, 2010; Alibayov *et al.*, 2014). *S. aureus* quickly develops ARGs, primarily by acquiring resistance genes and toxin genes from other *S. aureus* strains or different genera (Chen and Novick, 2009; Haaber *et al.*, 2017). While these genes are transferred through various MGEs, bacteriophage transduction has been suggested as the primary HGT mechanism in *S. aureus* (McCarthy *et al.*, 2014).

S. aureus strains typically contain 1-4 bacteriophages integrated into their genome as prophages, playing an important role in the *S. aureus* evolution and functional capabilities (Goerke *et al.*, 2009; Lindsay, 2014; McCarthy *et al.*, 2014; Xia and Wolz, 2014). These prophages can spontaneously detach or be induced by factors such as DNA damage, leading to a lytic cycle where the bacterium produces phage progeny and undergoes lysis. This cycle can result in bacterial DNA being packaged in a phage capsule, a process known as generalized transduction (ZINDER, 1955). Contrary to previous beliefs that transduction was an error, recent findings suggest that different antibiotics can vary the ratio of transducing particles (Stanczak-Mrozek *et al.*, 2017). Transduction utilizes specific sites on the bacterial chromosomal and plasmid DNA and can transfer around 1.5% (~45 kbp) of the *S. aureus* genome (Morse, 1959; Ubelaker and Rosenblum, 1978). Besides, phages also exhibit strain-specific transduction. The phage Φ187, for instance, specifically transduces *S. aureus* strains of type ST395, which possess a unique teichoic acid in their cell walls (Winstel *et al.*, 2013). Interestingly, it can also transduce other staphylococcal species (Uchiyama *et al.*, 2014) and even *Listeria monocytogenes* (Chen and Novick, 2009; Chen *et al.*, 2015), which share a similar cell wall structure. This transduction involves DNA injection and adsorption, but not the typical lytic action of the phages, resulting in phage-mediated intergeneric transfer of toxin genes (Chen and Novick, 2009). Thus, in addition to being a highly useful genetic tool, generalized transduction is also a major driver of *S. aureus* evolution and the spread of ARGs.

Given their host specificity and ability to carry a diverse array of functional elements (Koskella and Meaden, 2013), bacteriophages can profoundly influence the adaptability and pathogenicity of *S. aureus* strains. Such genetic variation not only enhances bacterial abilities to adapt to broader conditions but also leads to significant clinical implications. Previous studies showed that phage-bacteria interactions play a decisive role in bacterial diversity, virulence, and evolution (Naureen *et al.*, 2020) and accessory genes within phage serve as major determinants for *S. aureus* evolution (Xia and Wolz, 2014). Despite these growing

insights into the genetic diversity of *S. aureus* strains, there remains a significant gap in our understanding of *S. aureus* prophages in the context of AD. Therefore, focused efforts to profile these prophages are essential to explore the intricate relationship between genomic composition, phage-driven evolution, and the resulting phenotypic outcomes.

1.3.3 Prophage influence on *S. aureus* pathogenicity

Prophages wield a significant influence over the pathogenicity of *S. aureus*, which is crucial in the toxin production (Xia and Wolz, 2014; Zeaki *et al.*, 2015), lysogenic conversion (Brüssow *et al.*, 2004), immune resistance, and overall virulence, thereby affecting the adaptability and pathogenicity of *S. aureus* (Hatoum-Aslan, 2021) and holding substantial implications for disease treatment and prevention.

In *S. aureus*, lysogenic conversion, the process of incorporating prophage genetic material into *S. aureus* genome, often results in the addition of VFs, significantly augmenting the bacterium's ability to infect and cause disease in hosts. This process not only enhances the pathogenic potential of individual bacterial cells but also contributes to the genetic diversity and adaptability of *S. aureus* populations as a whole (Nepal *et al.*, 2021). Phages contribute notably to the pathogenicity of *S. aureus* through the introduction of toxin-producing genes. For example, prophages within *S. aureus* can possess genes responsible for various VFs, including Panton-Valentine leucocidin (PVL) toxin (Shallcross *et al.*, 2013), the cell-wall anchored protein SasX (Li *et al.*, 2012), staphylokinase (*sak*), enterotoxins (SEs), chemotaxis-inhibitory proteins (*chp*), exfoliative toxins (*eta*), and the staphylococcal complement inhibitor (*scn*) (De Haas *et al.*, 2004; Van Wamel *et al.*, 2006).

PVL is a potent toxin linked to severe skin infections (Zanger *et al.*, 2012; Shallcross *et al.*, 2013). The genes encoding PVL are situated on a prophage, thus exemplifying the pivotal role of phages in determining the toxin profile and, consequently, the virulence of *S. aureus* strains. Recent data suggest that these skin infections could lead to more severe bloodstream infections (Tattevin *et al.*, 2012). Moreover, these phages facilitate the methodical transfer of pathogenicity islands and play a central role in the horizontal transmission of both chromosomal and extra-chromosomal genes (Xia and Wolz, 2014).

SasX, a new cell wall-anchored VF, enhances nasal colonization, bacterial aggregation, and virulence in these bacteria. The gene for SasX is located as an accessory gene in a phiSPß-like prophage, which is notably larger than a typical siphovirus of *S. aureus* and resembles a prophage found in *S. epidermidis* strain RP62A (Holden *et al.*, 2010), which suggests that the

phage, and thus the new virulence trait, was acquired from *S. epidermidis*. Now the phage is also present in MRSA strains of the CC5 lineage by spreading among *S. aureus* strains (Li *et al.*, 2012). The emergence of MRSA strains carrying a phage that harbors SasX highlights the role of phages in spreading bacterial clones (Li *et al.*, 2012).

ETA causes bullous impetigo and its widespread variant, staphylococcal scalded-skin syndrome (SSSS). The ETA gene is housed in Sa1int phage genomes (Goerke *et al.*, 2009; Kahánková *et al.*, 2010), which have been linked to MRSA and Methicillin-sensitive *S. aureus* (MSSA) strain outbreaks of various CCs in countries like Japan and the Czech Republic (Yamaguchi *et al.*, 2002; Shi *et al.*, 2011; Růžičková *et al.*, 2012), and are found in a subset of CC121 strains associated with superficial infections (Kurt *et al.*, 2013). Infections with strains containing ETA-phages have manifested in varied clinical symptoms, ranging from blisters to more complicated conditions such as conjunctivitis and SSSS (Růžičková *et al.*, 2012).

While *sak* is typically found on Sa3int phages, it can also occasionally be located on Sa7int phages, such as in derivatives of the laboratory strain 8325-4 (Goerke *et al.*, 2006), which have been widely distributed worldwide in various labs. Sak-Sa7int phages are prevalent in numerous prominent hospital-associated MRSA clones in southern Germany (Schulte *et al.*, 2013), specifically those related to the ST5 lineage. These strains have proliferated and diversified rapidly over the years, revealing variations in PFGE patterns and antibiotic susceptibility. It's speculated that accessory elements, particularly varying phage content, might drive the proliferation of these strains, as no evident mutations associated with this characteristic have been identified in the core genome (Nübel *et al.*, 2008).

The heat-stable staphylococcal SEs play a crucial role not only in triggering toxic shock-like syndromes and causing food poisoning but also act as superantigens that promote T-cell proliferation, a function closely linked to the previously mentioned effects (Harris *et al.*, 1993; Pinchuk *et al.*, 2010). The gene encoding the *eta* toxin is located at the terminus of Sa1int phages, while genes *sea*, *sak*, *scn*, and *chp*, which are part of the immune evasion gene cluster (IEC), are situated at the end of Sa3int phages (Xia and Wolz, 2014). The gene for SEA is found within the genome of Siphoviridae bacteriophages. SEA is responsible for about 80% of Staphylococcal food poisoning cases and presents a challenge in regulation as its control does not depend on the *agr* or the staphylococcal accessory regulator, unlike most SEs (Hennekinne *et al.*, 2012; Sospedra *et al.*, 2013). A specific group of serotype F bacteriophages found in *S. aureus*, including strains from MRSA isolated in Irish hospitals and the historical strain PS42-D, have been shown to induce simultaneous expression of SEA, Sak, and β-lysin. The genes for SEA and Sak are closely associated and located near the phage attachment site (Coleman *et al.*, 1989).

Prophages may possess auxiliary metabolic genes (AMG) that are homologous to bacterial metabolic genes and these AMGs might alter bacterial metabolism in favor of infecting phages (Clokie *et al.*, 2011). Studies have shed light on the role of AMGs in contributing to the ecological fitness of phage-host systems (Thompson *et al.*, 2011; Q. Wang *et al.*, 2022), such as dUTP pyrophosphatase that is indicative of their similar capacities for accurate DNA replication, GT27 polypeptide alpha-N-acetylgalactosaminyltransferase, small subunit ribosomal proteins S6 and S18, and DNA-methyltransferase that are responsible for both protein or lipid modification and protein synthesis, regulation of gene expression, or evading host defense mechanisms, respectively (Thompson *et al.*, 2011; Q. Wang *et al.*, 2022). These genes could enhance the metabolic capabilities of host strains and confer them a greater selective advantage on more diverse environmental challenges.

Deep insights into how prophages influence *S. aureus* pathogenicity are pivotal for crafting effective treatment strategies. As phages introduce or amplify VFs, they offer fresh prospects for targeted medical interventions. Simultaneously, as the world grapples with escalating antibiotic resistance, there is a rejuvenated interest in using phages directly as a therapeutic measure, known as phage therapy (York, 2022; Strathdee *et al.*, 2023). An in-depth understanding of this relationship is not merely academically significant but is crucial for developing innovative and effective strategies for treating and preventing *S. aureus* infections in clinical settings. This knowledge is particularly pertinent in an era where antibiotic resistance is undermining the efficacy of traditional treatment approaches, necessitating the exploration and adoption of alternative therapeutic modalities.

1.4 Study aims and hypotheses

S. aureus harbors significant diverse strains, with varying genetic compositions and functionalities that have different associations with the severity of AD. However, the mechanisms by which certain strains isolated from AD exacerbate disease symptoms more than strains from HE remains largely underexplored. It also remains unclear whether *S. aureus* strains in the context of AD exhibit distinctive differentiation between global and local scales. Given the significant proportion and increasing trend of the AD population in the world, it is crucial to dissect whether systematic genome differences exist between *S. aureus* strains from AD and HE groups and between local and global scales, to pinpoint the genetic markers and functions that distinguish *S. aureus* strains associated with AD, and to understand the role of HGT and prophages in the virulence and antibiotic resistance of *S. aureus* strains. Addressing these gaps requires a concerted effort to utilize advanced genomic sequencing technologies, develop more effective computational methods, and employ various cutting-edge bioinformatic techniques to identify specific genetic indicators efficiently.

Whole-genome sequencing (WGS) offers an unprecedented resolution for investigating the subtle genetic differences among *S. aureus* strains in clinical contexts on the strain level. To elucidate the impact of *S. aureus* strain diversity on AD development and its treatment implications, we have cultured and deeply sequenced the whole genomes of 48 *S. aureus* strains from AD patients and HE (33 AD strains vs. 15 HE strains) from a cohort at Klinikum Augsburg, Germany, leveraging the third sequencing generation (TGS) technology by PacBio. Besides, we also collected a global dataset of 300 publicly available *S. aureus* strains (150 AD strains vs. 150 HE strains) from nine different countries to conduct an in-depth, *in silico* comparison of genomic and functional disparities at both local and global scales. To effectively distinguish between closely related *S. aureus* strains from AD and HE, we have developed and deployed a state-of-the-art machine learning (ML) model designed to identify potential marker genes. This approach not only allows for a comprehensive understanding of the strain-specific genetic factors contributing to AD, but also facilitates the development of novel diagnostic tools and targeted therapeutic strategies to manage this challenging skin condition more effectively.

Therefore, this thesis focuses on the following investigations:

1 As AD skin represents a unique microenvironment, we hypothesize it might promote adaptive selection towards a homogenous set of *S. aureus* strains, which make AD strains genetically and thus functionally differ from HE strains.

2 To address the knowledge gap resulting from most variations being observed locally, this study investigates the genomic and functional differences between *S. aureus* strains from AD and HE across global and local contexts. We aim to elucidate whether the genomic features of *S. aureus* strains from HE and AD share a similar biogeographical pattern and if not, what mechanisms drive the differentiation in terms of both health status (AD vs. HE) and geographical locations (global vs. local). We corroborate our global analysis with a local collection of 48 strains isolated from Augsburg, Germany.

3 Given the critical role of AMRs in the adaptation and pathogenicity of *S. aureus*, we hypothesize that many AMRs, frequently encoded within MGEs, are introduced into *S. aureus* through HGT events from bacterial species originating in environmental or animal reservoirs.

4 Because *S. aureus* strains cannot be differentiated based on the 16S rRNA gene-based and whole genome-based phylogenetic analysis in terms of health status, we hypothesize the observed differences in AD-associated strains could be attributed to specific marker genes, which could robustly differentiate AD and HE strains. To address the knowledge gap, we developed a random forest (RF) model in this study for the accurate identification of marker genes differentiating AD-related *S. aureus* strains from those in HE.

5 To assess the functional association of human-derived *S. aureus* strains and those originating from animals and the potential of zoonosis through host-switching transmission, we collected *S. aureus* strains from other mammalian hosts to investigate if *S. aureus* strains follow host-specific adaptation. Additionally, we further hypothesize that the marker genes identified using the RF classifier are also able to differentiate *S. aureus* strains from various origins of hosts.

6 Prophages can profoundly influence the genetic and functional variations, pathogenicity, and adaptability of their bacterial hosts *S. aureus* strains. To fill in the significant gap in our understanding of prophages of *S. aureus* and potential links to AD, we profile the prophages of *S. aureus* to explore the intricate relationship between genomic composition, phage-driven adaptation, and the role of prophages in the pathogenicity of AD strains. We hypothesize prophages play an important role in the differentiation of AD and HE strains.

2 Materials and Methods

2.1 Isolation, sequencing, and assembly of 48 strains from Augsburg

Here we cultured and sequenced 48 genome sequences of *S. aureus* strains, including 44 genomes with complete chromosomes and 4 genomes with incomplete chromosomes, isolated from the skin and nose of 10 subjects from a study cohort established at the Chair for Environmental Medicine of the University Center for Health Sciences at the Klinikum Augsburg, Germany. The study was approved by the ethics committee of the Technical University of Munich (112/16 S and 187/17 S). All participants signed informed consent statements.

The isolation of *S. aureus* was performed by our collaborators from Augsburg, Germany. Skin and nose swabs (Transwabs, Medical Wire, Wiltshire, England) were sampled from June 2018 to March 2019; potential *S. aureus* isolates were retrieved by initial streaking and subsequent purification on Mannitol-salt agar plates (Carl Roth, Karlsruhe, Germany). Identity was confirmed by MALDI-TOF after growth on Columbia blood agar plates with 5% sheep blood (Oxoid Ltd, Basingstoke, UK). The Genomic-tip kit 20/G (Qiagen, Hilden, Germany) was used for genomic DNA extraction from approximately 4.5 x 10^9 cells (equivalent to an OD_{600} of 0.5) freshly grown in LB broth (Carl Roth) at 37 °C for approximately 4 - 5 hours. Cells were harvested by centrifugation at 3,270 x g for 10 minutes. For extraction lysozyme was replaced by lysostaphin (20 µl of a 10 mg/ml stock solution, Sigma-Aldrich, 2019, Product L9043) for more efficient cell lysis.

Genome sequencing was carried out on a PacBio Sequel Platform (Pacific Biosciences, Menlo Park, USA) using Sequel Binding Kit 3.0, Sequencing Plate 3.0 (4 rxns), and SMRT cell 1M v2 (PacBio). Library preparation was performed by applying the SMRTbell Express Template Prep kit 2.0 and the Barcoded Adapter Kit 8A (PacBio) according to the protocol provided by the manufacturer. Genomic DNA was sheared to ~ 10 kb using g-TUBEs (Covaris Inc., Woburn, MA) and processed without additional size selection. SMRTBell® libraries (4 pools, each including 12 strains) were loaded according to the diffusion loading protocol (PacBio) at a concentration of 5 pm. Movie time was 10 h per SMRT cell after immobilization for 2 h and pre-extension for 2 h.

After data demultiplexing, genome assembly was performed with the HGAP4 pipeline as embedded in SMRTLink version 6.0.0.47841 (PacBio) utilizing 95,473 to 497,238 realigned subreads, with an average subread length of 3,707 to 5,765 bp. The mean genome coverage was 145 to 918-fold. For the remaining incomplete assemblies, the genome assembly and high-quality circular consensus sequence (CCS) reads generated by SMRTLink were used for

circularization by Circlator version 1.5.5 with the parameter --merge_min_length_merge 1000 (Hunt *et al.*, 2015). All complete chromosomal scaffolds were rotated to dnaA as the start position. If no dnaA gene was found for the incomplete assemblies, then Prodigal V2.6.3 (Hyatt *et al.*, 2010) was used to identify the gene nearest the center of the contig.

Genome assemblies of these 48 strains have been deposited in the DDBJ/ENA/GenBank databases under the following accession numbers: CP077852 - CP077932 (Z. Wang *et al.*, 2022). Raw reads (SAMN19550019 - SAMN19550066) have been deposited in the Sequence Read Archive (SRA) under the BioProject accession number PRJNA732579. Detailed information on all the 48 Augsburg strains is included in Table S2.

The qualities of all *S. aureus* genomes used in this thesis were assessed using CheckM v1.1.3 (Parks *et al.*, 2015). Default parameters were used for all tools described below unless otherwise specified.

2.2 Collection of the global dataset of 300 *S. aureus* from AD and HE

A total of 300 genomes of *S. aureus* strains from AD and HE were collected from public databases representing a broad global spectrum of sites where strains were originally isolated. For *S. aureus* strains from AD, we searched for *S. aureus* genomes from the NCBI BioSample database using the keywords “Atopic dermatitis and *Staphylococcus aureus*” (1,707 results, March 3, 2023); subsequently, these results were exported in the format “Accessions list”; these accession numbers were matched with the metadata of all *S. aureus* assemblies downloaded from the NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/154/). Furthermore, the matched assembly accession numbers were used to retrieve *S. aureus* genome sequences using ncbi-genome-download v0.3.1 with parameters “bacteria --section genbank --formats fasta --flat-output” (156 assemblies downloaded). After quality control (completeness >98%, contamination <3%), 150 *S. aureus* genomes from AD were retained for further analysis.

For *S. aureus* strains from HE, since there were not enough *S. aureus* assemblies with qualified metadata available on the NCBI Genome database, we searched the BioSample database with the keywords “Staphylococcus aureus and healthy skin and Homo sapiens” (364 results, March 3, 2023). The search results were exported and the Sequence Read Archive (SRA) accession numbers were used to download whole-genome sequencing raw read samples using Prefetch v2.8.0 (Sherry *et al.*, 2008), obtaining 338 SRA samples. These samples were then processed to split forward and reverse reads using Fasterq-dump v2.8.0 (Sherry *et al.*, 2008) and manually assembled using Spades v3.13.0 with parameters “--careful

-k 55,77,99,127 --cov-cutoff auto" (Bankevich *et al.*, 2012). After quality control, 209 assemblies were retained. To maintain a comparable number of genomes to the AD strains for further analysis, we subsampled 150 *S. aureus* genomes from HE.

Detailed information on the 300 public *S. aureus* strains is included in Table S1.

2.3 Collection of the public dataset of *S. aureus* from mammalian hosts

To investigate the association of human-derived and animal *S. aureus* strains, *S. aureus* assemblies from other mammalian animals were collected from the BV-BRC database (https://www.bv-brc.org, August 15, 2023). We selected *S. aureus* genomes isolated from cows, dogs, horses, pigs, sheep, and yaks due to their high availability which could ensure robust results. The accession numbers of the selected strains were used to retrieve genome sequences using Batch Entrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez). Genomes with abnormal sizes deviating from the normal size of *S. aureus* at ~2.8 Mb or with high contamination (>10%) were excluded, ensuring a comprehensive selection of high-quality *S. aureus* genomes for our study. After quality control, we included 333 strains from cows, 56 strains from dogs, 145 strains from horses, 391 strains from pigs, 34 strains from sheep, and 83 strains from yaks for the analysis. Detailed information on all mammalian strains is included in Table S3.

2.4 Orthologous gene clustering

Open reading frames (ORFs) of all genomes were predicted using Prodigal v2.6.3 with the parameters "-i -a -d" (Hyatt *et al.*, 2010). All ORFs were aligned with each other using BLASTP version 2.12.0+ with commands -evalue 1e-5 (Camacho *et al.*, 2009). BLAST result was then filtered to a percent identity of 70% and query coverage of 75% (Oyserman *et al.*, 2016). Then the three columns, including query genes, subject genes, and e-value were retained using Cut -f 1,2,12. The result was further processed using Mcxload with settings "-abc --stream-mirror --stream-neg-log10 -stream-tf 'ceil(200)' -o -write-tab". Finally, orthologous gene clusters were identified using MCL version 14-137 with an optimal inflation value of 2 (Schaeffer, 2007).

2.5 Pan-genome and core genome calculation

The gene abundance table of the 48 Augsburg *S. aureus* strains and the 300 global strains generated by MCL version 14-137 above was transformed into a gene presence-absence table (0/1 matrix). The binary 0/1 matrix was then used as the input file to calculate the pan-genome

and core genome using PanGP v1.0.1 with settings “sample algorithm Distance Guide, sample size 500, sample repeat 10, amplification coefficient 10 (Zhao *et al.*, 2014). Fitting functions of the core genome and pan-genome were also calculated with PanGP. The calculated pan-genome and core genome were outputted and the accumulation curves were visualized using ggplot2 (Wickham, 2011).

Core genes are defined as those that are present in more than 95% of strains; accessory genes are those that are present in more than one strain but not core; unique genes are those that are only present in a single strain. The statistical significance between AD and HE strains across core, accessory, and unique genes was calculated using the Wilcoxon rank sum test, following the normality test using the Shapiro-Wilk test, respectively.

2.6 Calculation of within-group distance

Within-group distance for the global AD strains and HE strains in terms of health status was calculated using Vegan 2.6-4 with the method Bray-Curtis, respectively (Dixon, 2003). Their normality was tested with the Asymptotic one-sample Kolmogorov-Smirnov test as there were over 5000 observations, followed by the Wilcoxon rank sum test to assess the statistical significance.

Within-group distance between different countries of origin was calculated using Vegan 2.6-4 with the method Bray-Curtis (Dixon, 2003), respectively. Only the four countries with more than 10 strains were used for comparison. Their normality was tested with the Shapiro-Wilk test, followed by the calculation of statistical significance using the Wilcoxon rank sum test.

2.7 Functional prediction of *S. aureus* genomes

Gene functions of all *S. aureus* strains were annotated using both the GhostKOALA version 2.2 (Kanehisa *et al.*, 2016) searching the “genus_prokaryotes + viruses” database online (April 1, 2022). The genomes of the main cluster and Cluster 2 were also functionally annotated using Prokka v1.14.6 with settings --genus Staphylococcus --gram + --evalue 1e-05 (Seemann, 2014).

The gene abundance for each KEGG functional category in each genome was calculated using a custom Python script (count_ko_per_genome.py) with genes functionally unannotated assigned into “unassigned”. based on the abundance table, the Principal Component analysis

(PCA) of differentiating functions was performed with R package Vegan v2.5-6 (Dixon, 2003). The top differentiating functional categories were calculated using the envfit function with 999 permutations with only the top five shown.

The additional two clusters were singled out to compare their average gene abundance in selected functional categories with the main cluster. For Cluster 1, the average gene copies of each gene function within the featured Replication and repair pathway were examined and compared with the main cluster. For Cluster 2, the average gene copies of each gene function within the other featured pathways, including Transporters, Quorum sensing, Two-component system, and Antimicrobial resistance genes (AMR) were examined and compared with the main cluster. Gene functions with significant differential copies were selected for further analysis.

The lantibiotic operon synteny was visualized in Easyfig 2.2.5 (Sullivan *et al.*, 2011), for which the input files were annotation files in the gbk format outputted by Prokka v1.14.6 (Seemann, 2014). Four representative strains were selected for comparison, with one AD strain (GCA_015714515) and one HE strain (ERS5272160) from the main cluster and two AD strains (GCA_020906905, GCA_015713015) from Cluster 2. Only the subregion of the lantibiotic operon and flanking sequences were compared.

2.8 Phylogenetic analysis

The phylogenetic tree based on the accessory genes of the 48 Augsburg strains were constructed using Roary version 3.11.2 with the parameters "-i 70 -cd 95 -iv 2.0 (Page *et al.*, 2015). The input data for Roary (Page *et al.*, 2015) was the gff3 files generated by Prokka v1.14.6 (Seemann, 2014).

The phylogenomic tree based on the 16S rRNA gene and whole genome for the 48 Augsburg strains was constructed using the TYGS (Type Strain Genome Server, https://tygs.dsmz.de) (Meier-Kolthoff and Göker, 2019).

The *nisB* gene, the lantibiotic biosynthesis protein, was used for the phylogenetic analysis of the lantibiotic operon as it is the longest gene in the operon. All *nisB* gene sequences of the *S. aureus* strains from this study were extracted by blasting all genes to the known *nisB* gene (WP_001092605) using BLASTP v2.12.0+ with the 70% of identify and 75% of coverage (Altschul *et al.*, 1990). Then all extracted genes were blast against the online Non-redundant database and the top 1000 best hits were outputted for further processing. Subsequently, the outputted *nisB* gene sequences were aligned using Mafft v7.310 (Katoh and Standley, 2013)

utilizing the parameters “--maxiterate 1000 –localpair” to maximize iterative refinement. All gap positions in the alignment with gaps in 50% or more of the sequences were removed using TrimAl v1.4.rev22 (Capella-Gutiérrez *et al.*, 2009) with –gt 0.5. This curated alignment was used to infer the phylogenetic tree using RAxML version 8.2.11 with the PROTGAMMAAUTO model and 200 bootstraps (Stamatakis, 2014). All phylogenetic trees were polished using iTOL (https://itol.embl.de) (Letunic and Bork, 2021).

The phylogenetic tree were constructed using CVTree version-3.0.0 (Zuo, 2021) based on the whole genomes of the 348 *S. aureus* strains in this study to show the clonal structure. Three genomes from *S. argenteus*, *S. epidermitis*, and *S. schweitzeri* were used as the outgroup.

2.9 Development of random forest classifier

The random forest (RF) classifier, a supervised machine learning model, was built to predict health status (AD and HE) of *S. aureus* genomes based on the presence-absence table of the orthologous gene family obtained from MCL version 14-137. This pipeline was built and optimized based on the Microbiome Helper (Comeau *et al.*, 2017), including preprocessing, performance evaluation, feature selection, and prediction. Initially, the *S. aureus* strains from public datasets were partitioned into the training and test datasets in various proportions (5:5, 6:4, 7:3, 8:2, and 9:1) across 10 iterations for each. The training dataset was subsampled randomly without replacement from the 300 public strains and the remaining strains were used as the test dataset. The subsampling was repeated 10 times for cross-validation to get more robust predictions. Based on the results of each partition, the optimal training-test partition ratio was determined (9:1). Then the training dataset and test dataset were used for the classifier construction and evaluation, respectively. The 48 *S. aureus* strains sequenced by our group were used as the real-world test dataset to further verify the performance of the classifier.

The model’s performance was robustly evaluated using multiple metrics, including precision, recall, F1-score, accuracy, and area under the curve - receiver operating characteristics (AUC-ROC). Recall, also known as sensitivity, is the fraction of relevant instances that were retrieved; Precision, or specificity, is the fraction of relevant instances among the retrieved instances; F1-score is the harmonic mean of the precision and recall; AUC stands for "Area under the ROC Curve". That is, ROC curves show the trade-off between sensitivity and specificity. Classifiers that give curves closer to the top-left corner indicate a better performance.

The calculation was as follows:

Prediction		
	AD	**HE**
AD	a	b
HE	c	d

Precision (p) = $\frac{a}{a+c}$, Recall (r) = $\frac{a}{a+b}$, F1-score = $\frac{2pr}{p+r}$ = $\frac{2a}{2a+b+c}$, Accuracy = $\frac{a+d}{a+b+c+d}$

As the gene families conserved across all strains are not associated with the AD status, they would be noise for model performance. Therefore, feature selection on the training dataset is essential to select an optimal subset of predictive biomarker genes that have decisive roles in the differentiation of *S. aureus* strains. For each RF model training, all gene features are ranked by their accuracy weights (feature importance) assigned by the classification model. Only the non-rare gene families (in over 10% of strains) that are common enough to provide meaningful associations were selected for classifier training to enhance the generalizability and robustness of the model while avoiding overfitting to rare features that might not be broadly representative or predictive. The receiver operating characteristic curve was calculated using R package pROC version 1.18.4 (Robin *et al.*, 2011). Then the classifier was further optimized using the top 50 biomarker gene features from the training set. The performance of the optimized classifier was also evaluated with the real-world test dataset.

2.10 Identification and annotation of prophages

The prophage genome sequences of the *S. aureus* strains from AD patients and healthy individuals were identified based on the combined annotations using PhiSpy v4.2.19 (Akhter *et al.*, 2012), VIBRANT v1.2.1 (Kieft *et al.*, 2020), and Phigaro v2.3.0 (Starikova *et al.*, 2020) with default parameters. The quality assessment of prophage sequences was performed using CheckV v0.8.1 (Nayfach *et al.*, 2020). The prophage species were identified using CD-HIT V4.8.1 with commands "-c 0.95 -n 10 -d 0 -M 0" (Fu *et al.*, 2012) by clustering high-quality prophage sequences (completeness > 90%) to each other, setting a 95% similarity as the species cutoff (Turner *et al.*, 2021). The number of common and unique prophage species of AD and HE strains were shown using InteractiVenn (Heberle *et al.*, 2015).

Taxonomic assignment of *S. aureus* high-quality prophages was performed using vConTACT v.2.0 (Bin Jang *et al.*, 2019). Prophage gene clusters were identified using the same workflow for orthologous gene clustering of *S. aureus* strains. The functional annotation of the

prophages identified was carried out using the PHROG database v3 (Terzian *et al.*, 2021). To reduce the prevalence of genes with unidentified functions, each unknown gene within a given cluster was assigned the most commonly observed known function within that same cluster. The auxiliary metabolic genes (AMGs) encoded by the prophage sequences were identified using DRAM-v1.4 (Shaffer *et al.*, 2020).

The virulence factors in prophage sequences were annotated with the online VFanalyzer in VFDB 2022 (http://www.mgc.ac.cn/VFs, May 26, 2023) (B. Liu *et al.*, 2022). Then the number of VFs per strain was calculated using a custom Python script (vf_counting_per_strain.py), which was then visualized using Pheatmap v1.0.12 with Euclidean as the row clustering method (Raivo Kolde, 2019).

2.11 Micro-evolutionary analysis of *S. aureus* strains

For each isolate of *S. aureus*, the sequence type was determined using MLST 2.19.0 (https://github.com/tseemann/mlst). The *S. aureus* strains, whose STs were found in both AD and HE groups, were defined as matched STs and retained for further analysis of micro-diversity to minimize the effect of asymmetrical sampling. The core genome alignment of *S. aureus* strains was obtained using Roary v3.13.0 (Page *et al.*, 2015), with the parameters “-e -n -cd 95 -r -v -i 70 -iv 2”. The input file for Roary was the annotation file in gff3 format generated using Prokka v1.14.6 (Seemann, 2014), with the parameters "--genus Staphylococcus --species aureus --cpus 30 --evalue 1e-05”. Pairwise single nucleotide polymorphisms (SNP) distance matrix for *S. aureus* strains was calculated using snp-dists v0.7.0 with default settings (https://github.com/tseemann/snp-dists).

We also assessed the strength of purifying selection on *S. aureus* strains using the ratio of nonsynonymous and synonymous substitutions (dN/dS), where values less than 1 indicate purifying selection and lower values suggest stronger purifying selection and higher values hint at greater genetic drift (weaker purifying selection). Based on the core genome alignment for strains from AD and HE, respectively, the pairwise dN/dS ratio was calculated using Maximum-likelihood approximation (codeML) within the PAML v4.10.6 package (Álvarez-Carretero *et al.*, 2023). We excluded dN/dS values where dS ≥1, as this suggests synonymous substitutions are approaching saturation.

2.12 Statistical analysis

Statistical analyses were conducted starting with assessing the normality of the data distribution using either the Shapiro-Wilk test or the Kolmogorov-Smirnov test when the observations exceed 5000. Subsequently, the Welch two sample t-test (normal distribution) or non-parametric Mann-Whitney U-test also called the Wilcoxon rank sum test (non-normal distribution), was employed to evaluate the statistical significance of observed differences. Any p-value < 0.05 was considered indicative of statistical significance.

2.13 Data availability

The datasets supporting the conclusions of this manuscript have been included in Supplementary Tables. The real-world dataset (48 Augsburg strains) was published by Wang *et al*., 2022. All Python and R scripts and the RF model workflow used in this thesis are available at Github: https://github.com/zhongjiew/Scripts4PhD.

3 Results

3.1 Data characterization

In our study, we analyzed a global dataset of 150 AD strains from 4 countries and 150 HE strains from 9 countries, complemented by a local German dataset of 33 AD and 15 HE strains. As shown in Table 1, the genome sizes varied, with the AD group having a narrower size range (2.66 to 2.97 Mb) compared to the HE group (2.69 to 3.12 Mb). The GC content was relatively consistent across both groups, with a minimum of 32.6% in AD and 32.2% in HE, and both reaching a maximum of 33%. Genome completeness was high in both groups, only slightly higher in the HE group (98.58% to 99.65%·) than in the AD group (98.01% to 99.51%). Contamination rates were comparable and low, with a maximum of 4.05% in AD and 4.15% in HE. The number of scaffolds and scaffold N50 values highlighted the AD group's genome assembly as more fragmented (1 to 416 scaffolds; N50 ranging from 12,677 bp to 2,855,872 bp) compared to the HE group (1 to 248 scaffolds; N50 ranging from 121,200 bp to 2,965,341 bp). Gene content analysis revealed a higher number of genes in the HE group (2465 to 3021) as opposed to the AD group (2432 to 2856), indicating a potential difference in gene repertoire related to health status. Additional details are in Tables S1&S2.

Table 1. Statistics of all human-derived *S. aureus* assemblies

	AD			**HE**		
	Min	Mean	Max	Min	Mean	Max
Number of strains	150[a] + 33[b]			150[a] + 15[b]		
Number of countries	4			9		
Genome size (Mb)	2.66	2.80	2.97	2.69	2.80	3.12
GC (%)	32.6	32.82	33	32.2	32.73	33
Completeness (%)	98.01	99.44	99.51	98.58	99.45	99.65
Contamination (%)	0.08	0.27	4.05	0.08	0.23	4.15
Number of scaffolds	1	37	416	1	45	248
Scaffolds N50 (bp)	12677	763555	2855872	121200	756086	2965341
Number of genes	2432	2588	2856	2465	2612	3021

Note: a, global dataset; b, local dataset.

3.2 Reduced gene content diversity of AD strains in the global dataset

To compare the genome difference between AD and HE strains in the global dataset, the gene content diversity of the 300 global *S. aureus* strains identified an “open” pan-genome (the number of new gene families continuously increasing as more genomes are included) and a stable core genome as indicated by the fitting functions (Figure 5A), consisting of 41% of core, 40% of accessory, and 19% of unique genes (Figure 5B).

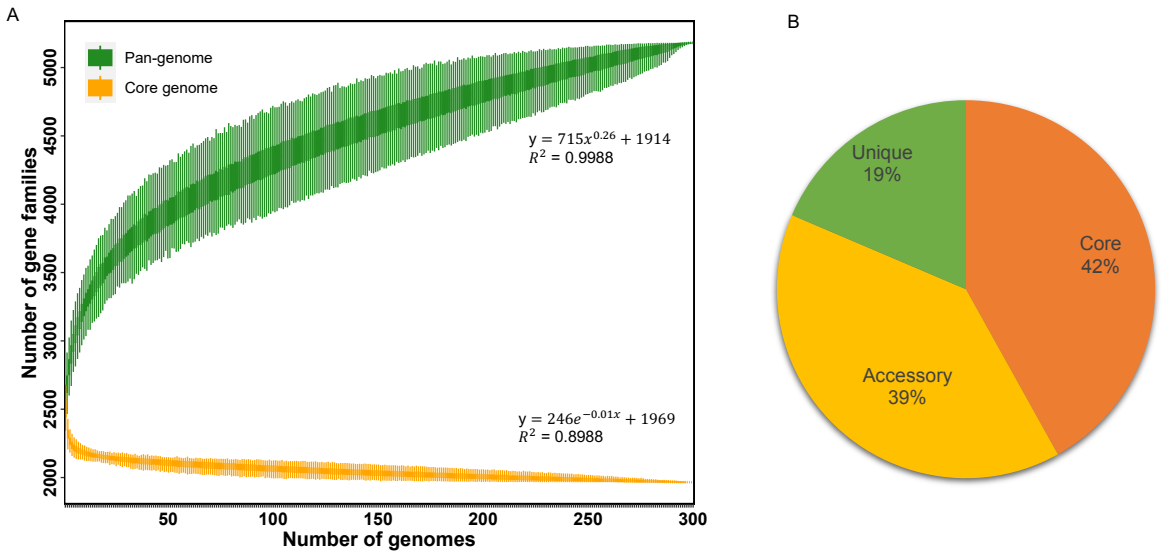

Figure 5. Pan-genome and core genome of 300 global *S. aureus* strains. (A) Pan-genome and core genome accumulation curves of the global 300 *S. aureus* strains from AD and HE as a function of the number of isolates, calculated by PanGP. (B) Pie chart of the percentage of gene families within the three gene categories (core: in >95% of strains, accessory: in > 1 strains but not core, and unique: only in 1 strain) for the 300 strains from AD and HE in the global dataset.

When analyzing the pan-genome and core genome of strains from AD and HE separately, the accumulation curves illustrated that the pan-genomes for both AD and HE strains increased rapidly and maintained an “open” status, respectively (Figure 6A). Intriguingly, the increasing rate for strains from HE outpaced strains from AD, achieving a greater pan-genome. Meanwhile, the core genome size for both groups showed a swift decline, yet eventually stabilized at a similar size.

To pinpoint genome composition differences between AD and HE, we categorized *S. aureus* genes into core (present in > 95% of strains), accessory (present in multiple strains but less prevalent than core), and unique (exclusive to one strain). As shown in Figure 6B, while the core genomes of both groups were notably similar, strains from HE had more accessory and unique genes than those from AD, accounting for the larger pan-genome size in HE strains. Examining the gene distribution at the strain level revealed a significantly elevated proportion of core genes in AD-associated strains (Mann-Whitney U-test, *p*-value < 0.01, Figure 6C), hinting at AD-specific selective pressure fostering strain homogeneity. In contrast, strains from HE were characterized by a significantly higher proportion of accessory genes (Mann-Whitney U-test, *p*-value < 0.01, Figure 6C). Both groups had a minimal proportion of unique genes.

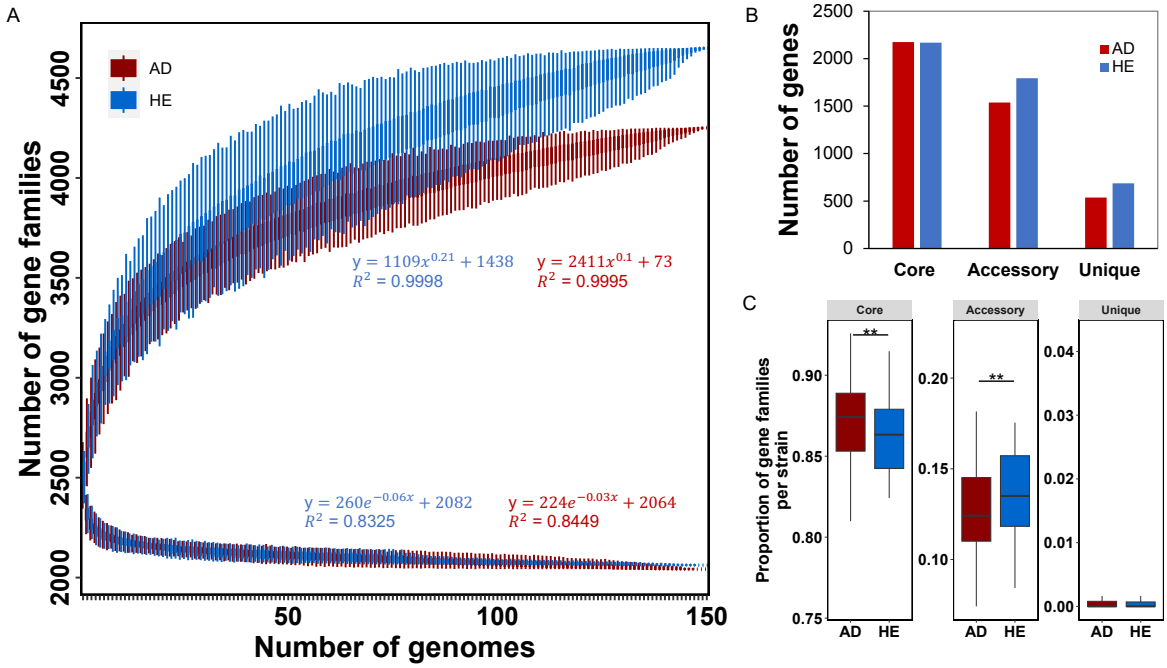

Figure 6. Gene content of the global 300 *S. aureus* strains. (A) Pan-genome and core genome accumulation curves of strains from AD and HE, separately, as a function of the number of isolates, calculated by PanGP. (B) Number of core, accessory, and unique genes for strains from AD and HE, respectively. (C) Proportion of gene families within the three gene categories at the strain level (core: in >95% of strains, accessory: in > 1 strains but not core, and unique: only in 1 strain). Statistical significance between AD and HE groups was calculated by the Mann-Whitney U-test. ** *p*-value < 0.01.

3.3 Greater functional variability of AD strains in the global dataset

We then profiled the functional repertoire of these strains to investigate whether the trend of reduced gene content in AD strains retained the functional variability. The top five prevalent functions encompassed Transporters, ABC transporters, DNA repair and recombination proteins, Enzymes with EC numbers, and Two-component system, with Transporters substantially dominating (Figure 7A). PCA analysis (Figure 7B) identified the top differentiating functions in the global strains as Replication and repair, Transporters, Quorum sensing, Two-component system, and AMR genes. However, contrary to the reduced gene content, AD strains exhibited a significantly broader functional variability (Mann-Whitney U-test, *p*-value < 0.001, Figure 7C).

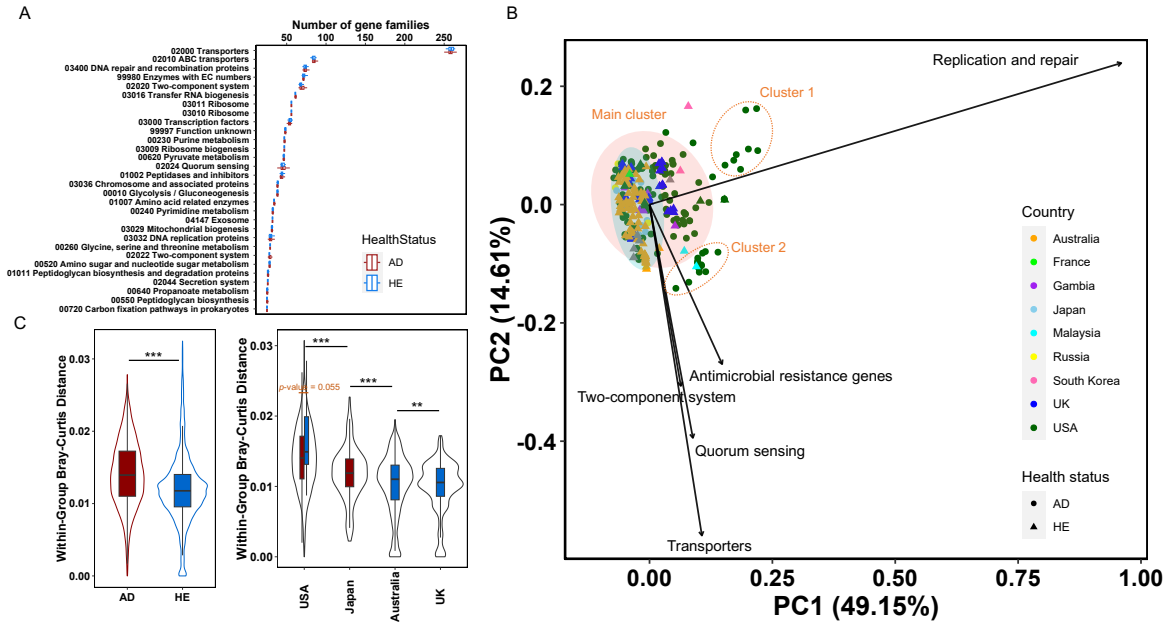

Figure 7. Functional divergence of the 300 global *S. aureus* strains from AD and HE. (A) Strain-level abundance of gene families associated with each KEGG functional category. Only the top 30 abundant functions are shown, ranked by median in descending order. (B) PCA analysis of the differentiating functions within the global *S. aureus* strains. Only the top 5 differentiating functions are shown. Colors represent the country of origin and shapes indicate health status. Shaded ellipses represent the 95% confidence interval for the AD (red) and HE (blue) groups. Two additional clusters, Cluster 1 and Cluster 2, are depicted with unfilled dotted ellipses. Variances explained in the two dimensions are expressed in the parenthesis. (C) Within-group Bray-Curtis distance in terms of health status and countries of origin. Only the four countries with more than 10 strains are shown. Statistical significance was calculated by the Mann-Whitney U-test for health status and the Kolmogorov-Smirnov test for countries. * *p*-value < 0.05; ** *p*-value < 0.01; *** *p*-value < 0.001.

The global 300 strains were derived from nine countries (Table 1). Interestingly, as depicted in Figure 7B, the variation in the AD group primarily arose from the American strains. Except for the main cluster housing the majority of the strains, two additional distinct clusters were identified that mainly contained AD strains: Cluster 1 predominantly featured Replication and repair, with 9 AD strains; Cluster 2, comprising 13 AD strains and 1 HE strain, was characterized by AMR, Transporters, Quorum sensing, and Two-component system. Both clusters were dominant with American AD strains, with only one Malaysian HE strain, highlighting the high strain diversity in the US. These two clusters were only slightly affected when strains from AD and HE were clustered separately (Figure 8).

Significant differences were revealed among strains from different countries including the USA, Japan, Australia, and the UK (Mann-Whitney U-test, all *p*-values < 0.01, Figure 7C), with the American strains exhibiting the highest diversity. An obvious distinction between AD and HE strains within the US was also observed, despite it was not significant (Mann-Whitney U-test, *p*-value = 0.055, Figure 7C), possibly due to the imbalance in strain numbers (124 AD vs. 6 HE strains).

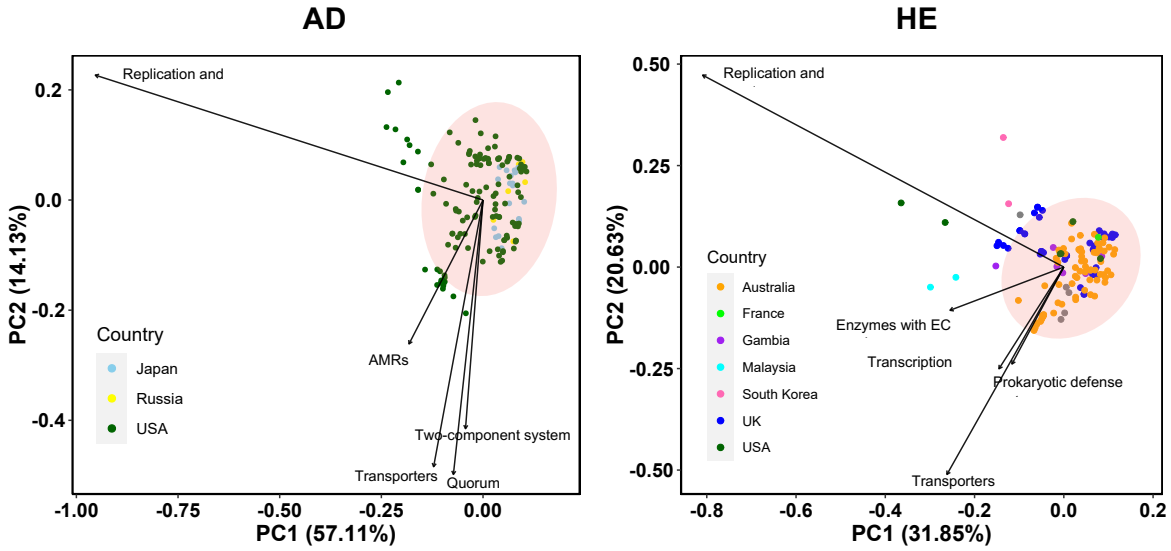

Figure 8. PCA analysis of differentiating functions within the global *S. aureus* strains for AD and HE groups, separately. Only the top 5 differentiating functions are shown. Colors represent the country of origin. Shaded ellipses represent the 95% confidence interval for the AD and HE groups. Variances explained in the two dimensions are expressed in the parenthesis.

To further assess the association of human-derived *S. aureus* strains and those originating from animals and the potential of zoonosis through host-switching transmission, we performed a PCA analysis on *S. aureus* strains from human and six other mammals (cow, dog, horse, pig, sheep, and yak) based on the KEGG functional annotation (metadata of animal strains in Table S3). The result showed a significantly pronounced distinction between human isolates and those from all other animals (Mann-Whitney U-test, *p*-value < 2.2e-16, Figure 9), implying a distinct origin and evolutionary trajectory of the human-associated strains and limited potential for zoonosis.

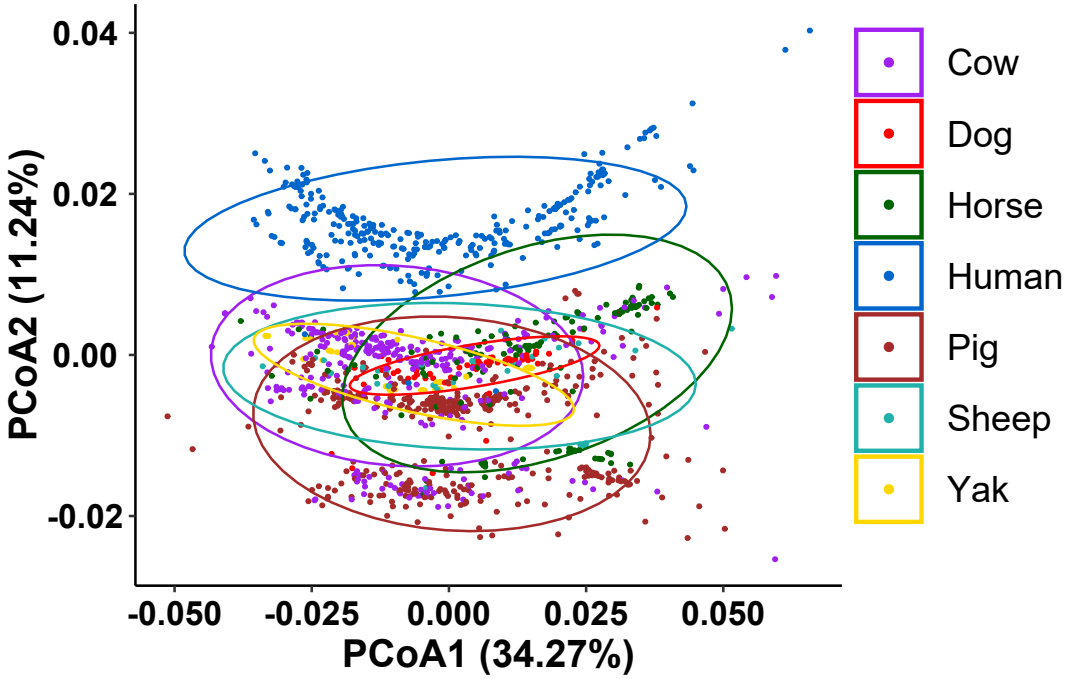

Figure 9. PCA analysis of *S. aureus* strains isolated from humans and other mammalian hosts, including Cow, Dog, Horse, Pig, Sheep, and Yak, based on the gene abundance table of KEGG functional categories. Colors represent hosts of origin. Ellipses represent the 95% confidence interval for different hosts. Variances explained in the two dimensions are expressed in the parenthesis.

3.4 Diversification of AD-dominant Cluster 1 and Cluster 2

To elucidate which functions drive the differentiation of Cluster 1 and Cluster 2, we examined the metadata and functional orthologs in the five featured categories. As depicted in Figure 10A, the distribution of strains across three different clusters, namely Main cluster, Cluster 1, and Cluster 2, were stratified by country of origin and health status of the individuals from which the strains were isolated. A diverse geographical representation is observed within the clusters, with strain contributions from Australia, France, Gambia, Japan, Malaysia, Russia, South Korea, the UK, and the USA, along with some strains of uncertain origin ('Not determined'). Notably, the Main cluster shows a higher diversity of countries compared to Clusters 1 and 2, which are predominantly composed of strains from the USA. The health status distribution within the clusters indicates a comparable dispersal of AD and HE strains in the Main cluster, while Cluster 1 and Cluster 2 exhibit a greater predominance of AD-associated strains. This distribution suggests potential geographical and health status-related patterns in the genetic clustering of the strains, which could reflect underlying ecological or evolutionary processes.

To further investigate what functional orthologs drive the diversification of Cluster 1 and 2, we compared the number of gene copies of each function within the five featured functional categories. Compared to the main cluster, Cluster 1 strains demonstrated a 6-9 fold higher abundance in three transposases (K07497, K07482, and K07483) from Replication and repair (Figure 10B). Cluster 2 strains were defined by twenty functional orthologs, predominantly found in this cluster or displaying more gene copies than the main cluster. These included 5 lantibiotic genes (*nisE, nisF, nisC, nisB, and nisP*), 2 *braSR* two-component system genes (*braS* and *braR*), 2 bacitracin transport genes (*bcrB* and *bcrA*), 1 argine antiporter (*arcD*), 4 peptide/nickel transport genes (ABC.PE.P, ABC.PE.P1, *ddpD*, and *ddpF*), 1 bacterial/archaeal transporter family 2 protein (TC.BAT2), and 5 AMR genes (*msrA*, *mph*, *aph3*-III, *sat4*, and *mecR1*). Strikingly, 60% of these genes are related to antibiotic resistance, with these genes denoting resistance to lantibiotics, bacitracin, macrolide, aminoglycoside, streptothricin, and methicillin antibiotics, respectively. The *braSR* two-component system and bacterial/archaeal transporter family 2 protein act as tools for transporting and regulating a diverse range of molecules and processes like stress response and antibiotic resistance. The arginine antiporter and peptide/nickel transport genes are involved in the transport of arginine, an amino acid essential for bacterial growth and survival, as well as peptides and nickel ions, playing a role in nutrient uptake and metal ion homeostasis.

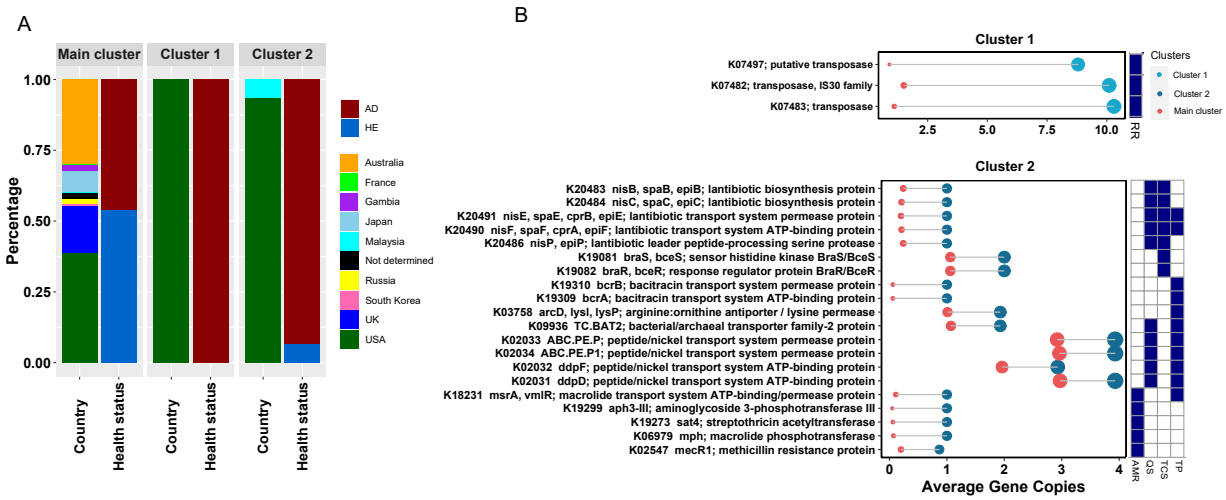

Figure 10. Metadata and functional orthologs diversifying *S. aureus* strains in Cluster 1 and Cluster 2. (A) Distribution of countries of origin and health status for the strains of Main cluster, Cluster 1, and Cluster 2. In each cluster, the proportions of strains derived from individuals with AD and HE are indicated by red and blue color blocks, respectively. (B) Functional genes diversifying the *S. aureus* strains of Cluster 1 and Cluster 2. Dot size indicates the number of average gene copies of strains in the respective cluster, as represented by the X-axis. Colors represent Main cluster (red), Cluster 1 (sky blue), and Cluster 2 (navy blue). Heatmap on the right shows pathways these genes are involved in, with dark blue representing presence. RR, Replication and repair; AMR, Antimicrobial resistance genes; QS, Quorum sensing; TCS, Two-component system; TP, Transporters.

3.5 HGT of the lantibiotic operon in Cluster 2 strains

To further unravel the mechanism propelling the diversification, we centered our analysis on the suite of lantibiotic-related genes of Cluster 2. Our result already showed that these genes were universally present in Cluster 2 strains with an identical number of gene copies but nearly absent in the main cluster (Figure 10B). The phylogenetic tree based on the *nisB* gene revealed robust evidence of HGT events of *nisB* across multiple families within the Bacillales order (Figure 11A). Specifically, the tree topology showed a closer relationship between Bacillaceae and Staphylococcaceae. Within the Bacillaceae clade, multiple genera harbored this gene. However, within the Staphylococcaceae clade, the *nisB* gene was exclusively detected in the single *Staphylococcus* genus. Given the little probability of a gene spreading from one single genus to multiple disparate families through HGT, it is plausible that certain *S. aureus* strains acquired the lantibiotic operon from other *Staphylococcus* species, which had likely obtained it from the Bacillaceae family. Interestingly, the closest Bacillaceae hits included

Alkalihalobacillus clausii, *Priestia megaterium*, and *Bacillus cereus*, primarily environmental bacteria and widely distributed in nature, implying potential interactions between human-associated and environmental bacteria.

Subsequently, we compared the lantibiotic operon synteny in the genomic sequences of representative *S. aureus* strains from both Cluster 2 and Main cluster (Figure 11B). The operon, arranged in the *nisA, B, C, D, P, F, E,* and *G* order, was inserted within the genome sequences of Cluster 2 strains, but absent in the corresponding region of strains from the main cluster. This further supports that HGT events led to the acquisition of the lantibiotic operon and propelled the diversification of Cluster 2 strains.

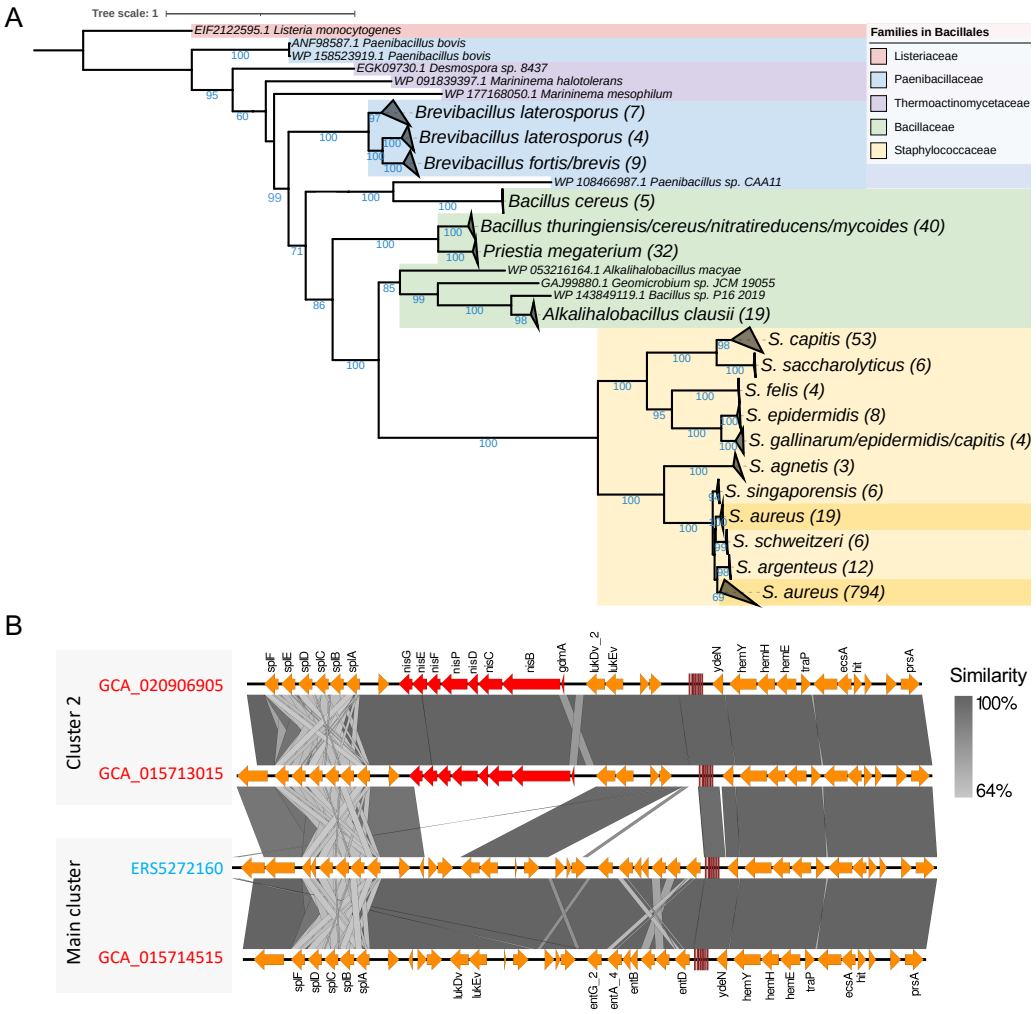

Figure 11. HGT analysis of the lantibiotic operon of human-derived *S. aureus* strains. (A) Phylogenetic tree based on the *nisB* gene of the lantibiotic operon. Colors of the clades show the respective affiliated families under the order Bacillales. *S. aureus* is especially marked in darker yellow. Numbers in the

parentheses beside the labels indicat the number of gene sequences included. Numbers beside nodes indicate bootstrap values. (B) Synteny of the lantibiotic operon in representative *S. aureus* strains in Main cluster and Cluster 2. The lantibiotic operon is highlighted in red. Gene functions obtained from Prokka are labeled above and below. On the right, AD strains are denoted in red and HE strains in blue. The gradient grey bar at the right denotes the pairwise similarity of the alignment.

3.6 Gene content diversity and functional variation of the Augsburg collection

To verify whether the gene content diversity and functional variation patterns found in the global dataset hold on a local scale, we utilized 48 *S. aureus* strains (33 AD vs 15 HE) isolated from Augsburg, Germany. Both the pan-genome and core genome of the local strains displayed stability (Figure 12) with the increase of genomes analyzed. The pan-genome comprised a higher proportion of core genes (64%) than the global strains (41%), comparable accessory genes (local vs global: 35% vs 39%), and much fewer unique genes (local vs global: 1% vs 19%).

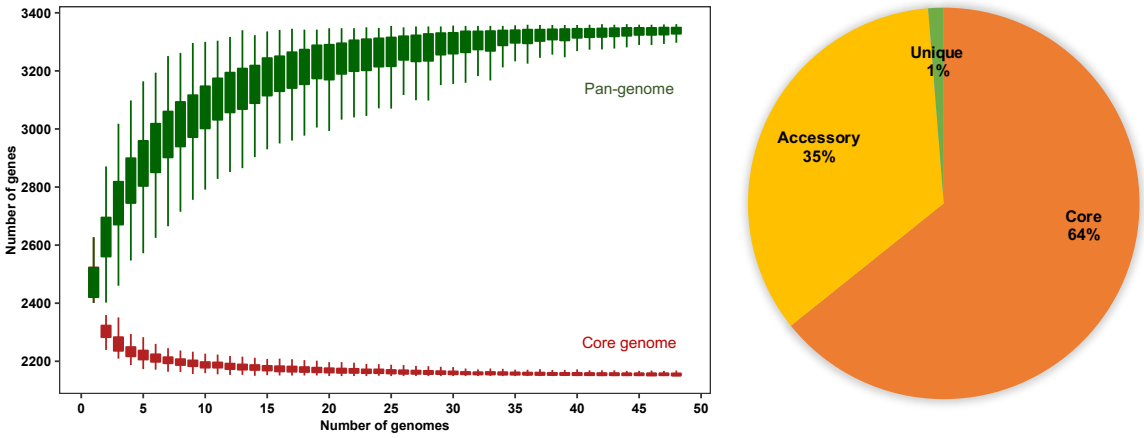

Figure 12. Pan-genome and core genome of the 48 local *S. aureus* strains. (A) Pan-genome and core genome accumulation curves of the local 48 *S. aureus* strains from AD and HE isolated from Augsburg. The fitting functions of the curves were not shown as both curves remain stable with the increase of strains. (B) Pie chart of the percentage of gene families within the three gene categories (core: in >95% of strains, accessory: in > 1 strains but not core, and unique: only in 1 strain) for strains from AD and HE in the local dataset.

We utilized accessory genes of the local strains for phylogenetic analysis as their differential presence in genomes can enhance resolution in distinguishing closely related *S. aureus* strains. The phylogenetic tree (Figure 13A) exhibited a distinct pattern between the strains from AD and HE groups with only one exception for each group. A distinction between different subjects was also observed. However, there was no difference between isolates from the nose and skin for both analyses (Figure 13A). The pan-genome and core genome of the local AD and HE strains (Figure 13B) mirrored the global pattern that HE strains showed a larger pan-genome than AD strains and both groups exhibited a similar core genome. At the strain level, AD strains had more core genes, while HE had more accessory genes. Both groups exhibited minimal unique genes. No significant difference was detected between the two groups across all three gene categories (Figure 13C). Functional distribution for each local strain showed a clear functional divergence between strains from AD and HE (Figure 13D), with the most abundant functional categories annotated as Transporters, ABC transporters, DNA repair and recombination proteins, Enzymes with EC numbers, and Two-component system, congruent with the global strains. Contrary to the global dataset, PCA analysis delineated that AD strains were more functionally homogenous than HE strains at the local scale (Wilcoxon rank sum test, *p*-value < 0.001, Figure 13E&F). The most distinct functional categories, ranked by the contribution, were Replication and repair, Transporters, Transcription factors, Enzymes with EC numbers, and Lysine biosynthesis.

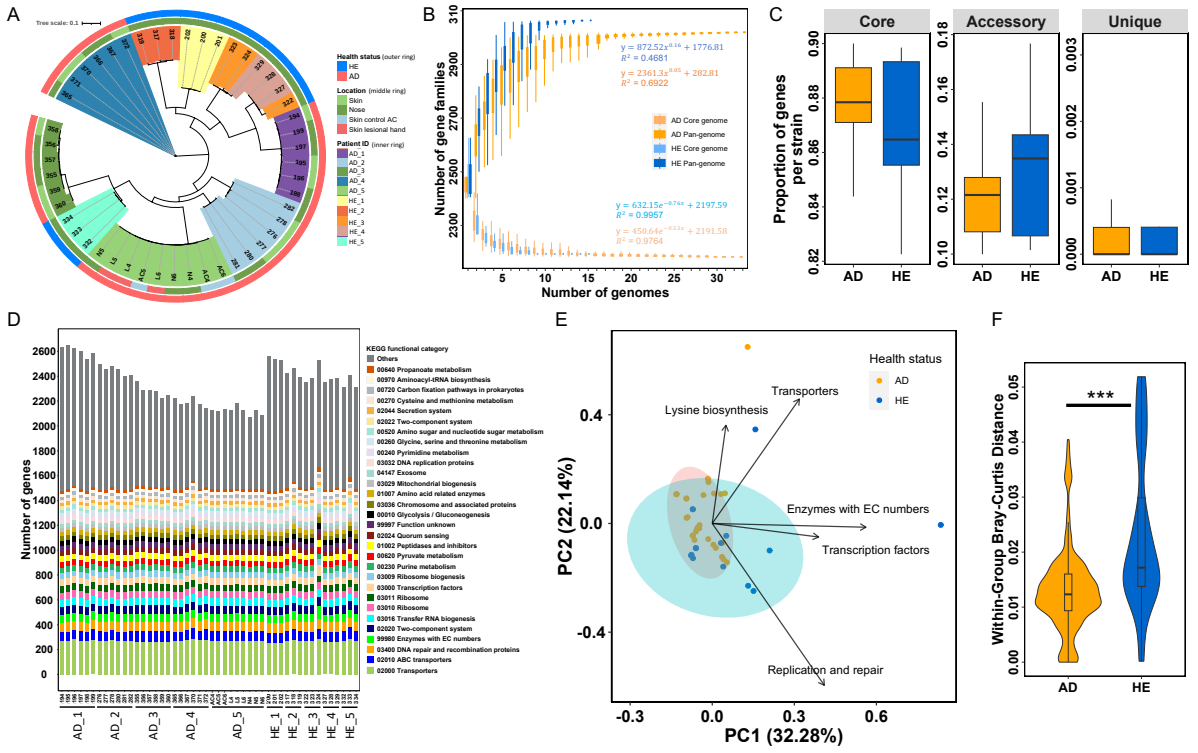

Figure 13. Gene content diversity and functional variation of the 48 local *S. aureus* strains from Germany. (A) Phylogenetic tree based on the accessory genes identified by Roary. Bootstrap values are shown beside nodes. Colors in the outer ring, middle ring, and clades represent health status, skin location, and patient ID, respectively. Strains are shown per individual. (B) Core genome and pan-genome accumulation curves of the AD and HE strains, respectively, as a function of the number of isolates. Fitting functions of the curves are shown beside the curves. (C) Proportion of gene families at the strain level within the three gene categories (core: in >95% of strains, accessory: in > 1 strains but not core, and unique: only in 1 strain) for both AD and HE strains, respectively. (D) Gene abundance of KEGG functional categories for each strain. Only the top 30 abundant functional categories are shown, and the remaining functional categories are shown as Others in grey. The functional categories are ranked by abundance from the bottom to the top. (E) PCA analysis based on the abundance of KEGG functional categories. The top 5 differential functional categories are shown. Ellipses are plotted for the AD (pink) and HE (sky blue) cluster, representing the 95% confidence interval within each group. Variances explained in the two dimensions are expressed in the parenthesis. (F) The within-group Bray-Curtis distance of health status. Statistical comparisons between AD and HE groups were performed by the Mann-Whitney U-test. *** p-value < 0.001.

Examination of the top 10 differentiating orthologs (Table 2) exhibited largely deviation from global ones. Therein, Replication and repair featured 4 transposases (K07497, K07483, K07485, and K07498); Transporters featured 2 genes, including K03893 *arsB* arsenical pump membrane protein and K18104 *abcA*/*bmrA* ATP-binding cassette subfamily B; Transcription factors featured 2 transcriptional regulators, including K21903 *cadC*/*smtB* lead/cadmium/zinc/bismuth-responsive transcriptional repressor and K03892 *arsR* arsenate/arsenite/antimonite-responsive transcriptional repressor; Enzymes with EC numbers featured 2 ion transporter genes, including K01533 *copB* P-type Cu2+ transporter and K01534 *zntA* Zn2+/Cd2+-exporting ATPase. Although transposases play the most important role in the differentiation in both global and local datasets, the other functions related to Transporters, Transcription factors, and Enzymes with EC numbers were uniquely revealed in the local dataset. Notably, half of the top ten differential orthologs were involved in the transport and regulation of metal ions. Overall, although local strains somewhat mirrored global strains in gene content, they exhibited unique functional differentiation processes.

Table 2. Top 10 differentiating orthologs of the 48 Augsburg strains

Functional categories	Top 10 differentiating functional genes
Replication and repair	Four transposases (**K07497**, **K07483**, **K07485**, and **K07498**)
Transporters	**K03893** arsB arsenical pump membrane protein
	K18104 abcA/bmrA ATP-binding cassette subfamily B
Transcription factors	**K21903** cadC/smtB lead/cadmium/zinc/bismuth-responsive transcriptional repressors
	K03892 arsR arsenate/arsenite/antimonite-responsive transcriptional repressors
Enzymes with EC numbers	**K01533** copB P-type Cu2+ transporter
	K01534 zntA Zn2+/Cd2+-exporting ATPase

3.7 Phylogenetic tree of closely related *Staphylococcus* species based on 16S rRNA

In the phylogenetic analysis of the 16S rRNA gene sequences of *S. aureus* strains isolated from Augsburg, the result presents a clear delineation of the *S. aureus* clade from other *Staphylococcus* species, as illustrated in the phylogenetic tree (Figure 14). Notably, the tree demonstrates a clustering of various *Staphylococcus* species that highlights the genetic proximity within the genus. However, the genetic homogeneity, particularly among the coagulase-negative Staphylococci, points to the limitation of 16S rRNA sequencing in serving as an effective barcode for *Staphylococcus* species identification, exemplified by the insufficient separation between *S. aureus* and other *Staphylococcus* species, such as *S. argenteus* and *S. schweitzeri*. Besides, whole genome-based phylogenomic analysis was also not be able to differentiate *S. aureus* strains according to health status (AD lesional skin, AD non-lesional skin, and healthy skin) (Obata *et al.*, 2023). Therefore, the failure of both 16S rRNA gene-based and whole genome-based phylogenies to differentiate AD and HE strains highlights the need for additional discriminative genetic markers and a more robust, multi-locus approach to enhance species identification and accurately barcode the strains of this clinically significant species.

Figure 14. Phylogenetic tree based on the 16S rRNA genes of the 48 Augsburg *S. aureus* strains and other closely related *Staphylococcus* species. Bootstrap values are indicated at the nodes. The tree scale denotes genetic distance. Colored bars in the left, middle, and right side correspond to distinct patient IDs, skin locations, and health status, respectively. Strains were shown per individual.

3.8 RF classifier performance in identifying marker genes in *S. aureus* strains

To pinpoint potential marker genes that lead to phenotypic variations, we developed an RF classifier based on the gene presence-absence table of *S. aureus* strains. For this study, genomes of 348 *S. aureus* strains were used, including 183 from AD patients and 165 from HE. Therein, 150 strains from AD and HE, respectively, were used for the RF classifier construction and 48 Augsburg strains (33 AD vs 15 HE) were used for the further verification of the RF performance. We first evaluated the performance of the RF classifier using different training and test dataset ratios. As depicted in Figure 15A, the classifier demonstrated comparable F1-scores for both AD and HE groups. Specifically, AD had a better recall, while HE exhibited better precision across all partitions. Performance metrics constantly improved, especially accuracy rose from 86.93% in the 5:5 set to 91.67% in the 9:1 set, and AUC increased from 94.62% to 97.69%. Notably, AUC surpassed 95% in all partition sets except the 5:5, underscoring the classifier's excellent performance on the *S. aureus* classification. We then opted for the 9:1 partition ratio since its F1-score, accuracy, and AUC surpassed those of other partitions. It achieved an overall accuracy of 90% and an AUC of 94.67% for the test dataset (Figure 15B).

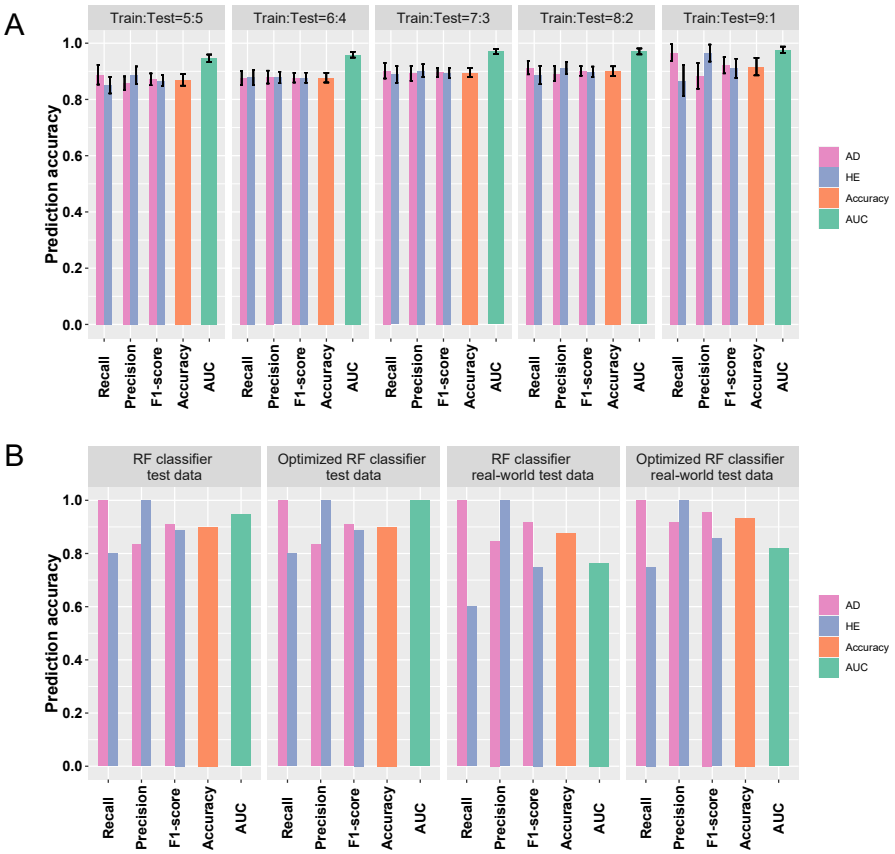

Figure 15. Determination of the train-test partition and performance of the RF classifier. (A) The dataset was divided into training and test datasets in 5:5 to 9:1 proportions across 10 iterations for each. Recall, precision, F1-score, accuracy, and AUC were used to evaluate the performance of the RF models. The black bars indicate the 95% confidence intervals. (B) Performance of the RF classifier and the optimized RF classifier on both the test and the real-world test datasets. The test dataset was divided into training and test datasets in 9:1 proportion. Recall, precision, F1-score, accuracy, and AUC were used to evaluate the performance of the RF models.

Based on the k-fold cross-validation results (Figure 16), selecting the top 50 biomarkers yielded a balance between minimizing the prediction error rate and using a concise gene set. To further optimize the classifier, we conducted a feature importance analysis, pinpointing the 50 most informative gene features for classification. The optimized classifier consistently outperformed its initial version. Specifically, the optimized RF classifier achieved an accuracy of 90% and an AUC of 100% for the test dataset (Figure 15&16). We also assessed the performance of both classifier versions on the real-world test dataset. The initial classifier attained an overall accuracy of 87.5% and an AUC of 76.36%, while the optimized classifier reached an accuracy of 93.33% and an AUC of 81.82% (Figure 15&16). Overall, the RF classifier demonstrated high reliability in distinguishing AD-related strains from healthy isolates using the top 50 biomarker genes.

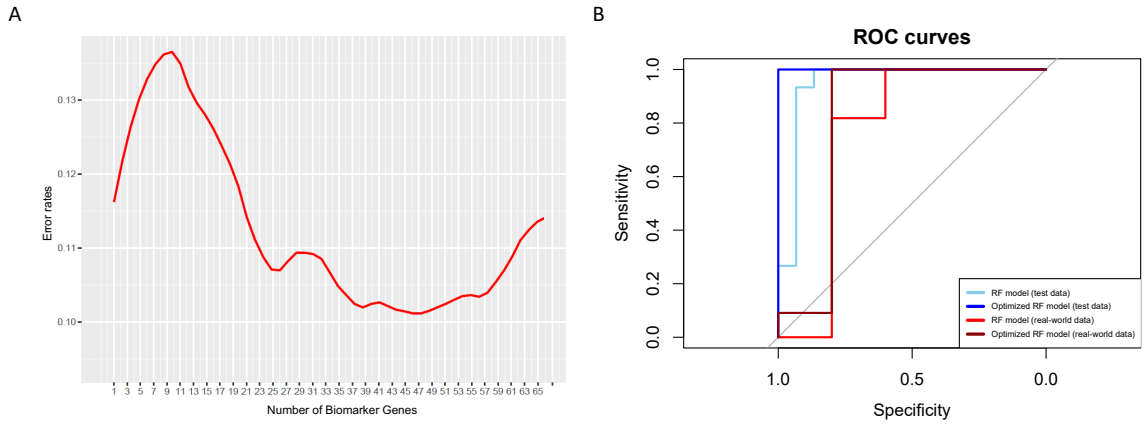

Figure 16. Selection of key marker genes via 10-fold cross-validation. The optimal prediction error rate was achieved with 50 genes. Consequently, a refined classifier was developed using the top 50 marker genes, ranked by their mean decrease in accuracy. Receiver operating characteristic curves (ROC) for prediction of *S. aureus* strains from AD and HE groups based on the presence-absence table of gene

families generated by MCL. Sensitivity is also recall. Specificity is also precision. AUC (Area under the curve) was 0.9467 and 1 for the test data when using the RF model and the optimized version, respectively, while the real-world test data achieved 0.7636 and 0.8182, respectively.

3.9 Distribution and function of the marker genes in *S. aureus*

Based on our analysis of the distribution of the 50 marker genes, *S. aureus* strains formed two clusters using the genomes obtained from the 300 isolates (Figure 17). One cluster predominantly consisted of strains from HE (77.7%, HE cluster), while the other cluster was predominantly based on strains from AD (73.9%, AD cluster). In addition, four distinctive co-occurrence patterns emerged among the 50 feature genes (Figure 17); two groups were predominantly found in the HE cluster, whereas one group was more prevalent in the AD cluster, and the last group showed varied prevalence among its within-groups genes.

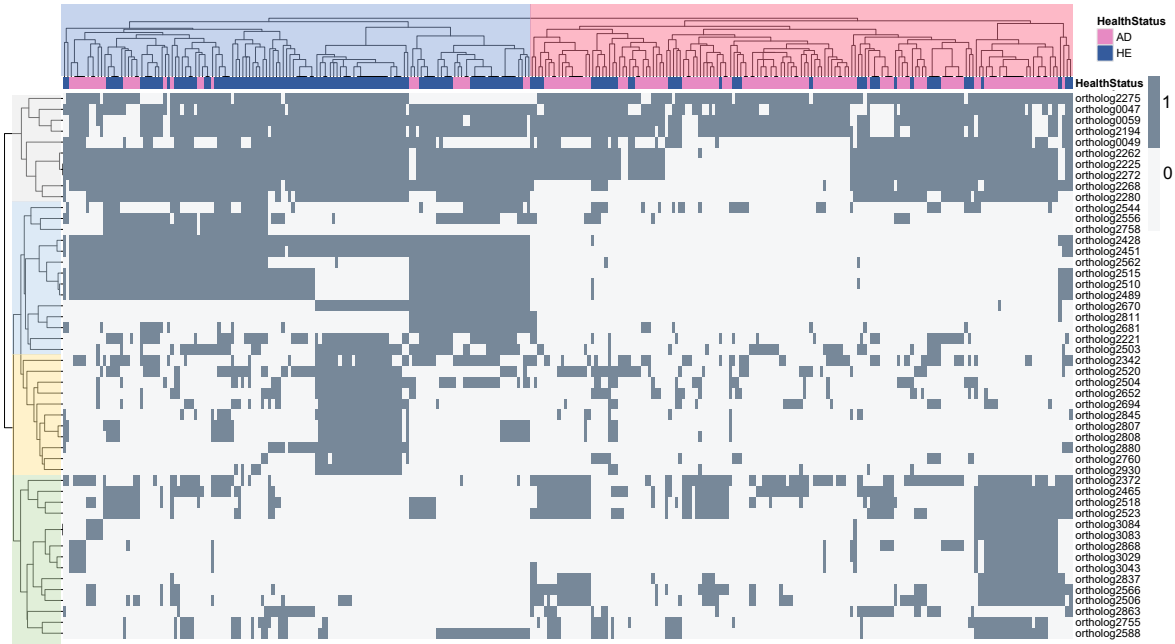

Figure 17. Heatmap of the 50 marker genes in the public dataset. Dark grey indicates presence and light grey indicates absence of marker genes. The strains (columns) form two main clusters shadowed in blue (HE cluster) and red (AD cluster), respectively. Colors for health status are the same in B. The marker genes (rows) form four clusters. Both rows and columns were clustered using the Euclidean method.

For the real-world dataset, all strains from AD patients were accurately classified (Figure 18), with most prediction probabilities for each strain of the AD group surpassing 90%. The strains from HE had varying prediction probabilities between 50% and 90%, despite three instances of misclassification. However, the absence of 10 feature genes, particularly the top two, might explain the relatively lower prediction probabilities for some strains in this dataset.

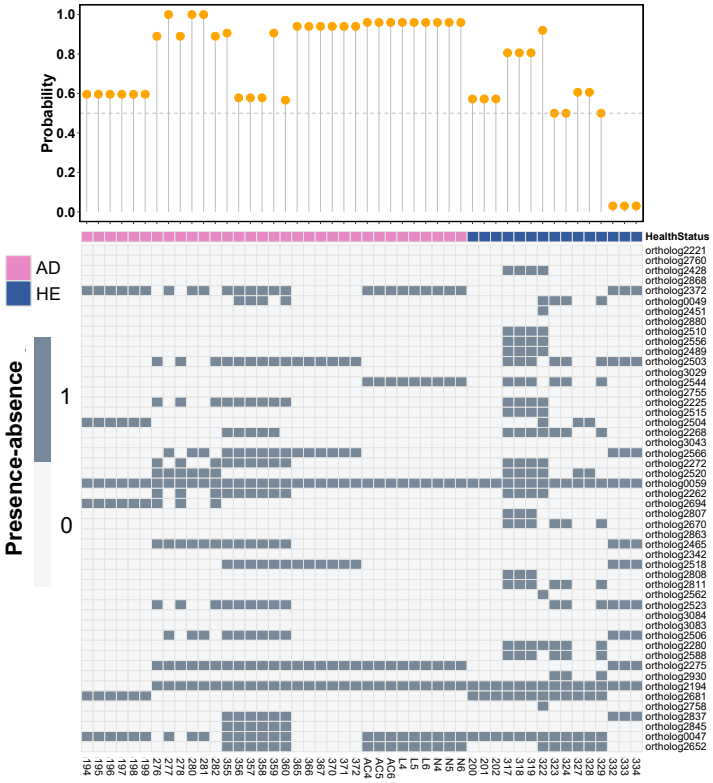

Figure 18. Heatmap of the presence-absence of the 50 marker genes of the real-world dataset and the prediction probability of the optimized classifier for each strain. The 50 marker genes (rows) are ranked by importance.

We subsequently examined the importance and functions of the marker genes. As shown in Figure 19, a significant proportion of the genes, 37 out of the 50 (74%), were more commonly found in strains from HE, suggesting that the differentiation was predominantly influenced by gene families exclusive or more abundant in the HE group. Upon functional characterization of the genes, 32 genes (64%) were associated with Staphylococcal phages. Specifically, phage holin was given the highest classification weight and had a higher presence in strains from HE. Toxin genes, representing 10% of the feature genes, also emerged as keys in the

differentiation, with many originating from prophages. Interestingly, 20% of the genes mainly found in strains derived from AD patients were hypothetical, suggesting a need to further explore *S. aureus* functions in AD, even though *S. aureus* is a widely studied model organism.

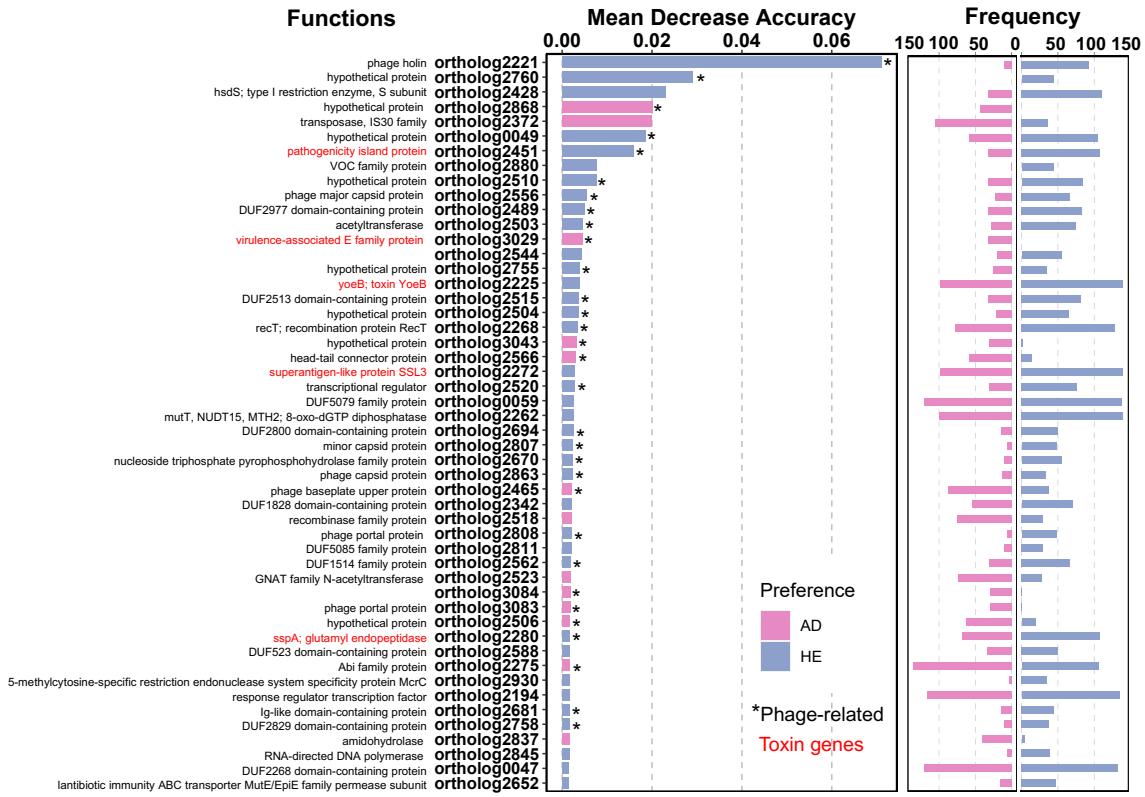

Figure 19. Function and frequencies of the most discriminative 50 marker genes as assigned by the RF classifier. Orthologs were assigned using MCL, and functions annotated using KEGG and Refseq. Toxin genes are labeled in red, assigned using the virulence factor database (VFDB). Middle panel shows marker genes ranked by mean decrease in accuracy, as determined by an RF classifier. Bar color indicates the prevalence of genes in all AD (pink) or HE (light blue) strains. Genes marked with asterisks (*) are phage-related, as assigned by blasting against predicted prophage genes. The right panel shows gene frequency in all AD and HE strains, respectively.

Since *S. aureus* possesses diverse STs, we also explored whether feature genes correlate more with health status (AD vs. HE) or clonal structure (STs) due to asymmetrical ST sampling between AD and HE. ST analysis revealed that 78% of our strains exhibit matching STs (Table S1), indicating that those with asymmetrical STs represent a relatively small subset. Furthermore, the identification of biomarkers with matched ST samples from AD and HE revealed that 31 biomarkers, accounting for 62%, remained consistent with the original set. Notably, the top 22 biomarkers from this repeated analysis overlapped with those identified in

the original study, contributing to an 84% prediction accuracy for these biomarkers. This classifier (Figure 20A) also demonstrated highly similar performance with the original version (Figure 15B). Moreover, the clustering analysis (Figure 20B) based on the presence of feature genes revealed a balanced distribution of major STs (e.g., ST1, 5, 8, 15) across strain clusters, and the phylogenetic analysis (Figure 20C) illustrated only a small proportion of strains exhibiting near-clonal similarity based on phylogenetic distance, further supporting our findings. Taken together, these findings suggest that the identified biomarkers are more significantly correlated with the health status of individuals rather than the asymmetrical distribution of STs, although the latter does have influence on the results to some extent.

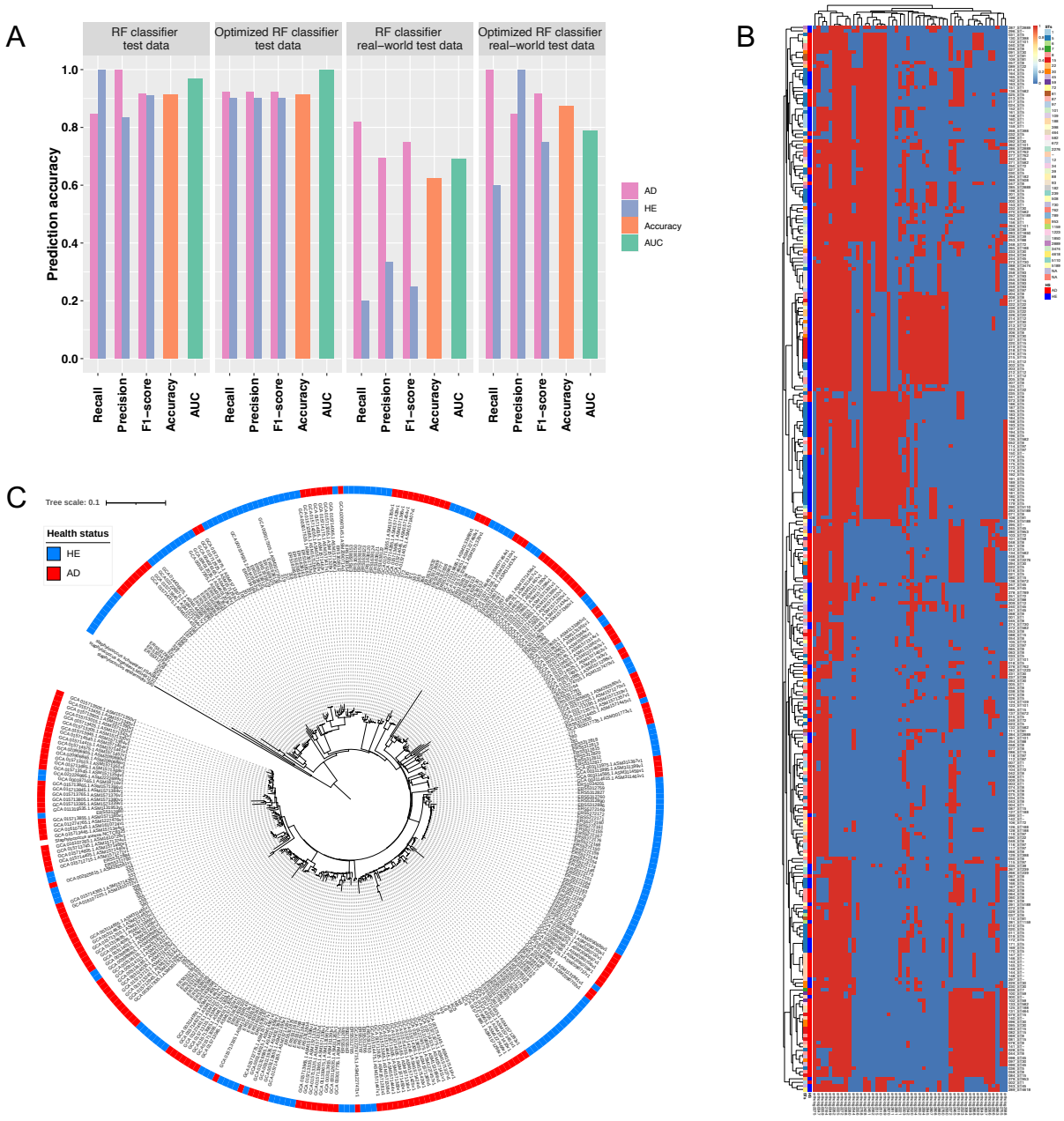

Figure 20. RF classifier performance using strains with matched STs and influence of STs. (A) Performance of the RF classifier and the optimized RF classifier based on the strains with matched STs. The test dataset was divided into training and test datasets in 9:1 proportion. Recall, precision, F1-score, accuracy, and AUC were used to evaluate the performance of the RF models. (B) Heatmap of the clustering analysis of 300 publicly available strains, each labeled with its respective ST, based on the presence or absence of the 50 marker genes. STs with a high number of strains are approximately evenly distributed across clusters of strains (refer to clustering of rows), exemplified by ST1, 5, 8, 15, et al. Red indicates presence and blue indicates absence of marker genes. STs and health status of all strains (columns) were annotated beside row clusters. Labels in the right consists of two parts: genome number and specific ST assigned using MLST. Both rows and columns were clustered using the Euclidean method. (C) Phylogenetic tree based on the whole genomes of the 348 *S. aureus* strains in this study to show the clonal structure. Health status is shown for AD (red) and HE (blue) strains. Three genomes from *S. argenteus*, *S. epidermitis*, and *S. schweitzeri* were used as the outgroup.

To further investigate whether these genes can be used to discriminate *S. aureus* strains based on their hosts of origin, we did Principal Coordinates Analysis (PCoA) including *S. aureus* strains isolated from pigs, poultry, and yaks (metadata of non-human isolates in Table S3). The PCoA result showed human-derived strains (both AD and HE groups) clustered separately (Figure 21A). This suggests the ability of the "feature" genes to distinguish strains based on host origin. Additionally, significant within-group distance differences among hosts (Wilcoxon rank sum test, all *p*-values < 0.001, Figure 21B) emphasize the role of these genes in host-specific adaptations.

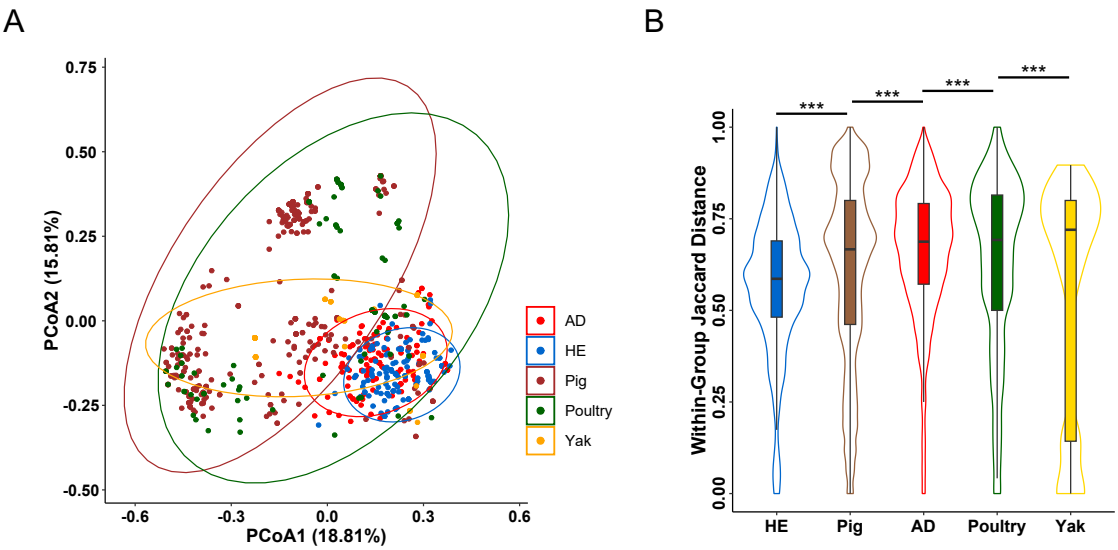

Figure 21. Feature genes can discriminate *S. aureus* strains based on their hosts of origin. (A) PCoA analysis based on the marker genes of *S. aureus* strains from different hosts. Strains from pig (dark red), poultry (dark green), and yak (yellow), along with the strains from AD patients (red) and healthy controls (blue), are included for comparison. The ellipses denote the 95% confidence interval. Variances explained in the directions are shown in the parenthesis. (B) Within-group distance of the five different hosts. The Jaccard method was used for the calculation. The hosts are ranked by the medians in ascending order. The significance comparison of different hosts was performed using the Wilcoxon rank sum test. The significance of differences is shown between all pairs of groups, indicated by p-values < 0.001 (***).

3.10 Micro-diversity in AD and HE *S. aureus* strains through SNP and dN/dS analysis

To further explore the micro-diversity in strains from AD and HE, we assessed the number of SNPs and dN/dS ratio derived from the core genome of *S. aureus* strains with matched STs present in both AD and HE groups, which could minimize the potential bias induced by asymmetrical clonal structure. A total of 45 STs were identified, of which 14 STs (244 strains) were present in both AD and HE groups (AD: 137 vs HE: 107 strains, Table S1). SNPs shed light on genetic variability while dN/dS reveals evolutionary pressures. Typically, a dN/dS < 1 suggests purifying or negative selection, with values closer to zero indicating a more intense purifying selection (Kryazhimskiy and Plotkin, 2008). The AD group exhibited significantly fewer SNPs and a lower dN/dS ratio than the HE group (Mann–Whitney U test, *p*-value < 0.001 for both, Figure 22), suggesting stronger purifying selection pressures in the specific environment exposed. Conversely, the HE group's higher SNPs and dN/dS suggest exposure to a wider range of conditions and evolutionary pressure, leading to increased genetic variations.

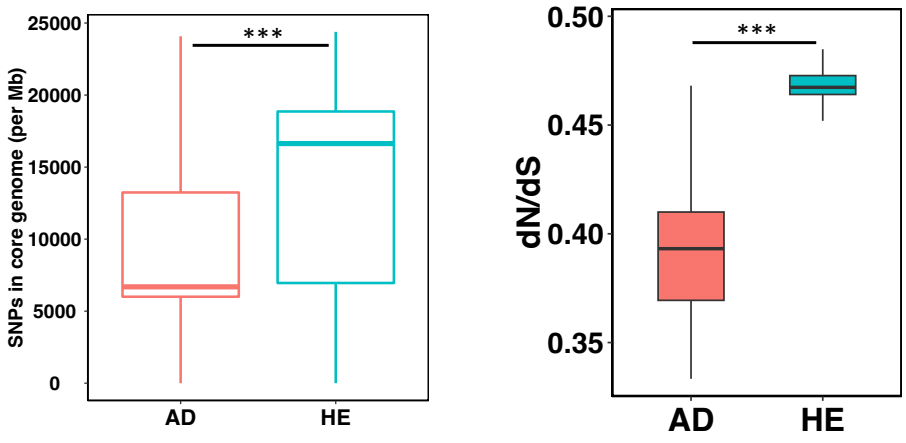

Figure 22. Collective results of SNPs and dN/dS analysis using strains with matched STs. Number of SNPs in core genome per Mb calculated using SNP-dists, and ratio of nonsynonymous and synonymous substitutions (dN/dS) calculated using codeML based on the core genomes of AD and HE strains with matched STs, respectively. The significance was performed by the Wilcoxon rank sum test, following the Kolmogorov-Smirnov test for the normality test of the data. *** means a *p*-value < 0.001.

To further reduce the influence of relative genetic distance, we performed a detailed examination at the ST level that highlighted four STs (ST1, 5, 30, and 45) with a relatively balanced distribution between AD and HE groups (Table S1). For SNPs, ST1, 5, and 45 demonstrated significantly fewer numbers in AD strains, aligning with the combined result, while ST30, despite a higher SNP count in HE strains, did not show a significant disparity (Figure 23A-D). However, individual analysis of the four STs revealed divergent trends in terms of dN/dS ratio, with ST1 and ST5 mirroring the collective trend, whereas ST30 and ST45 deviating (Figure 23E-H).

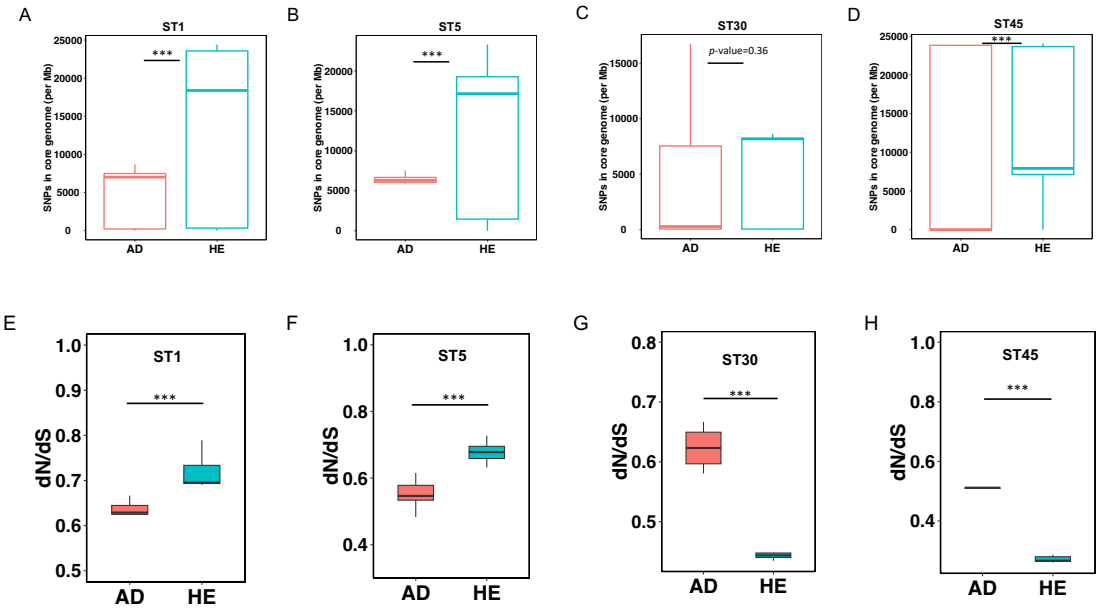

Figure 23. ST-level results of SNPs and dN/dS analysis. (A-D) Number of SNPs calculated using SNP-dists based on the core genomes of the four balanced STs (ST1, 5, 30, 45), respectively. The significance was performed by the Wilcoxon rank sum test, following the Kolmogorov-Smirnov test for the normality test of the data. *** means a *p*-value < 0.001. (E-H) Ratio of nonsynonymous and synonymous substitutions (dN/dS) calculated using codeML based on the core genomes of the four balanced STs (ST1, 5, 30, 45), respectively. The significance was performed by the Wilcoxon rank sum test, following the Kolmogorov-Smirnov test for the normality test. *** means a *p*-value < 0.001.

3.11 Prophages significantly contribute to the genetic differences of *S. aureus* strains

Having observed the significant contribution of prophages to the feature gene set, we sought to investigate how prophage genes influence the gene content of *S. aureus* genomes. We only focused on the non-rare gene families (present in >10% of strains) to enhance the generalizability and robustness of the model and reduce stochastic effects, as the RF classifier assigned an importance value to each gene family in the process, which was classified into two categories as differential (positive contribution to differentiation between AD and HE groups) and non-differential. Of these, 838 gene families were differential and 2144 non-differential (Figure 24A). We identified prophage genomes in most human-derived *S. aureus* strains (Table 3). By blasting the *S. aureus* genes against the predicted prophage genes, we revealed a striking shift from 5.2% phage-related genes in the non-differential gene set to an elevated 46.8% in the differential set. Besides, gene content analysis revealed a significantly higher number of gene families in the strains from HE compared to the strains from AD (Figure 24B), despite similar genome size (Figure 24C).

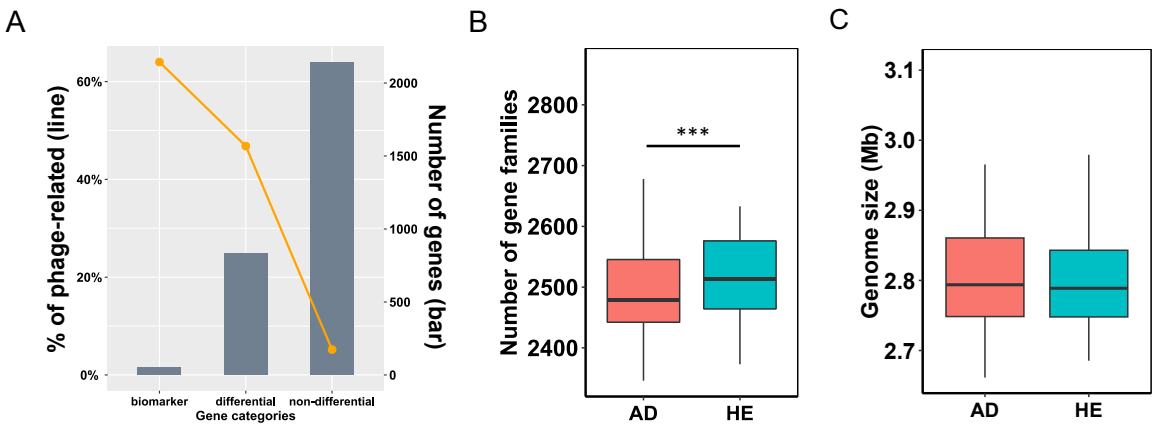

Figure 24. Prophages significantly contribute to the genetic differences of *S. aureus* strains. (A) Proportion (left) and number (right) of prophage-related genes in the marker, differential, and non-differential gene sets. Differential genes are those that contribute to the differentiation of AD and HE strains indicated by the RF classifier. Non-differential genes are the opposite. (B) Number of gene families of AD and HE strains, (C) Genome size of *S. aureus* strains from AD and HE groups. No significant difference was detected between the two groups. The significance was performed by the Wilcoxon rank sum test, following the Kolmogorov-Smirnov test for the normality test of the data. *** means a *p*-value < 0.001.

3.12 Prophage profiles of AD and HE *S. aureus*

Our findings highlighted the significant role of prophages in the genetic diversity of *S. aureus* strains. By profiling the prophages of *S. aureus*, defined with high-quality (HQ) genomes (completeness > 90%) and all predicted sequences (including remnants), we sought to uncover their prevalence and potential genetic influence, offering precise functional insights and a deeper understanding of their evolutionary and functional roles in *S. aureus*.

Table 3. Information on the predicted prophages of all 348 *S. aureus* strains.

	AD	**HE**
Number of strains	183	165
Number of strains having prophage	176	165
Number of total prophages predicted	555	541
Number of HQ prophages	133	163
Mean of prophages per strain	3.15	3.28
Median of prophages per strain	3	3
prophage average length (bp)	25633	29631
prophage max length (bp)	74286	67389
prophage min length (bp)	2391	4305

First, we observed the omnipresence of prophage sequences in our strains, with every strain from HE and all but 7 strains from AD containing predictable prophage sequences (Table 3). When focusing on HQ prophages, we identified 133 and 163 HQ prophage genomes in AD and HE strains, respectively (Table 3). A significantly higher number of HQ prophages per strain was observed in the HE group (Wilcoxon rank sum test, p-value < 0.05, Figure 25A). However, no major difference in HQ prophage length was observed between the two groups (Figure 25B). Our network analysis of the HQ prophages did not reveal dissimilarities at the genus level between AD and HE strains of *S. aureus* (Figure 25C).

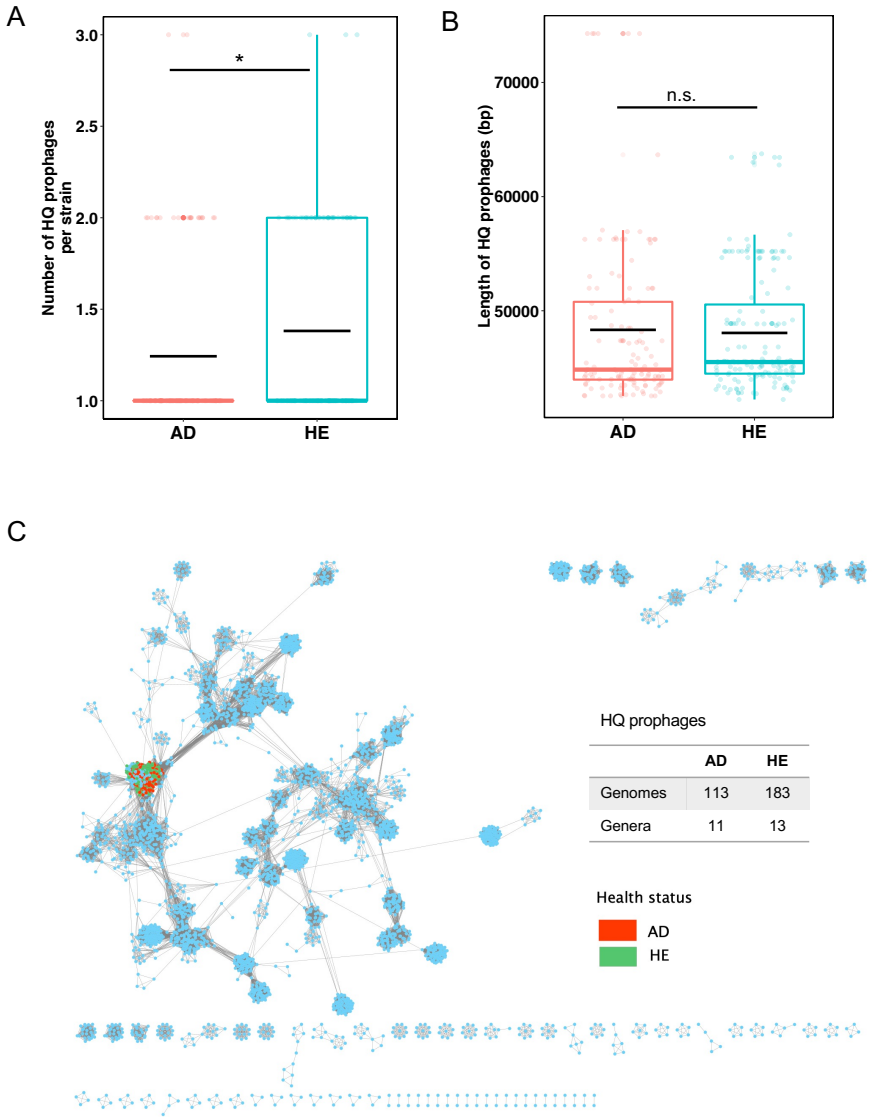

Figure 25. High-quality prophages in AD and HE groups display comparable size and taxonomy but differ in number. (A) Number of HQ prophages per strain, based on the combined prediction results of PhiSpy, VIBRANT, and Phigaro. Black crossbar in the box indicates the average number. (B) Length of high-quality prophage sequences predicted in AD and HE groups. Black crossbar in the box indicates the average number. The significance of the difference was performed by the Wilcoxon rank sum test. * means a *p*-value < 0.05. (C) Clustering analysis of high-quality prophages using vConTACT2. Each circle (node) represents a prophage sequence and connecting lines (edges) represent the similarity between sequences based on shared clusters of proteins. Sequences are clustered at the genus level. Red and green dots represent prophage sequences predicted in the *S. aureus* strains from AD and HE groups, respectively. Light blue dots represent prophage sequences from the reference database. Number of prophage sequences and genera for AD and HE strains clustered in the network is shown on the right.

To fully characterize the prophage gene content as well as functional implications, we analyzed all predicted prophage sequences in more detail. Prophage gene content in HE strains exhibited significantly richer diversity (Wilcoxon rank sum test, *p*-value < 0.0001, Figure 26A), with 976 clusters identified (257 unique to the tested strains), compared to 867 clusters (148 unique) in the AD group (Figure 26B). This result, combined with the pronounced core genome diversity of HE-derived strains (Figure 22), points to a potential co-evolution between the bacterial genome and its prophages in HE strains. Using the PHROG database for functional implications of the prophages, most functional categories were more abundant in HE-associated prophages than in AD (Figure 26C), except for the moron, auxiliary metabolic gene, and host takeover that slightly predominated in the prophages from AD. This underscores that strains from HE might have encountered more diverse environmental challenges or possess a longer evolutionary lineage, which led to the assimilation of versatile functional genes. Examining differential functional genes of prophages between AD and HE, we pinpointed 45 genes with a differential count greater than 30 using the PHROG database (Figure 26D); 7 genes were more prevalent in prophages from AD, while 38 were predominant in prophages from HE. Notably, the category related to DNA, RNA, and nucleotide metabolism was more abundant in HE prophages, while virulence traits like enterotoxin type A were more concentrated in AD prophages, suggesting these prophages might amplify the virulence potential of their bacterial hosts.

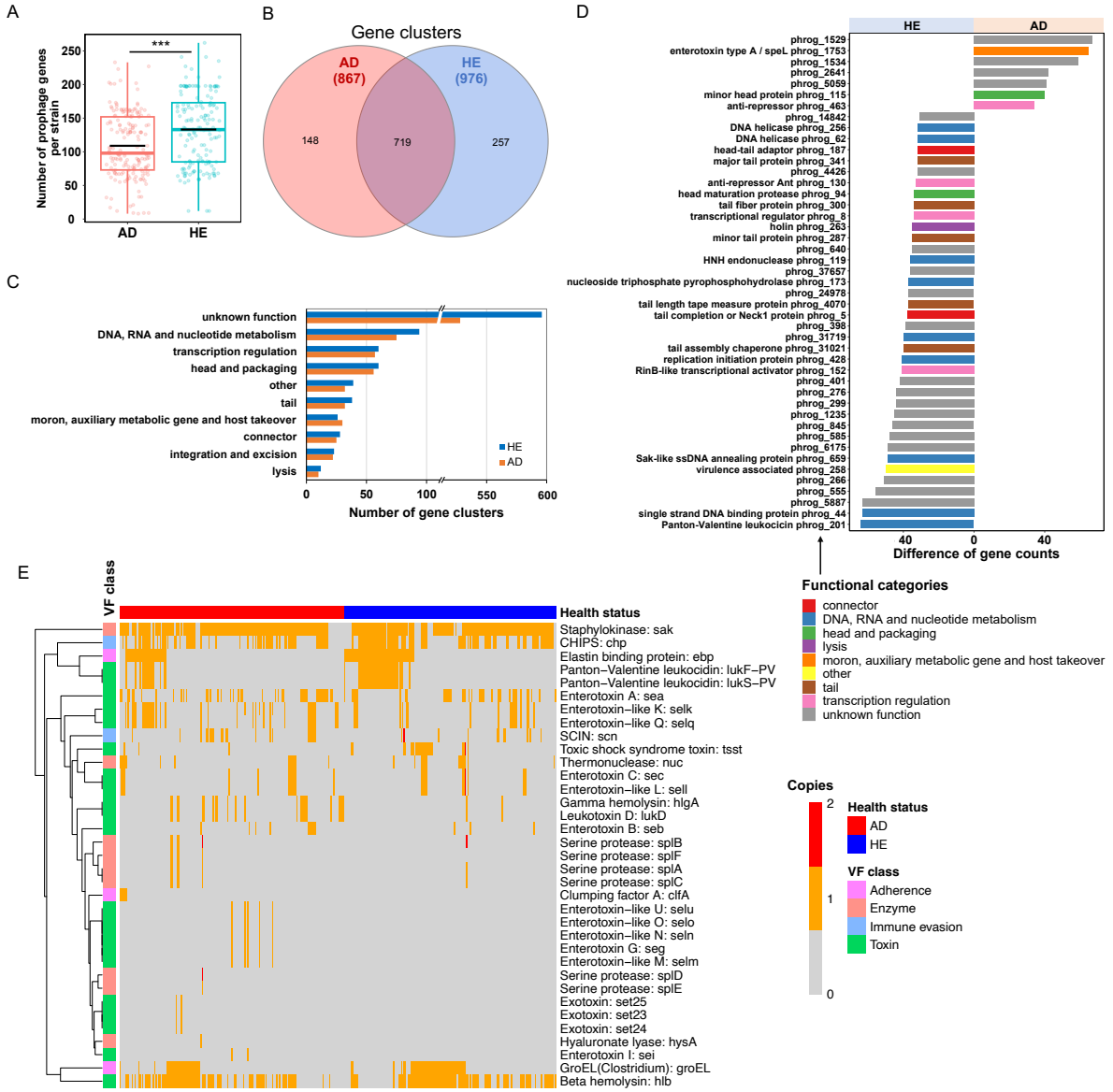

Figure 26. Prophage genes in AD and HE groups differ in gene content and associated functional implications based on all predicted prophage sequences. (A) Number of prophage gene clusters each strain possesses, assigned using MCL. Black crossbar in the box indicates the average number. * means a *p*-value < 0.05. (B) Number of shared and unique prophage gene clusters in AD and HE groups, respectively. The number in the parenthesis represents the total number of prophage clusters identified. (C) Number of prophage gene clusters assigned into PHROG functional categories for AD (orange) and HE (blue) prophages. (D) Differential functions between AD and HE prophages (gene count difference > 30), annotated with PHROG database. Bar colors show functional categories. (E) Heatmap of virulence factors carried in prophages of *S. aureus* strains in AD and HE groups, identified using VFanalyzer. Heatmap colors indicate the number of gene copies. Row clustering was calculated with the Euclidean method.

We further profiled HQ prophage genomes in *S. aureus*, analyzing gene content, species, and functional implications for a more biologically meaningful perspective. We identified 107 prophage species using the 95% similarity cutoff. The number of prophage species per strain was significantly higher in HE than in AD group (Figure 27A), with AD-associated strains in total housing 60 species (50 unique to the tested strains). HE-associated strains possessed 57 species, of which 47 were unique (Figure 27B). Additionally, HQ prophages in the HE group demonstrated a significantly richer diversity of gene clusters per strain, which is consistent with the result of all predicted sequences, while a total of 556 clusters (92 unique) were identified in the AD group, compared to 547 clusters (83 unique) in the HE group (Figure 27C&D). Notably, two more known functional categories, transcription regulation as well as integration and excision, were more abundant in the HQ prophages of AD strains, besides moron, auxiliary metabolic gene and host takeover (Figure 27E). Differential functional gene analysis identified enterotoxin type A was also more prevalent in HQ prophages of AD strains, while the category related to connector and DNA, RNA, and nucleotide metabolism was more abundant in HE prophages (Figure 27F). These findings largely align with those from all prophage sequences.

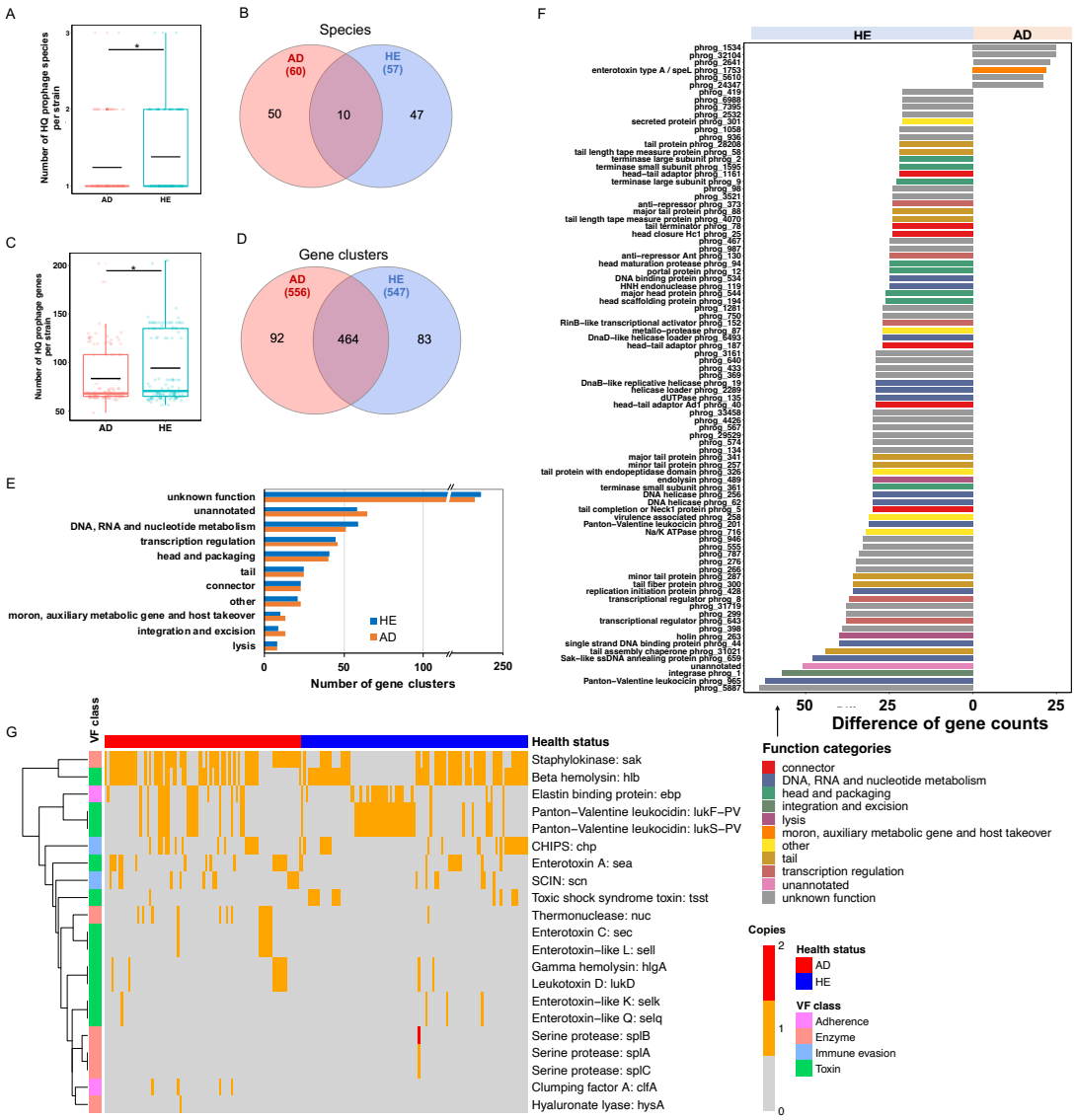

Figure 27. High-quality prophages in AD and HE groups differ in gene content and associated functional implications. (A) Number of high-quality phage species each strain harbored, identified by sequence clustering using CD-HIT with the 95% similarity cutoff. Black crossbar in the box indicates the average number. * means a *p*-value < 0.05. (B) Number of shared and unique species for AD and HE prophages, respectively. The number in the parenthesis represents the total number of prophage species identified. (C) Number of high-quality prophage gene clusters each strain possesses, assigned using MCL. Black crossbar in the box indicates the average number. * means a *p*-value < 0.05. (D) Number of shared and unique prophage gene clusters in AD and HE groups, respectively. The number in the parenthesis represents the total number of prophage clusters identified. (E) Number of prophage gene clusters assigned into PHROG functional categories for AD (orange) and HE (blue) prophages. (F) Differential functions between AD and HE prophages (gene count difference > 20), annotated with PHROG database. Bar colors show functional categories. (G) Heatmap of virulence factors carried in prophages of *S. aureus* strains in AD and HE groups, identified using VFanalyzer. Heatmap colors indicate the number of gene copies. Row clustering was calculated with the Euclidean method.

3.13 VFs enriched in prophages from AD-associated *S. aureus* strains

Having observed a strong correlation between virulence factors (VFs) and prophage genes, we speculated VFs in prophages might significantly influence *S. aureus* virulence. Examination of VFs carried in all prophage sequences indicated a higher representation of enzyme and toxin-linked VFs in AD strains, whereas adherence and immune evasion VFs were more prevalent in HE strains (Figure 28). In HQ prophages, enzyme-related VFs were more common in the AD group, while toxin-related VFs predominated in the HE group (Figure 28).

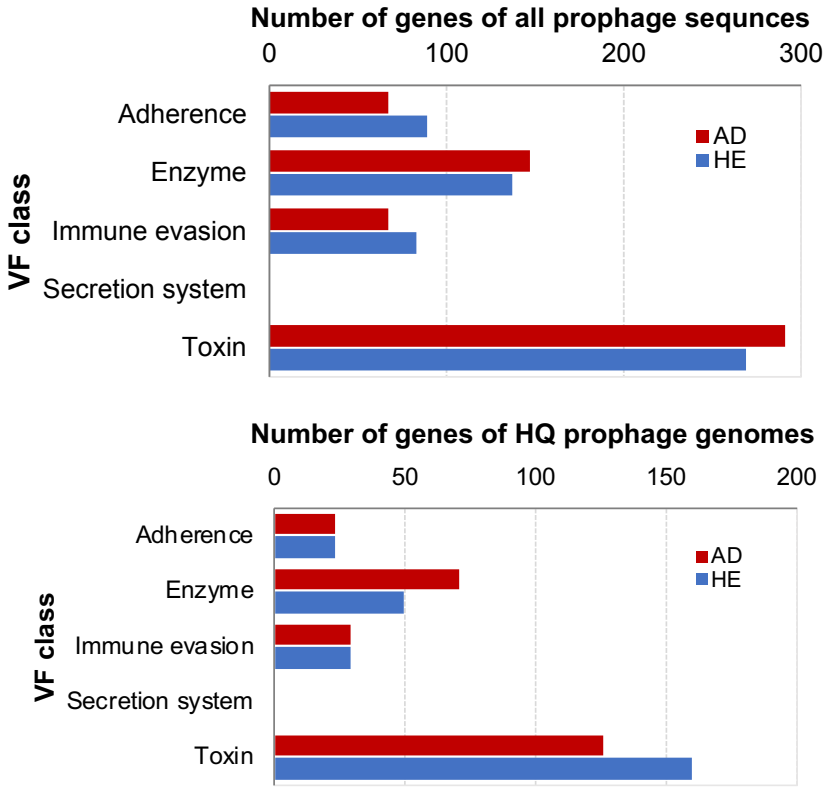

Figure 28. Number of genes coded by all prophage sequences and HQ prophage genomes, assigned into each virulence factor class in AD and HE groups, respectively.

Analyzing the specific VF genes in all prophage sequences in more detail, over 96% of prophage-containing strains in both groups encode VFs in their prophages, underscoring the omnipresence of prophage-carrying VFs. As Figure 26E depicted, over 70% of VFs were predominantly or exclusively associated with AD prophages, while a mere 16% were more

prevalent in HE prophages. AD-associated VFs predominantly include serine protease-like enzymes (*splA - splF*), which degrade fibrinogen and plasma proteins, weakening skin barriers and enhancing inflammation. Enterotoxins (*sea - sec, seg, sei*) and enterotoxin-like enzymes (*SelKMNOPQU*), also AD-dominant, cause food poisoning and provoke severe inflammatory responses. The role of exotoxins (Set23 - Set25), unique to AD prophages, might induce a substantial immune response and inflammation. Clumping factor A (*clfA*), exclusive in AD prophages, aids bacterial adherence and enhances colonization. Hyaluronate lyase (*hysA*) deteriorates the skin barrier, promoting bacterial invasion. Three other significant AD-dominant VFs are gamma hemolysin (*hlgA*) and leukotoxin D (*lukD*), both of which damage immune cells and weaken skin defense, and staphylococcal complement inhibitor (SCIN), which protects *S. aureus* from the immune system. Overall, the prominence of these VFs in AD-associated prophage sequences underscores the critical role of prophages in the pathogenicity of *S. aureus* strains in AD patients.

HE-associated prophages primarily contain VFs such as the chemotaxis inhibitory protein (CHIPS), which blocks immune cells from infection sites. Elastin binding protein (*ebp*) promotes bacterial adherence and tissue invasion. Panton-Valentine leukocidin (*LukF-PV* and *LukS-PV*) targets leukocytes, contributing to severe skin infections. The toxic shock syndrome toxin (*tsst*), highly abundant in HE prophages, is a potent superantigen triggering a strong immune response, potentially causing toxic shock syndrome.

The analysis of VFs encoded by HQ prophages indicates a divergence in prevalence between AD and HE strains (Figure 27G). VFs such as Staphylokinase (*sak*), Thermonuclease (*nuc*), and SCIN are predominantly found in AD strains. Toxins-related genes like *hlgA* and *lukD* also show higher occurrence in AD. Uniquely present in AD strains are genes like *clfA*, *hysA*, *sec*, and *sell*. In contrast, genes such as *ebp*, Beta hemolysin (*hlb*), CHIPS, and both *lukF-PV* and *lukS-PV* genes exhibit a higher presence in HE strains, echoing the trends observed across all prophage sequences. Notably, *splA - splC*, *selk*, and *selq* were unexpectedly more prevalent in HE strains based on HQ prophages, deviating from the overall prophage findings. These patterns underscore the potential role of specific genes in the pathogenesis of AD, largely aligning with the comprehensive prophage data.

4 Discussion

In this study, we conducted a comprehensive WGS-based analysis on a collection of 348 *S. aureus* strains in the context of AD and HE. These results have yielded pivotal insights into the genomic and functional divergences between *S. aureus* strains from AD and HE, at both local and global scales. The study not only illuminated a notable discrepancy in the gene content and functional variations between AD and HE strains, but also highlighted distinct mechanisms of functional differentiation at local and global scales. Additionally, the investigation has revealed strong evidence of HGT events and gene enrichment as key drivers for *S. aureus* diversification. By utilizing the state-of-the-art ML algorithm, we construct and employ RF classifiers and in-depth prophage profiling, our research has identified a collection of marker genes that can accurately differentiate AD-associated strains from those associated with HE, which underscores the critical role of prophages in the genetic differentiation and pathogenicity of *S. aureus* in the context of AD conditions, offering unprecedented insight into the subtle genomic distinctions of this bacterium in clinical settings. The findings from this study enhance our comprehension of the genomic and functional complexity of *S. aureus* strains associated with AD.

4.1 Advances in sequencing and computational techniques enhance microbiome analysis

The field of microbial genomics has experienced an unprecedented surge in the past decade, thanks to the rapid advancements in sequencing technologies and bioinformatics methods. This transformative progress has not only reshaped our understanding of microbial genomics but has also offered novel insights into their interactions and therapeutic strategies.

Modern microbiology heavily leans on sequencing technologies to decipher the genetic blueprints of microorganisms and conduct large-scale microbial studies from basic research to clinical diagnostics (Bansal and Boucher, 2019). The next-generation sequencing (NGS) platforms, such as Illumina, deliver rapid and high-throughput sequencing capabilities, offering unprecedented sequencing depth at reduced costs to facilitate large-scale genomic studies (Metzker, 2010; Slatko *et al.*, 2018). For microbiologists, NGS has unlocked the potential to study microbial communities *en masse* and transformed our understanding of microbial diversity in various niches, from the human organs to the deepest oceans (Wensel *et al.*, 2022). Another pivotal advancement lies in genome assembly and annotation, where bioinformatics tools and algorithms construct and characterize genomic sequences by managing a

combination of massive short reads, enhancing the assembly's completeness and accuracy. From these data, information on microbial typing, resistance, and virulence is garnered, enabling comprehensive functional elements mapping across diverse species and aiding outbreak investigations (Deurenberg *et al.*, 2017).

With the advent of TGS platforms like PacBio and Oxford Nanopore, the shortcomings of short-read sequencing are being addressed (van Dijk *et al.*, 2018). By providing considerably longer read lengths compared to NGS, TGS techniques have provided an opportunity to understand the genetic and functional diversity of the skin microbiota at the finest resolution by producing genome assemblies of unprecedented quality (Rhoads and Au, 2015). Furthermore, the absence of PCR amplification reduces bias and leads to more uniform coverage across the genome. The introduction of PacBio Single Molecule Real-Time (SMRT) sequencing has empowered researchers to delve deeper into complex genomic structures and unravel intricate genomic patterns. The ability of TGS to provide reads that span several kilobases enables sequencing through repetitive and GC-rich regions, facilitating the resolution of structurally complex genomic areas, and enhancing de novo genome assembly (Bansal and Boucher, 2019).

WGS has revolutionized microbial genomics, offering an unrivaled depth into the genetic landscapes of microorganisms (Kwong *et al.*, 2015). WGS deciphers the entire DNA sequence of a microbial organism, offering not only critical insights beyond mere species identification but also the capability to differentiate between closely related microbial strains (Lan and Reeves, 2000). Through WGS, researchers can elucidate intricate functional details like VFs and antibiotic resistance mechanisms (Mason *et al.*, 2018). WGS also aids in the discovery of novel genes, potentially leading to the uncovering of new metabolic pathways and microbial capabilities (Kuroda *et al.*, 2001). This comprehensive genetic blueprint not only aids in understanding microbial biology and evolution but also empowers applications in public health, where tracking and managing disease outbreaks hinge on the genomic characterization of pathogens (Quainoo *et al.*, 2017). Combined with NGS and TGS as they advance and become more accessible, WGS stands as a cornerstone in microbial studies, bridging gaps in knowledge and paving the way for innovative solutions in clinical microbiology and medicine.

ML, with its capability to unearth patterns from complex and high-dimensional data, holds immense potential in microbial genomics, offering novel insights and fostering predictive modeling (Ghannam and Techtmann, 2021; Hernández Medina *et al.*, 2022). In modern microbiology studies, the complexity of experimental data has reached a point where the incorporation of ML is not only necessary but also presents an array of opportunities that span from medical diagnostics to the discovery of valuable biomarkers. ML algorithms have been

proposed as powerful and informative tools in the phenotyping and microbial feature classification (Gupta and Gupta, 2021; Marcos-Zambrano *et al.*, 2021). Biological samples contain both outcomes (information about the sample and its context) and extractable features (e.g., gene sequences, phenotypic data). Together, outcomes and features are used as training examples, which serve as input for the ML model. During the training phase, the algorithm uses these examples to train the model, enabling it to make predictions subsequently for new samples with known features but unknown outcomes. For instance, an ML classifier trained on gene presence/absence data could predict the species of a new isolate based on its gene features, even if taxonomic information is lacking, or to predict functional or phenotypic characteristics (pathogenic and non-pathogenic status). For example, an ML classifier, trained using 50 gut microbial gene markers, successfully differentiated individuals with type 2 diabetes from healthy controls (Qin *et al.*, 2012).

Typical applications of ML approaches that utilize genomic data as features are to predict antibiotic resistance. ML prediction performance often relies on the training set size. In the case of detecting antibiotic resistance genes in *Escherichia coli*, where ample experimental data exist, the effectiveness of the approach is specifically notable (Moradigaravand *et al.*, 2018). Powered by NGS and WGS, high-dimensional quantitative features can depict a sample, such as the relative abundance of numerous species, gene presence or absence profiles, or detected single-nucleotide variants in specific genes. In this context, ML methods often aim to differentiate between clinical cases and controls to guide mechanistic experiments or the development of novel diagnostic tools (Asnicar *et al.*, 2024). In functional genomics, ML algorithms process vast amounts of genomic data to discern patterns and associations often overlooked by traditional statistical methods, with significant applications in personalized medicine. By examining patient-specific genomic data, ML models not only help map out potential therapeutic strategies and identify drug targets but also forecast patient reactions to particular treatments. For example, researchers developed an ML model to predict the risk of MRSA in hospitalized community-acquired pneumonia patients, finding that ICU-admitted patients needing ventilation had a higher MRSA risk, while certain factors reduced this risk; the model proved highly accurate in its predictions (Rhodes *et al.*, 2023).

The RF algorithm is a versatile and robust ML method (Breiman, 2001) and has found significant application in microbial studies (Méric *et al.*, 2018; S. Wang *et al.*, 2022). Its ensemble nature, which involves multiple decision trees voting on an outcome, offers high accuracy and handles the complexity of microbial data effectively. For instance, it has been employed to predict phenotypic outcomes, such as disease susceptibility or antibiotic resistance in bacteria based on genomic markers (Méric *et al.*, 2018; S. Wang *et al.*, 2022),

thereby pushing the boundaries of precision medicine. Beyond resistance prediction, RF has also been utilized in microbiome analyses, classifying different microbial communities based on environmental or clinical factors (Beck and Foster, 2014; Ghannam and Techtmann, 2021; Zhu *et al.*, 2022). In different applications, the RF method was employed to assess the health status of dental implants, which can provide diagnostic and prognostic potential (Ghensi *et al.*, 2020). It also enabled predictions of overall mortality risk through training on human gut microbiome features (Salosensaari *et al.*, 2021). By identifying subtle genetic differences between closely related microbial strains, RF offers novel perspectives on creating models for classification and prediction in microbiology, deducing host phenotypes for disease prediction, and categorizing patients based on distinct microbial signatures (Marcos-Zambrano *et al.*, 2021).

Several bioinformatics tools now integrate ML algorithms. Among them, platforms that incorporate RF, like Microbiome Helper (Comeau *et al.*, 2017), are widely used for genomic data classification and feature importance analysis. The combination of RF and microbial genomics promises enhanced predictive modeling, offering avenues for personalized medical interventions and better microbial resource management. However, as with all ML models, careful validation and interpretation are crucial to avoid overfitting and ensure meaningful results.

4.2 Gene content analysis implicates selective adaptation and functional diversity

The comparative genomic analysis elucidated that strains of *S. aureus* associated with AD were distinguished by a reduced pan-genome size, coupled with a significantly elevated proportion of core genes at the strain level, observable both within local and global contexts. This phenomenon is attributed mainly to the diminished presence of accessory genes in AD-associated strains. The constrained pan-genome and augmented core gene presence in AD-associated strains suggest that the unique microenvironment of AD-affected skin exerts selective pressure, favoring the survival of *S. aureus* strains equipped with a highly specialized or adaptively optimized gene set. Supporting this observation, a recent investigation by Conte and his colleagues identified a comparable trend in AD-group strains isolated from a local clinic in Rome, Italy , reinforcing the notion that such selective pressures refine the genetic composition of these bacteria to enhance their adaptability to the AD microenvironment (Conte *et al.*, 2023).

On the other hand, the higher distance within AD strains in the global dataset suggests that, despite a shared core genome, there is a significant diversity in the accessory genome among

the AD strains. This diversity in the accessory genes could be driven by HGT events, or be the result of differing local adaptation pressures. AD skin presents a unique environment, e.g., altered skin pH (Ali and Yosipovitch, 2013; Hülpüsch *et al.*, 2020), different sebum composition (Rajka, 1974; Foti *et al.*, 2015; Yin *et al.*, 2023), compromised skin barrier (Marsella *et al.*, 2011; Agrawal and Woodfolk, 2014; Nakatsuji *et al.*, 2016), or distinctive immune response (Nakamura *et al.*, 2013; Riethmuller *et al.*, 2015; Werfel *et al.*, 2016), that provides a selective advantage for certain *S. aureus* strains depending on their specific microenvironment. Studies have shown that CC1 strains (Geoghegan *et al.*, 2018) or ST1 and ST8 (Saheb Kashaf *et al.*, 2023) were significantly enriched in AD patients, while CC30 strains are common in the general population (Geoghegan *et al.*, 2018). The observed outcome aligns with the presence of two additional clusters, thereby elucidating the diversification of AD strains, a process underpinned by gene enrichment and horizontal gene transfer.

Despite that AD strains demonstrated more variability than HE strains for the global dataset, *S. aureus* strains from HE in the local German cohort inversely exhibited a broader functional diversity compared to their AD counterparts. For local strains, the observed genetic diversity in HE strains could be elucidated through the lens of transmission dynamics. HE-associated strains are predominantly linked with asymptomatic carriage and colonization (Knox *et al.*, 2015; Chmielowiec-Korzeniowska *et al.*, 2020), potentially facilitating elevated transmission rates within these communities. The increased transmission and exchange of genetic material among *S. aureus* strains from HE in the local community may contribute to their higher diversity. However, It is also plausible that the variance between local and global *S. aureus* strain diversity patterns might also be influenced by the limited size of the cohorts examined in both the present study (n=48) and the research conducted in Rome (n=38) (Conte *et al.*, 2023), suggesting a need for broader-scale studies to fully understand the dynamics at play.

Furthermore, the implications underscore the critical need to account for both microenvironmental conditions and community dynamics in understanding the evolution and adaptation of *S. aureus* strains. This underscores the necessity of a multifaceted approach in microbial genomics research, integrating considerations such as genetic drift, selective pressure, and HGT, to decode the complex interactions between host environments and bacterial genetics. Insights into the transmission dynamics and genetic diversity of strains from HE are invaluable for elucidating the epidemiology of *S. aureus* and devising strategies to prevent its spread within communities. Identifying genetic markers that provide *S. aureus* strains with adaptability and survival advantages in AD-affected skin, while concurrently reducing their gene content diversity, offers promising targets for the development of innovative treatments. Future research should focus on enlarging the cohort sizes and

enhancing the geographical diversity of the sampled strains, employing advanced computational tools to affirm these findings and assess their relevance across wider populations. Such efforts will not only augment our understanding of the adaptive and evolutionary mechanisms of *S. aureus* but also aid in creating more effective public health interventions.

4.3 Multiple drivers shape *S. aureus* strains

The diversification of gene content, potentially influenced by natural selection at both micro and macro levels, is pivotal in determining the functional capabilities of *S. aureus* isolates and the environmental constraints imposed by the host (Van Rossum *et al.*, 2020). This study not only highlights the differences between AD and HE groups but also uncovers distinctions among host individuals at the microscale level and pronounced disparities across different countries, indicating that *S. aureus* diversification is influenced by a multifaceted array of factors.

Extant research underscores the significant impact of both health status (Kong *et al.*, 2012; Byrd *et al.*, 2017; Geoghegan *et al.*, 2018) and geographical location (Carter *et al.*, 2018; Saheb Kashaf *et al.*, 2023) on the genetic diversity of *S. aureus* strains, revealing a nuanced landscape of strain variation across different regions. Our analysis further elucidates this by demonstrating a pronounced diversity in American strains relative to those from Japan, Australia, and the UK. Divergent STs of *S. aureus*, as reported in recent comprehensive studies (Saheb Kashaf *et al.*, 2023), predominate differently by country, highlighting the geographical influence on strain prevalence and diversity. For instance, while ST1 is widespread in Europe and Singapore, it is notably absent in the US, where ST5 prevails. We also examined the role of health status within-country, which revealed greater diversity in the HE group for the US cohorts (Mann-Whitney U-test, p-value = 0.055, Figure 7C) and significantly greater diversity in the HE group for the German local cohorts (Figure 13F). This finding suggests this geographical patterning extends to within-country analyses of health status and disease status influences strain variation independently of geographical factors.

Moreover, recent studies have illuminated the phenomenon that individuals with AD are stably colonized by a singular major lineage of *S. aureus* (Key *et al.*, 2023), emphasizing the role of host-specific factors in determining gene content diversity. This insight is paralleled by comprehensive genomic and metagenomic analyses of *Staphylococcus epidermidis*, a closely related species also implicated in AD (Byrd *et al.*, 2017), which have identified a clear pattern of host specificity influencing gene content diversity (Zhou *et al.*, 2020). Our own research,

through phylogenetic analysis and PCA clustering of *S. aureus* strains from Augsburg (Figure 13), reinforces these findings by demonstrating a distinct pattern of gene content diversity that appears to be personalized to each host. This suggests that *S. aureus* strains may manifest host-specific adaptations. However, the scope of our study, limited by the small number of samples analyzed per individual (n = 3-9), introduces a degree of uncertainty regarding the generalizability of these host-specific characteristics across broader *S. aureus* populations. Despite this limitation, the insights garnered provide a crucial foundation for future research aimed at exploring this phenomenon with a more extensive dataset and a larger number of strains. The potential implications of these findings are profound, suggesting a pathway towards the development of personalized therapeutic approaches tailored to the unique microbial colonization patterns of individuals with AD. These findings highlight the need for an integrated approach that considers the complex interplay between host environments, health status, and geographical factors in microbial genomics research.

4.4 Scale-dependent mechanisms in *S. aureus* diversification

This study's functional analysis elucidates critical dimensions of *S. aureus* functional differentiation, evident at both local and global levels, revealing intricate mechanisms of bacterial adaptation and resistance. Primarily, the Replication and Repair pathway emerged as a pivotal differentiator across datasets, with a notable emphasis on the role of transposase enrichment. This pathway integrates a range of enzymatic processes and molecular mechanisms that are essential for DNA replication, repair, and recombination (Alberts, 2003; Ha and Edwards, 2021). Intriguingly, research suggests that various antibiotic classes commonly used to treat staphylococcal infections (Rayner and Munckhof, 2005; Maiques *et al.*, 2006) may lead to bacterial cell death by causing DNA damage (Kohanski *et al.*, 2007; Liu *et al.*, 2012; Dwyer *et al.*, 2014). Transposases are believed to play a crucial role in this context, aiding in the repair of DNA damage and facilitating genetic variability through the excision and subsequent repair of damaged DNA segments (Cuecas *et al.*, 2017). The observed transposase enrichment, particularly within AD-dominant Cluster 1 strains, suggests a strategic bacterial response to counteract antibiotic-induced DNA damage. This mechanism potentially facilitates bacterial survival by enabling the repair and recombination of DNA, thereby contributing to genetic diversity and resilience against antibiotic pressure.

Furthermore, the Transporters pathway was identified as another critical pathway distinguishing *S. aureus* strains. This pathway's significance was underscored by the identification of distinct functional orthologs in local versus global contexts. Local strains

demonstrated a propensity for genes related to arsenical and ATP-binding cassette transporters such as *arsB* (arsenical pump membrane protein) and *abcA* (ATP-binding cassette, subfamily B), whereas global strains exhibited a diverse array of genes involved in the 'transport of bacitracin, arginine, and peptides/nickel, bacitracin' (*bcrA* and *bcrB*), arginine (*arcD*), and peptide/nickel transport (ABC.PE.P, ABC.PE.P1, *ddpD*, and *ddpF*). These transport mechanisms are crucial for bacterial survival, facilitating the efflux of harmful substances, including antibiotics, and thereby contributing to antimicrobial resistance (Broer *et al.*, 1993; Schrader-Fischer and Berger-Bächi, 2001).

Additionally, the analysis highlighted a significant focus on metal-related genes in the local dataset, including *cadC*, *arsR*, *copB*, and *zntA*, indicating the pivotal role of metal ion transport and regulation in bacterial differentiation at the local scale. In contrast, the global dataset revealed the prominence of ARGs against macrolide, aminoglycoside, streptothricin, and methicillin, along with lantibiotic genes as the important orthologs driving the diversification of the AD-dominant Cluster 2 strains from the US, reflecting the impact of high antibiotic prescription rates in regions like the US and the consequential selection pressure on bacterial populations (Centers for Disease Control and Prevention, 2017). However, the distinction between the significance of metal ions in local strains versus antibiotic resistance in global strains calls for further investigation to understand these differential survival strategies more comprehensively.

Overall, these findings underscore the complexity of *S. aureus* functional differentiation, marked by scale-dependent divergence mechanisms. These insights not only enhance our understanding of bacterial adaptability and resistance but also highlight potential targets for therapeutic intervention, emphasizing the need for continued research to dissect these complex bacterial survival strategies further.

4.5 HGT in *S. aureus* diversification

HGT is a pivotal process in bacterial evolution and adaptation, acting as a fundamental mechanism through which bacteria can swiftly acquire new genetic information. This process enables the rapid dissemination of a wide range of genes, encompassing those responsible for antibiotic resistance, pathogenicity, and various metabolic functions (Jain, 2003; Soucy *et al.*, 2015; Arnold *et al.*, 2022). The impact of HGT on bacterial evolution is profound, offering a means for bacteria to swiftly adapt to fluctuating environmental conditions, such as the development of antibiotic-resistant strains in the wake of antibiotic utilization. This capacity for rapid adaptation renders bacteria remarkably resilient, allowing them to flourish across a

multitude of ecological settings. Additionally, HGT is a driving force behind the genetic diversity observed within bacterial populations, fostering evolutionary dynamics through the introduction of new genetic variations (Arnold *et al.*, 2022). This genetic diversity is crucial for survival under environmental pressures and presents considerable challenges in clinical contexts, especially regarding the control of bacterial infections and the formulation of effective antibiotic strategies (Lindsay, 2014).

The lantibiotic operon, known for the production of lantibiotics like nisin, consists of a gene cluster crucial for the biosynthesis and regulation of these antibacterial peptides (McAuliffe *et al.*, 2001). The detection of HGT involving the lantibiotic operon from the Bacillaceae family, prevalent in the environment (Mandic-Mulec *et al.*, 2015), underscores HGT's role in the spread of AMR genes. This finding underscores the intricate interactions between human-associated and environmental bacterial populations, shaping microbial diversity and adaptation. The implication of HGT in the spread of AMR genes extends beyond the lantibiotic operon, serving as a critical mechanism through which *S. aureus* strains acquires resistance to a broad spectrum of antibiotics. For instance, the *mecA* gene, responsible for methicillin resistance, is another example where HGT plays a pivotal role in the dissemination of resistance traits among bacterial populations (Berglund and Söderquist, 2008; Maree *et al.*, 2022). This gene is carried on the staphylococcal cassette chromosome *mec* (SCC*mec*), which has been transferred between different strains and species of *Staphylococcus*, leading to the emergence of MRSA strains (Ito *et al.*, 2001). Additionally, the *vanA* gene cluster, which confers resistance to vancomycin, has been observed to transfer between *Enterococcus* species and *S. aureus* through HGT, further complicating the treatment of staphylococcal infections (Courvalin, 2006). These examples underscore the crucial role of HGT in the acquisition of AMR genes by *S. aureus*, enabling the bacterium to adapt to and overcome antimicrobial pressures. This adaptive capability, facilitated by the exchange of genetic material between different bacterial species and strains, poses significant challenges to public health and underscores the need for ongoing research and development of novel antimicrobial strategies. Understanding the mechanisms and consequences of HGT is essential for predicting bacterial evolutionary trends and for the formulation of effective interventions to combat the spread of AMR.

HGT in bacterial communities can be mediated through three primary pathways: conjugation, which involves direct cell-to-cell DNA transfer; transformation, the uptake of extracellular naked DNA by competent cells; and transduction, the bacteriophage-mediated DNA transfer (Michaelis and Grohmann, 2023). Phage transduction is particularly noteworthy in *S. aureus*, as most isolates are lysogenized by bacteriophages, suggesting its prominence as an HGT

mechanism (Cafini *et al.*, 2017). The role of bacteriophages in diversifying the gene pool of *S. aureus* and thereby its differentiation, warrants further exploration to fully understand the contributions of phages to bacterial evolution and adaptation.

4.6 Identification of marker genes for AD-associated *S. aureus* using RF model

Observations have linked specific *S. aureus* clades with the health status of patients suffering from AD, indicating a correlation that underscores the complexity of strain differentiation within AD and HE contexts (Byrd *et al.*, 2017; Catriona P. Harkins *et al.*, 2018; Key *et al.*, 2023). Despite these correlations, achieving precise differentiation of *S. aureus* strains from AD and HE using whole-genome analysis poses significant challenges (N. Liu *et al.*, 2022). This difficulty suggests the presence of specific genes responsible for the variation observed, making the identification of such marker genes essential for understanding the mechanisms underlying strain differentiation.

In our study, we identified 50 marker genes that demonstrated high reliability in distinguishing AD-associated strains, achieving an overall accuracy and AUC of over 90% in both the test and real-world test datasets (Figure 13). This achievement marks a significant leap forward in distinguishing AD-associated *S. aureus* strains based on genetic markers. Our results show a notable improvement over previous studies that primarily focused on phenotyping for antibiotic resistance and susceptibility (Hirano *et al.*, 2021; Wang *et al.*, 2021; Rhodes *et al.*, 2023) or analyzing AD-derived microbiota (Jiang *et al.*, 2022). For instance, a machine learning model developed by Rhodes *et al*. predicted the risk of MRSA in patients with community-acquired pneumonia, achieving a ROC score of 77.5% (Rhodes *et al.*, 2023). Similarly, Wang *et al*. employed support vector machine models to differentiate between resistant and susceptible profiles for 10 antimicrobials, obtaining AUC scores ranging from 66.5% to 96.2 (Wang *et al.*, 2021). In clinical practice, the high accuracy and AUC values of machine learning models are essential for the prompt and precise identification of pathogenic strains, crucial for effective patient management. These performance metrics significantly reduce diagnostic errors, decrease healthcare costs, and foster clinician confidence, thereby enhancing the overall quality of healthcare delivery. The exceptional performance of our classifier highlights the critical role of machine learning in advancing microbial genomics research related to disease, suggesting its potential to transform future studies in this domain.

Prior research has explored the association between specific *S. aureus* CCs or STs and AD, yet has not succeeded in pinpointing a dominant clone or type among AD patient isolates. Techniques such as Pulsed Field Gel Electrophoresis (PFGE) have been utilized for

genotyping *S. aureus* strains, categorizing them into specific pulsotypes like A, B, or C (He *et al.*, 2014). Research has shown variability in the prevalence of these pulsotypes among AD patients, with one study identifying pulsotype B as the most common AD-associated type, accounting for 48% of isolates type (Lo *et al.*, 2010), and another study recognizing an additional 28 pulsotypes linked to AD (Lomholt *et al.*, 2005). Similarly, Multi-Locus Sequence Typing (MLST) has highlighted sequence types ST1 (19.4%) and ST5 (13.9%) as notably associated with AD (Kim *et al.*, 2009), though a thorough, more recent analysis has indicated that no single sequence type dominates the AD landscape globally (Saheb Kashaf *et al.*, 2023). One investigation using CC typing found CC1 to be most prevalent in AD cases at 20% (Catriona P. Harkins *et al.*, 2018), whereas another identified CC30 as making up a larger portion, at 33%, in a different cohort (Key *et al.*, 2023). These findings highlight the limitations of traditional genotyping methods in effectively distinguishing AD-related *S. aureus* strains, as the highest prevalence of any clone or type identified in these studies did not exceed 50%.

In contrast, our study introduces a novel approach by identifying a precise set of marker genes with exceptional accuracy in differentiating AD-associated *S. aureus* strains (Figure 15). This breakthrough significantly advances our ability to distinguish between strains associated with disease and health, offering new pathways for research into the genetic markers that underpin these differences. Additionally, these marker genes hold promise as therapeutic targets, suggesting that altering their expression could shift the microbiome towards a state more characteristic of healthy individuals, potentially alleviating AD symptoms or preventing its progression.

However, the challenge of 10 missing feature genes in some strains within the real-world dataset highlights a limitation in predicting strains from HE accurately. This challenge points to the genetic specificity and variability among *S. aureus* strains, suggesting a need for refinement of the marker gene set to better capture the diversity of strains associated with healthy states. Consequently, future applications of our model, particularly in AD research, may benefit from customizing the feature gene set. This might include a thorough reassessment of the marker genes to incorporate additional ones that reflect the broader genetic diversity of *S. aureus* strains in various contexts. Moreover, the methodology and principles underlying our model hold potential for broader applications beyond *S. aureus*, indicating an opportunity to adapt this approach for studying other bacterial species or strains. By leveraging our model, future research can develop tailored marker gene sets for diverse microbial studies, enhancing the model's utility across different research scenarios and providing a flexible framework for genomic analysis and strain differentiation.

4.7 Influence of clonal structure on *S. aureus* microdiversity

Purifying selection, also known as negative selection, is an evolutionary process that removes deleterious alleles from a population, thereby preserving the genetic stability and fitness of a species (Martinez-Gutierrez *et al.*, 2019). This mechanism favors the retention of beneficial or neutral mutations while eliminating harmful ones, ensuring the survival and reproductive success of the organism. Given the abnormally high abundance of *S. aureus* strains in AD conditions (Kong *et al.*, 2012; Altunbulakli *et al.*, 2018), purifying selection may significantly influence the population structure of *S. aureus* (Zapotoczna *et al.*, 2018; Howden *et al.*, 2023).

Our examination of *S. aureus* strains elucidates the significance of purifying selection in shaping bacterial microdiversity across varied environments. While previous research did not find notable differences in single SNPs between AD and HE groups (Catriona P. Harkins *et al.*, 2018; Edslev *et al.*, 2018), our findings reveal a distinct pattern. Specifically, strains from AD patients, when matched by STs, exhibited significantly fewer SNPs and a reduced ratio of dN/dS compared to those from HE. This indicates a stronger purifying selection within AD environments, as evidenced by a dN/dS ratio less than 1, which signifies negative selection (Kryazhimskiy and Plotkin, 2008) and points to a decrease in gene content diversity. In contrast, strains from HE showed higher SNPs and dN/dS ratios, reflecting the varied evolutionary pressures they encounter, likely due to a broader range of environmental and microbial interactions. These differences in evolutionary dynamics between AD and HE groups are crucial for understanding *S. aureus* pathogenesis in different host settings.

S. aureus presents a diversity of CCs or STs, with varying prevalence between AD patients and healthy individuals (Hwang *et al.*, 2021; Ogonowska *et al.*, 2021), raising questions about the consistency of microdiversity patterns across all STs: is the observed collective pattern of microdiversity consistent across all STs, or does it vary among them? Upon closer examination of specific STs, we found that SNP distributions at the ST level generally adhere to the overall observed trend. Among the studied STs (ST1, ST5, ST30, and ST45) in AD strains, all displayed reduced SNPs, though for ST30 difference in SNP counts between AD and HE strains was not significant. For accurate dN/dS comparisons between closely related strains or species, it is essential to consider the time elapsed since their divergence (Rocha *et al.*, 2006). To minimize the effect of divergence time, we conducted intra-ST comparisons. The result revealed that ST1 and ST5 align with the broader trend, whereas ST30 and ST45 diverged, possibly due to varying levels of purifying selection pressure or a potential sample size bias, especially for ST30 and ST45.

These findings indicate a generally consistent SNP pattern across different STs in *S. aureus*, yet reveal variability in dN/dS ratios among these STs, underscoring the complex evolutionary dynamics within the species. This suggests that clonal lineages significantly influence *S. aureus*'s microdiversity. However, confirming this pattern necessitates further analysis with a more diverse and systematically collected set of strains, to deepen our understanding of the evolutionary forces shaping *S. aureus* in different host environments.

4.8 Different prophages infect strains from AD and HE

Phages play a crucial role in regulating bacterial population dynamics through various mechanisms, significantly impacting bacterial abundance and diversity (Breitbart and Rohwer, 2005; Voigt *et al.*, 2021). Their most notable effect on bacterial populations is a substantial decrease in their numbers (Breitbart and Rohwer, 2005). Phages, which outnumber bacteria by approximately 20-30%, are responsible for around 10^{24} bacterial infections every second, leading to the removal of about 4-50% of the bacterial population in natural settings at any given time (Fuhrman and Noble, 1995). Furthermore, prophages have a significant influence on the structure of bacterial genomes, serving as sites for genomic rearrangements and gene disruptions. Prophages also mark historical interactions between host bacteria and phages, reflecting their evolutionary adaptations (Bobay *et al.*, 2014). In the case of *S. aureus* in AD, it has been demonstrated that specific *S. aureus* phages can significantly reduce the abundance of *S. aureus*, thereby mitigating disease severity (Shimamori *et al.*, 2020)

Our research identified significant differences in the prophage species and gene compositions of *S. aureus* strains between AD patients and those from HE, with HE strains being more influenced by their prophages. This interaction suggests a co-evolutionary relationship that equips HE strains with a broader range of responses to environmental challenges, enhancing their adaptability. This enhanced adaptability is crucial for survival in varied environments or for resistance to a spectrum of antibiotics or other treatments. In contrast, AD-associated strains appear to be subject to selective pressures that favor certain *S. aureus* strains adapted to the unique conditions of AD, leading to a more restricted gene content. This variation in gene content, particularly noted in the marker genes—over 60% of which are associated with prophages—underscores the differences between the strains. Notably, the phage holin gene stands out as a key discriminative factor for differentiating strains, showing significant enrichment in HE strains.

Holins, which are encoded by bacteriophage genomes, are crucial for terminating the phage replication cycle by creating pores in the bacterial cytoplasmic membrane (Wang *et al.*, 2000;

Saier and Reddy, 2015). This action allows for the release of endolysins, enzymes that degrade the bacterial cell wall, leading to cell lysis and the subsequent release of new virions (Young and Bläsi, 1995). The prevalence of holins in HE-associated *S. aureus* strains might be explained by the "Killing the Winner" hypothesis, which posits a direct correlation between bacterial growth rates and phage activity, leading to a regulation of bacterial abundance (Winter *et al.*, 2010). The diminished presence of these genes in AD strains may reflect an evolutionary adaptation, where specific prophage-encoded virulence genes are retained, and others are lost, indicating a selective pressure within the AD environment. Alternatively, the over-presence of holins in HE strains could indicate a higher phage activity level, potentially reducing *S. aureus* growth rates and limiting their proliferation in HE compared to AD-affected skin. This phenomenon may explain why strains less affected by prophages are more prevalent in AD conditions. Further research is needed to validate these findings experimentally.

4.9 Functional implications of different prophages for *S. aureus* from AD and HE

Prophages, which are bacteriophage sequences integrated into bacterial genomes, play a crucial role in molding the physiological characteristics and introducing novel functions to bacteria through HGT (Arnold *et al.*, 2022; Borodovich *et al.*, 2022). Specifically, in *S. aureus*, prophages contribute significantly to the bacterium's arsenal by encoding various virulence factors and metabolic pathways, thereby influencing its adaptability and pathogenicity (Hatoum-Aslan, 2021). Our analysis reveals that prophages in HE strains exhibit a more varied gene content. However, a closer examination of high-quality prophage genomes and all prophage sequences, including remnants, uncovers a pronounced link between specific VFs and prophages associated with AD. This includes *clfA, nuc*, and multiple toxins (*sea, sec, sell, hlgA, lukD*), suggesting these VFs confer a selective advantage to AD strains. ClfA, a protein that binds fibrinogen, enhances bacterial attachment to host tissues, significantly contributing to the infection's pathogenesis by facilitating cellular clustering (Foster, 2009). Nuc, a thermostable enzyme, plays roles in evading immune responses or forming biofilms (Sultan *et al.*, 2019; Yu *et al.*, 2021). Meanwhile, enterotoxins and enterotoxin-like enzymes (*sea, sec,* and *sell*) are implicated in food poisoning incidents and elicit strong inflammatory reactions due to their superantigenic properties (Ortega *et al.*, 2010). HlgA, part of a binary toxin, causes extensive host cell lysis, leading to tissue damage (Dalla Serra *et al.*, 2005), and LukD, a pore-forming cytotoxin, targets leukocytes, undermining the host's immune defenses (Rojo *et al.*, 2014).

The dataset encompassing all prophage sequences reveals a broader array of VFs, notably serine proteases, enterotoxin-like substances, and exotoxins, which are predominantly found in prophages associated with AD. This indicates that prophage remnants play a significant role in enhancing virulence variability and possibly the adaptability of *S. aureus*. VFs, particularly serine proteases, are known to facilitate deeper bacterial infiltration, leading to more profound and challenging-to-treat inflammation (Nakatsuji *et al.*, 2016; Liu *et al.*, 2017). While the literature has not fully explored the specific functions of exotoxins (Dinges *et al.*, 2000), several Staphylococcal Enterotoxin Types (SETs) function as superantigens. These superantigens can activate a substantial fraction of the T-cell population, causing extensive cytokine release and inflammatory responses (Dinges *et al.*, 2000). The contribution of VFs within prophage remnants to the overall bacterial physiology, or whether these prophages can be induced, remains uncertain. Intriguingly, our analysis also identified that certain VFs (CHIPs, *ebp*, *lukFS-PV*, and *tsst*), previously associated with AD, were more prevalent in prophages from HE in both datasets. This inconsistency calls for further exploration to elucidate the underlying mechanisms.

The enhanced variety of VFs observed in the comprehensive prophage dataset underscores the pivotal role of HGT via prophage sequences in the evolution of *S. aureus*. This has significant repercussions for managing infections and developing treatment approaches for patients with AD. Additionally, our study discovered VFs typically located in the *S. aureus* chromosome or Pathogenicity Islands (SaPIs), such as *ebp, nuc, and hlb*, within high-quality prophage genomes. The presence of these VFs in prophages may indicate a wider dissemination of these genes or suggest HGT events deserving of further investigation. Notably, *ebp* and β-hemolysin have been identified in prophages (Naorem *et al.*, 2021). Research also suggests that phages can facilitate the movement of superantigen-encoding SaPIs carrying various VFs, like TSST-1 and enterotoxin B (Novick and Ram, 2017). This could amplify the exchange of VF-coding genes not only within and between strains but also between phages and SaPIs. Such findings highlight the intricate and dynamic nature of the *S. aureus* genome, suggesting that gene mobility and HGT may obscure the conventional demarcations of VF gene locations. Future research aims to unravel these complexities, enhancing our grasp of the genetic architecture and its impact on the pathogenicity of *S. aureus*.

5 Conclusions and outlook

Our research detailed in this thesis has successfully characterized the genomic and functional variability among *S. aureus* strains associated with AD compared to those from HE. Advanced genomic sequencing and innovative ML approaches have illuminated significant differences in genetic markers and functionalities that could potentially transform our understanding and treatment of AD. The genomic analysis revealed that AD-associated *S. aureus* strains exhibit narrowed genetic diversity and greater proportion in core genes compared to HE strains, pointing towards a selective pressure in the specific AD microenvironment that favors certain strains capable of thriving and exacerbating the disease condition. This trend towards adaptive evolution is reflected across both local and global scales, uncovering that the differences in gene content and function are not only preserved across different geographical locations but also exhibit certain biogeographical patterns.

Key functional categories like Replication and repair, and Transporters, play a central role in diversifying these strains. The heightened presence of transposases on both global and local levels suggests them as an important driver for *S. aureus* diversification. Distinctly, local strains showed enrichment in metal-related genes, whereas global strains were more associated with antimicrobial resistance. This highlights the nuanced roles of metal transport and antibiotic resistance in the scale-dependent diversification process. A critical finding from this research is the role of HGT in the diversification of *S. aureus* strains. The detection of HGT event of the lantibiotic operon further highlights a dynamic genetic interchange between human-associated and environmental bacteria, enhancing our understanding of bacterial evolution in different ecosystems. Despite a reduced gene content, AD strains demonstrated broader functional variability, indicating that HGT might significantly contribute to the survival and virulence of these strains in the unique environment of AD-affected skin. These findings enrich the current understanding of *S. aureus* evolution and its genetic underpinnings in AD, offering critical insights for clinical perspectives. Future studies should delve deeper into the precise factors and functions influencing the observed genetic diversity.

Our investigation unveils the profound influence of prophages in shaping the genomic landscape and virulence profiles of strains associated with AD. The identified marker genes, especially those associated with prophages, exhibit remarkable discriminative power in distinguishing AD-related strains from HE-related isolates. Furthermore, the enrichment of VFs within AD-associated prophages illustrates the significant impact of these prophages on the heightened pathogenicity of *S. aureus* in AD conditions. Conversely, the unique genes and functions in HE-associated prophages shed light on their adaptive evolution, possibly driven

by diverse environmental challenges. These findings not only elucidate the co-evolutionary dynamics between *S. aureus* and its prophages but also pave the way for targeted therapeutic interventions for AD.

Given the insights gained, several promising avenues for future research and application emerge. Phage therapy could target prophages to disrupt genetic elements in AD-associated strains, offering a novel treatment avenue that circumvents traditional ABR mechanisms. The identification of specific genetic markers for AD-associated *S. aureus* strains paves the way for precision medicine approaches, where treatments can be tailored based on the genetic profile of the infecting strain. Continuous monitoring of *S. aureus* strains through genomic sequencing could lead to improved diagnostic tools that effectively detect and differentiate AD-related strains, enabling timely and precise interventions.

Combining insights from genomics, bioinformatics, dermatology, and immunology will be crucial in developing comprehensive strategies that address not just the microbial component but also the host responses in AD. Increasing awareness about the microbial factors in AD and promoting hygiene and care practices among the population can help mitigate the spread and impact of virulent *S. aureus* strains. Expanding the research to include functional studies on the proteins encoded by the identified marker genes could elucidate their specific roles in the disease process, providing targets for therapeutic intervention. This thesis sets a foundation for future research that could lead to significant improvements in the management and treatment of AD, using cutting-edge genomic and bioinformatic tools to explore how microbial diversity and functionality affect disease outcomes.

References

Abeck, D. and Mempel, M. (1998) *Staphylococcus aureus* colonization in atopic dermatitis and its therapeutic implications. *Br J Dermatology, Suppl* **139**: 13–16.

Agrawal, R. and Woodfolk, J.A. (2014) Skin barrier defects in atopic dermatitis topical collection on allergic skin diseases. *Curr Allergy Asthma Rep* **14**:.

Akhter, S., Aziz, R.K., and Edwards, R.A. (2012) PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies. *Nucleic Acids Res* **40**: e126.

Al-Afif, K.A.M., Buraik, M.A., Buddenkotte, J., Mounir, M., Gerber, R., Ahmed, H.M., et al. (2019) Understanding the Burden of Atopic Dermatitis in Africa and the Middle East. *Dermatol Ther (Heidelb)* **9**: 223–241.

Al-Shobaili, H.A., Ahmed, A.A., Alnomair, N., Zeiad,), Alobead, A., and Rasheed, Z. (2016) Molecular Genetic of Atopic dermatitis: An Update. *Int J Health Sci (Qassim)* **10**: 96.

Alberts, B. (2003) DNA replication and recombination. *Nature* **421**: 431–435.

Ali, S.M. and Yosipovitch, G. (2013) Skin pH: From basic science to basic skin care. *Acta Derm Venereol* **93**: 261–267.

Alibayov, B., Baba-Moussa, L., Sina, H., Zdeňková, K., and Demnerová, K. (2014) *Staphylococcus aureus* mobile genetic elements. *Mol Biol Rep* **41**: 5005–5018.

Alsterholm, M., Strömbeck, L., Ljung, A., Karami, N., Widjestam, J., Gillstedt, M., et al. (2017) Variation in *Staphylococcus aureus* colonization in relation to disease severity in adults with atopic dermatitis during a fivemonth follow-up. *Acta Derm Venereol* **97**: 802–807.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Altunbulakli, C., Reiger, M., Neumann, A.U., Garzorz-Stark, N., Fleming, M., Huelpuesch, C., et al. (2018) Relations between epidermal barrier dysregulation and *Staphylococcus* species–dominated microbiome dysbiosis in patients with atopic dermatitis. *J Allergy Clin Immunol* **142**: 1643-1647.e12.

Álvarez-Carretero, S., Kapli, P., and Yang, Z. (2023) Beginner's Guide on the Use of PAML to Detect Positive Selection. *Mol Biol Evol* **40**:.

Apfelbacher, C.J., Diepgen, T.L., and Schmitt, J. (2011) Determinants of eczema: Population-based cross-sectional study in Germany. *Allergy Eur J Allergy Clin Immunol* **66**: 206–213.

Appiah, M.M., Haft, M.A., Kleinman, E., Laborada, J., Lee, S., Loop, L., et al. (2022) Atopic dermatitis Review of comorbidities and therapeutics. *Ann Allergy, Asthma Immunol* **129**: 142–149.

Arents, B.W.M., van Zuuren, E.J., Vermeulen, S., Schoones, J.W., and Fedorowicz, Z. (2022) Global Guidelines in Dermatology Mapping Project (GUIDEMAP), a systematic review of atopic dermatitis clinical practice guidelines: are they clear, unbiased, trustworthy and evidence based (CUTE)?*. *Br J Dermatol* **186**: 792–802.

Arnold, B.J., Huang, I.T., and Hanage, W.P. (2022) Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* **20**: 206–218.

Asnicar, F., Thomas, A.M., Passerini, A., Waldron, L., and Segata, N. (2024) Machine learning for microbiologists. *Nat Rev Microbiol* **22**: 191–205.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Bansal, V. and Boucher, C. (2019) Sequencing Technologies and Analyses: Where Have We Been

and Where Are We Going? *iScience* **18**: 37–41.

Barnetson, R.S.C. and Rogers, M. (2002) Childhood atopic eczema. *Br Med J* **324**: 1376–1379.

Beck, D. and Foster, J.A. (2014) Machine Learning Techniques Accurately Classify Microbial Communities by Bacterial Vaginosis Characteristics. *PLoS One* **9**: e87830.

Belay, N. and Rasooly, A. (2002) *Staphylococcus aureus* Growth and Enterotoxin A Production in an Anaerobic Environment. *J Food Prot* **65**: 199–204.

De Benedetto, A., Agnihothri, R., McGirt, L.Y., Bankova, L.G., and Beck, L.A. (2009) Atopic dermatitis: a disease caused by innate immune defects? *J Invest Dermatol* **129**: 14–30.

Berglund, C. and Söderquist, B. (2008) The origin of a methicillin-resistant *Staphylococcus aureus* isolate at a neonatal ward in Sweden - Possible horizontal transfer of a staphylococcal cassette chromosome mec between methicillin-resistant *Staphylococcus haemolyticus* and *Staphylococcus aureus*. *Clin Microbiol Infect* **14**: 1048–1056.

Berube, B.J. and Wardenburg, J.B. (2013) *Staphylococcus aureus* α-toxin: Nearly a century of intrigue. *Toxins (Basel)* **5**: 1140–1166.

Bessa, G.R., Machado, D.C., Weber, M.B., D'Azevedo, P.A., Quinto, V.P., Lipnharski, C., and Bonamigo, R.R. (2016) *Staphylococcus aureus* resistance to topical antimicrobials in atopic dermatitis. *An Bras Dermatol* **91**: 604–610.

Bieber, T., Paller, A.S., Kabashima, K., Feely, M., Rueda, M.J., Ross Terres, J.A., and Wollenberg, A. (2022) Atopic dermatitis: pathomechanisms and lessons learned from novel systemic therapeutic options. *J Eur Acad Dermatology Venereol* **36**: 1432–1449.

Bikard, D. and Marraffini, L.A. (2012) Innate and adaptive immunity in bacteria: Mechanisms of programmed genetic variation to fight bacteriophages. *Curr Opin Immunol* **24**: 15–20.

Bobay, L.M., Touchon, M., and Rocha, E.P.C. (2014) Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A* **111**: 12127–12132.

Borodovich, T., Shkoporov, A.N., Ross, R.P., and Hill, C. (2022) Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. *Gastroenterol Rep* **10**: 012.

Breiman, L. (2001) Random forests. *Mach Learn* **45**: 5–32.

Breitbart, M. and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* **13**: 278–284.

Breuer, K., Häussler, S., Kapp, A., and Werfel, T. (2002) *Staphylococcus aureus*: Colonizing features and influence of an antibacterial treatment in adults with atopic dermatitis. *Br J Dermatol* **147**: 55–61.

Broer, S., Ji, G., Broer, A., and Silver, S. (1993) Arsenic efflux governed by the arsenic resistance determinant of *Staphylococcus aureus* plasmid pI258. *J Bacteriol* **175**: 3480–3485.

Brown, S.J., Asai, Y., Cordell, H.J., Campbell, L.E., Zhao, Y., Liao, H., et al. (2011) Loss-of-function variants in the filaggrin gene are a significant risk factor for peanut allergy. *J Allergy Clin Immunol* **127**: 661–667.

Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004) Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol Mol Biol Rev* **68**: 560–602.

Burian, M., Rautenberg, M., Kohler, T., Fritz, M., Krismer, B., Unger, C., et al. (2010) Temporal expression of adhesion factors and activity of global regulators during establishment of *staphylococcus aureus* nasal colonization. *J Infect Dis* **201**: 1414–1421.

Byrd, A.L., Belkaid, Y., and Segre, J.A. (2018) The human skin microbiome. *Nat Rev Microbiol* **16**: 143–155.

Byrd, A.L., Deming, C., Cassidy, S.K.B., Harrison, O.J., Ng, W.I., Conlan, S., et al. (2017) *Staphylococcus aureus* and *Staphylococcus epidermidis* strain diversity underlying pediatric atopic dermatitis. *Sci Transl Med* **9**: eaal465.

Cafini, F., Thi Le Thuy, N., Román, F., Prieto, J., Dubrac, S., Msadek, T., and Morikawa, K. (2017) Methodology for the study of horizontal gene transfer in *Staphylococcus aureus*. *J Vis Exp* **2017**: 55087.

Caldwell, R., Zhou, W., and Oh, J. (2022) Strains to go: interactions of the skin microbiome beyond its species. *Curr Opin Microbiol* **70**: 102222.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* **10**: 1–9.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Carter, G.P., Schultz, M.B., Baines, S.L., Da Silva, A.G., Heffernan, H., Tiong, A., et al. (2018) Topical antibiotic use coselects for the carriage of mobile genetic elements conferring resistance to unrelated antimicrobials in *staphylococcus aureus*. *Antimicrob Agents Chemother* **62**:.

Centers for Disease Control and Prevention (2017) Outpatient Antibiotic Prescriptions — United States, 2017 | Community | Antibiotic Use | CDC. *22102019*.

Chambers, H.F. and DeLeo, F.R. (2009) Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat Rev Microbiol* **7**: 629–641.

Chen, J., Carpena, N., Quiles-Puchalt, N., Ram, G., Novick, R.P., and Penadés, J.R. (2015) Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by cos phages. *ISME J* **9**: 1260–1263.

Chen, J. and Novick, R.P. (2009) Phage-mediated intergeneric transfer of toxin genes. *Science (80-)* **323**: 139–141.

Chevallereau, A., Pons, B.J., van Houte, S., and Westra, E.R. (2022) Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol* **20**: 49–62.

Chiu, L.S., Ho, M.S.L., Hsu, L.Y., and Tang, M.B.Y. (2009) Prevalence and molecular characteristics of *Staphylococcus aureus* isolates colonizing patients with atopic dermatitis and their close contacts in Singapore. *Br J Dermatol* **160**: 965–971.

Chmielowiec-Korzeniowska, A., Tymczyna, L., Wlazło, Ł., Nowakowicz-Dębek, B., and Trawińska, B. (2020) *Staphylococcus aureus* carriage state in healthy adult population and phenotypic and genotypic properties of isolated strains. *Postep Dermatologii i Alergol* **37**: 184–189.

Cho, S.H., Strickland, I., Boguniewicz, M., and Leung, D.Y.M. (2001) Fibronectin and fibrinogen contribute to the enhanced binding of *Staphylococcus aureus* to atopic skin. *J Allergy Clin Immunol* **108**: 269–274.

Cho, S.H., Strickland, I., Tomkinson, A., Fehringer, A.P., Gelfand, E.W., and Leung, D.Y.M. (2001) Preferential binding of *Staphylococcus aureus* to skin sites of Th2-mediated inflammation in a murine model. *J Invest Dermatol* **116**: 658–663.

Clarke, S.R., Brummell, K.J., Horsburgh, M.J., McDowell, P.W., Syed Mohamad, S.A., Stapleton, M.R., et al. (2006) Identification of in vivo-expressed antigens of *Staphylococcus aureus* and their use in vaccinations for protection against nasal carriage. *J Infect Dis* **193**: 1098–1108.

Clarke, S.R., Mohamed, R., Bian, L., Routh, A.F., Kokai-Kun, J.F., Mond, J.J., et al. (2007) The *Staphylococcus aureus* Surface Protein IsdA Mediates Resistance to Innate Defenses of Human Skin. *Cell Host Microbe* **1**: 199–212.

Clausen, M.L., Edslev, S.M., Andersen, P.S., Clemmensen, K., Krogfelt, K.A., and Agner, T. (2017) *Staphylococcus aureus* colonization in atopic eczema and its association with filaggrin gene mutations. *Br J Dermatol* **177**: 1394–1400.

Clausen, M.L., Edslev, S.M., Nørreslet, L.B., Sørensen, J.A., Andersen, P.S., and Agner, T. (2019) Temporal variation of *Staphylococcus aureus* clonal complexes in atopic dermatitis: a follow-up study. *Br J Dermatol* **180**: 181–186.

Clokie, M.R.J., Millard, A.D., Letarov, A. V., and Heaphy, S. (2011) Phages in nature. *Bacteriophage* **1**: 31–45.

Coleman, D.C., Sullivan, D.J., Russell, R.J., Arbuthnott, J.P., Carey, B.F., and Pomeroy, H.M. (1989) *Staphylococcus aureus* bacteriophages mediating the simulataneous lysogenic conversion of β-lysin, staphylokinase and enterotoxin A: Molecular mechanism of triple conversion. *J Gen Microbiol* **135**: 1679–1697.

Comeau, A.M., Douglas, G.M., and Langille, M.G.I. (2017) Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. *mSystems* **2**:.

Conte, A.L., Brunetti, F., Marazzato, M., Longhi, C., Maurizi, L., Raponi, G., et al. (2023) Atopic dermatitis-derived *Staphylococcus aureus* strains: what makes them special in the interplay with the host. *Front Cell Infect Microbiol* **13**: 1194254.

Conwill, A., Kuan, A.C., Damerla, R., Poret, A.J., Baker, J.S., Tripp, A.D., et al. (2022) Anatomy promotes neutral coexistence of strains in the human skin microbiome. *Cell Host Microbe* **30**: 171-182.e7.

Corrigan, R.M., Miajlovic, H., and Foster, T.J. (2009) Surface proteins that promote adherence of *Staphylococcus aureus* to human desquamated nasal epithelial cells. *BMC Microbiol* **9**: 1–10.

Courvalin, P. (2006) Vancomycin Resistance in Gram-Positive Cocci. *Clin Infect Dis* **42**: S25–S34.

Cuecas, A., Kanoksilapatham, W., and Gonzalez, J.M. (2017) Evidence of horizontal gene transfer by transposase gene analyses in Fervidobacterium species. *PLoS One* **12**: e0173961.

Czarnowicki, T., Krueger, J.G., and Guttman-Yassky, E. (2017) Novel concepts of prevention and treatment of atopic dermatitis through barrier and immune manipulations with implications for the atopic march. *J Allergy Clin Immunol* **139**: 1723–1734.

Dalla Serra, M., Coraiola, M., Viero, G., Comai, M., Potrich, C., Ferreras, M., et al. (2005) *Staphylococcus aureus* bicomponent gamma-hemolysins, HlgA, HlgB, and HlgC, can form mixed pores containing all components. *J Chem Inf Model* **45**: 1539–1545.

Davis, D.M.R., Drucker, A.M., Alikhan, A., Bercovitch, L., Cohen, D.E., Darr, J.M., et al. (2022) American Academy of Dermatology Guidelines: Awareness of comorbidities associated with atopic dermatitis in adults. *J Am Acad Dermatol* **86**: 1335-1336.e18.

Deckers, I.A.G., McLean, S., Linssen, S., Mommers, M., van Schayck, C.P., and Sheikh, A. (2012) Investigating international time trends in the incidence and prevalence of atopic eczema 1990-2010: A systematic review of epidemiological studies. *PLoS One* **7**:.

Deurenberg, R.H., Bathoorn, E., Chlebowicz, M.A., Couto, N., Ferdous, M., García-Cobos, S., et al. (2017) Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol* **243**: 16–24.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018) The Third Revolution in Sequencing Technology. *Trends Genet* **34**: 666–681.

Dinges, M.M., Orwin, P.M., and Schlievert, P.M. (2000) Exotoxins of *Staphylococcus aureus*. *Clin Microbiol Rev* **13**: 16–34.

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**: 927–930.

Domenico, E.G. Di, Cavallo, I., Capitanio, B., Ascenzioni, F., Pimpinelli, F., Morrone, A., and Ensoli, F. (2019) *Staphylococcus aureus* and the cutaneous microbiota biofilms in the pathogenesis of atopic dermatitis. *Microorganisms* **7**: 301.

Durr, H.A. and Leipzig, N.D. (2023) Advancements in bacteriophage therapies and delivery for bacterial infection. *Mater Adv* **4**: 1249–1257.

Dwyer, D.J., Belenky, P.A., Yang, J.H., Cody MacDonald, I., Martell, J.D., Takahashi, N., et al. (2014) Antibiotics induce redox-related physi. *Proc Natl Acad Sci U S A* **111**: E2100–E2109.

Edslev, S.M., Clausen, M.L., Agner, T., Stegger, M., and Andersen, P.S. (2018) Genomic analysis reveals different mechanisms of fusidic acid resistance in *Staphylococcus aureus* from Danish atopic dermatitis patients. *J Antimicrob Chemother* **73**: 856–861.

Elmose, C. and Thomsen, S.F. (2015) Twin Studies of Atopic Dermatitis: Interpretations and Applications in the Filaggrin Era. *J Allergy* **2015**: 1–7.

Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., et al. (2017) Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* **49**: 1752–1757.

Fleury, O.M., McAleer, M.A., Feuillie, C., Formosa-Dague, C., Sansevere, E., Bennett, D.E., et al. (2017) Clumping Factor B Promotes Adherence of *Staphylococcus aureus* to Corneocytes in Atopic Dermatitis. *Infect Immun* **85**: 12.

Flohr, C. and Mann, J. (2014) New insights into the epidemiology of childhood atopic dermatitis. *Allergy Eur J Allergy Clin Immunol* **69**: 3–16.

Foster, T.J. (2009) Colonization and infection of the human host by staphylococci: Adhesion, survival and immune evasion. *Vet Dermatol* **20**: 456–470.

Foster, T.J. (2005) Immune evasion by staphylococci. *Nat Rev Microbiol* **3**: 948–958.

Foster, T.J., Geoghegan, J.A., Ganesh, V.K., and Höök, M. (2014) Adhesion, invasion and evasion: The many functions of the surface proteins of *Staphylococcus aureus*. *Nat Rev Microbiol* **12**: 49–62.

Foti, C., Romita, P., Borghi, A., Angelini, G., Bonamonte, D., and Corazza, M. (2015) Contact dermatitis to topical acne drugs: A review of the literature. *Dermatol Ther* **28**: 323–329.

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

Fuhrman, J.A. and Noble, R.T. (1995) Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol Oceanogr* **40**: 1236–1242.

Geoghegan, J.A., Irvine, A.D., and Foster, T.J. (2018) *Staphylococcus aureus* and Atopic Dermatitis: A Complex and Evolving Relationship. *Trends Microbiol* **26**: 484–497.

Ghannam, R.B. and Techtmann, S.M. (2021) Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. *Comput Struct Biotechnol J* **19**: 1092–1107.

Ghensi, P., Manghi, P., Zolfo, M., Armanini, F., Pasolli, E., Bolzan, M., et al. (2020) Strong oral plaque microbiome signatures for dental implant diseases identified by strain-resolution metagenomics. *npj Biofilms Microbiomes* **6**: 1–12.

Glatz, M., Buchner, M., Bartenwerffer, W. Von, Schmid-Grendelmeier, P.S., Worm, M., Hedderich, J., and Fölster-Holst, R. (2015) Malassezia SPP.-Specific immunoglobulin E level is a marker for severity of atopic dermatitis in adults. *Acta Derm Venereol* **95**: 191–196.

Gnanamani, A., Hariharan, P., and Paul-Satyaseela, M. (2017) *Staphylococcus aureus*: Overview of Bacteriology, Clinical Diseases, Epidemiology, Antibiotic Resistance and Therapeutic Approach. In *Frontiers in Staphylococcus aureus*. InTech.

Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., et al. (2009) Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *J Bacteriol* **191**: 3462–3468.

Goerke, C., Wirtz, C., Flückiger, U., and Wolz, C. (2006) Extensive phage dynamics in *Staphylococcus aureus* contributes to adaptation to the human host during infection. *Mol Microbiol* **61**: 1673–1685.

Graham, E.B., Paez-Espino, D., Brislawn, C., Hofmockel, K.S., Wu, R., Kyrpides, N.C., et al. (2019) Untapped viral diversity in global soil metagenomes. *bioRxiv* 583997.

Gummalla, V.S., Zhang, Y., Liao, Y. Te, and Wu, V.C.H. (2023) The Role of Temperate Phages in Bacterial Pathogenicity. *Microorganisms* **11**: 541.

Gupta, M.M. and Gupta, A. (2021) Survey of artificial intelligence approaches in the study of anthropogenic impacts on symbiotic organisms – a holistic view. *Symbiosis* **84**: 271–283.

Guzik, T.J., Bzowska, M., Kasprowicz, A., Czerniawska-Mysik, G., Wójcik, K., Szmyd, D., et al. (2005) Persistent skin colonization with *Staphylococcus aureus* in atopic dermatitis: Relationship to clinical and immunological parameters. *Clin Exp Allergy* **35**: 448–455.

Ha, K.P. and Edwards, A.M. (2021) DNA Repair in *Staphylococcus aureus*. *Microbiol Mol Biol Rev* **85**:.

Haaber, J., Penadés, J.R., and Ingmer, H. (2017) Transfer of Antibiotic Resistance in *Staphylococcus aureus*. *Trends Microbiol* **25**: 893–905.

Haag, A.F., Fitzgerald, J.R., and Penadés, J.R. (2019) *Staphylococcus aureus* in Animals. *Microbiol Spectr* **7**: 19.

De Haas, C.J.C., Veldkamp, K.E., Peschel, A., Weerkamp, F., Van Wamel, W.J.B., Heezius, E.C.J.M., et al. (2004) Chemotaxis Inhibitory Protein of *Staphylococcus aureus*, a Bacterial Antiinflammatory Agent. *J Exp Med* **199**: 687–695.

Harkins, C.P., Holden, M.T.G., and Irvine, A.D. (2019) Antimicrobial resistance in atopic dermatitis: Need for an urgent rethink. *Ann Allergy, Asthma Immunol* **122**: 236–240.

Harkins, C. P., McAleer, M.A., Bennett, D., McHugh, M., Fleury, O.M., Pettigrew, K.A., et al. (2018) The widespread use of topical antimicrobials enriches for resistance in *Staphylococcus aureus* isolated from patients with atopic dermatitis. *Br J Dermatol* **179**: 951–958.

Harkins, Catriona P., Pettigrew, K.A., Oravcová, K., Gardner, J., Hearn, R.M.R., Rice, D., et al. (2018) The Microevolution and Epidemiology of *Staphylococcus aureus* Colonization during Atopic Eczema Disease Flare. *J Invest Dermatol* **138**: 336–343.

Harris, T.O., Grossman, D., Kappler, J.W., Marrack, P., Rich, R.R., and Betley, M.J. (1993) Lack of complete correlation between emetic and T-cell-stimulatory activities of staphylococcal enterotoxins. *Infect Immun* **61**: 3175–3183.

Hatoum-Aslan, A. (2021) The phages of staphylococci: critical catalysts in health and disease. *Trends Microbiol* **29**: 1117–1129.

He, Y., Xie, Y., and Reed, S. (2014) Pulsed-Field Gel Electrophoresis Typing of *Staphylococcus aureus* Isolates. *Methods Mol Biol* **1085**: 103–111.

Heberle, H., Meirelles, V.G., da Silva, F.R., Telles, G.P., and Minghim, R. (2015) InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**: 1–7.

Hennekinne, J.-A., De Buyser, M.-L., and Dragacci, S. (2012) *Staphylococcus aureus* and its food poisoning toxins: characterization and outbreak investigation. *FEMS Microbiol Rev* **36**: 815–836.

Hernández Medina, R., Kutuzova, S., Nielsen, K.N., Johansen, J., Hansen, L.H., Nielsen, M., and Rasmussen, S. (2022) Machine learning and deep learning applications in microbiome research. *ISME Commun* **2**: 1–7.

Higaki, S., Morohashi, M., Yamagishi, T., and Hasegawa, Y. (1999) Comparative study of staphylococci from the skin of atopic dermatitis patients and from healthy subjects. *Int J Dermatol* **38**: 265–269.

Hirano, Y., Shinmoto, K., Okada, Y., Suga, K., Bombard, J., Murahata, S., et al. (2021) Machine Learning Approach to Predict Positive Screening of Methicillin-Resistant *Staphylococcus aureus* During Mechanical Ventilation Using Synthetic Dataset From MIMIC-IV Database. *Front Med* **8**: 2222.

Hirasawa, Y., Takai, T., Nakamura, T., Mitsuishi, K., Gunawan, H., Suto, H., et al. (2010) *Staphylococcus aureus* extracellular protease causes epidermal barrier dysfunction. *J Invest Dermatol* **130**: 614–617.

Holden, M.T.G., Feil, E.J., Lindsay, J.A., Peacock, S.J., Day, N.P.J., Enright, M.C., et al. (2004) Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A* **101**: 9786–9791.

Holden, M.T.G., Lindsay, J.A., Corton, C., Quail, M.A., Cockfield, J.D., Pathak, S., et al. (2010) Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). *J Bacteriol* **192**: 888–892.

House, W. (2022) Global Report on Atopic Dermatitis 2022 atopicdermatitisatlas.org International League of Dermatological Societies (ILDS) International League of Dermatological Societies.

Howard-Varona, C., Lindback, M.M., Bastien, G.E., Solonenko, N., Zayed, A.A., Jang, H. Bin, et al. (2020) Phage-specific metabolic reprogramming of virocells. *ISME J* **14**: 881–895.

Howden, B.P., Giulieri, S.G., Wong Fok Lung, T., Baines, S.L., Sharkey, L.K., Lee, J.Y.H., et al. (2023) *Staphylococcus aureus* host interactions and adaptation. *Nat Rev Microbiol* **21**: 380–395.

Howell, M.D., Kim, B.E., Gao, P., Grant, A. V., Boguniewicz, M., DeBenedetto, A., et al. (2009) Cytokine modulation of atopic dermatitis filaggrin skin expression. *J Allergy Clin Immunol* **124**: R7–R12.

Hoyles, L., McCartney, A.L., Neve, H., Gibson, G.R., Sanderson, J.D., Heller, K.J., and van Sinderen, D. (2014) Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res Microbiol* **165**: 803–812.

Hülpüsch, C., Tremmel, K., Hammel, G., Bhattacharyya, M., de Tomassi, A., Nussbaumer, T., et al. (2020) Skin pH–dependent *Staphylococcus aureus* abundance as predictor for increasing atopic dermatitis severity. *Allergy Eur J Allergy Clin Immunol* **75**: 2888–2898.

Hunt, M., Silva, N. De, Otto, T.D., Parkhill, J., Keane, J.A., and Harris, S.R. (2015) Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol* **16**:.

Hwang, J., Thompson, A., Jaros, J., Blackcloud, P., Hsiao, J., and Shi, V.Y. (2021) Updated understanding of *Staphylococcus aureus* in atopic dermatitis: From virulence factors to commensals and clonal complexes. *Exp Dermatol* **30**: 1532–1545.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Irvine, A.D., McLean, W.H.I., and Leung, D.Y.M. (2011) Filaggrin Mutations Associated with Skin and Allergic Diseases. *N Engl J Med* **365**: 1315–1327.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C., and Hiramatsu, K. (2001) Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* **45**: 1323–1336.

Iwamoto, K., Moriwaki, M., Miyake, R., and Hide, M. (2019) *Staphylococcus aureus* in atopic dermatitis: Strain-specific cell wall proteins and skin immunity. *Allergol Int* **68**: 309–315.

Iwamoto, K., Moriwaki, M., Niitsu, Y., Saino, M., Takahagi, S., Hisatsune, J., et al. (2017) *Staphylococcus aureus* from atopic dermatitis skin alters cytokine production triggered by monocyte-derived Langerhans cell. *J Dermatol Sci* **88**: 271–279.

Jain, R. (2003) Horizontal Gene Transfer Accelerates Genome Innovation and Evolution. *Mol Biol Evol* **20**: 1598–1602.

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., et al. (2019) Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing

networks. *Nat Biotechnol* **37**: 632–639.

Jiang, Z., Li, J., Kong, N., Kim, J.H., Kim, B.S., Lee, M.J., et al. (2022) Accurate diagnosis of atopic dermatitis by combining transcriptome and microbiota data with supervised machine learning. *Sci Rep* **12**: 1–13.

K. Patel, H. Wyatt, E. M. Kubiak, S, G. (2001) *Staphylococcus aureus* Colonization of Children with Atopic Eczema and Their Parents. *Acta Derm Venereol* **81**: 366–367.

Kadariya, J., Smith, T.C., and Thapaliya, D. (2014) *Staphylococcus aureus* and Staphylococcal Food-Borne Disease: An Ongoing Challenge in Public Health. *Biomed Res Int* **2014**: 1–9.

Kahánková, J., Pantůček, R., Goerke, C., Růžičková, V., Holochová, P., and Doškař, J. (2010) Multilocus PCR typing strategy for differentiation of *Staphylococcus aureus* siphoviruses reflecting their modular genome structure. *Environ Microbiol* **12**: 2527–2538.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**: 726–731.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

Kelleher, M., Dunn-Galvin, A., Hourihane, J.O.B., Murray, D., Campbell, L.E., McLean, W.H.I., and Irvine, A.D. (2015) Skin barrier dysfunction measured by transepidermal water loss at 2 days and 2 months predates and predicts atopic dermatitis at 1 year. *J Allergy Clin Immunol* **135**: 930-935.e1.

Key, F.M., Khadka, V.D., Romo-González, C., Blake, K.J., Deng, L., Lynn, T.C., et al. (2023) On-person adaptive evolution of *Staphylococcus aureus* during treatment for atopic dermatitis. *Cell Host Microbe* **31**: 593-603.e7.

Kezic, S., O'Regan, G.M., Yau, N., Sandilands, A., Chen, H., Campbell, L.E., et al. (2011) Levels of filaggrin degradation products are influenced by both filaggrin genotype and atopic dermatitis severity. *Allergy Eur J Allergy Clin Immunol* **66**: 934–940.

Kieft, K., Zhou, Z., and Anantharaman, K. (2020) VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**: 1–23.

Kim, B.S., Kim, J.Y., Lim, H.J., Lee, W.J., Lee, S.J., Kim, J.M., et al. (2011) Colonizing features of *Staphylococcus aureus* in early childhood atopic dermatitis and in mothers: A cross-sectional comparative study done at four kindergartens in Daegu, South Korea. *Ann Allergy, Asthma Immunol* **106**: 323–329.

Kim, D.W., Park, J.Y., Park, K.D., Kim, T.H., Lee, W.J., Lee, S.J., and Kim, J. (2009) Are there predominant strains and toxins of *Staphylococcus aureus* in atopic dermatitis patients? Genotypic characterization and toxin determination of *S. aureus* isolated in adolescent and adult patients with atopic dermatitis. *J Dermatol* **36**: 75–81.

Kim, J., Kim, B.E., Ahn, K., and Leung, D.Y.M. (2019) Interactions between atopic dermatitis and *Staphylococcus aureus* infection: Clinical implications. *Allergy, Asthma Immunol Res* **11**: 593–603.

Kim, J.P., Chao, L.X., Simpson, E.L., and Silverberg, J.I. (2016) Persistence of atopic dermatitis (AD): A systematic review and meta-analysis. *J Am Acad Dermatol* **75**: 681-687.e11.

Knox, J., Uhlemann, A.C., and Lowy, F.D. (2015) *Staphylococcus aureus* infections: Transmission within households and the community. *Trends Microbiol* **23**: 437–444.

Kobayashi, T., Glatz, M., Horiuchi, K., Kawasaki, H., Akiyama, H., Kaplan, D.H., et al. (2015) Dysbiosis and *Staphyloccus aureus* Colonization Drives Inflammation in Atopic Dermatitis. *Immunity* **42**: 756–766.

Kohanski, M.A., Dwyer, D.J., Hayete, B., Lawrence, C.A., and Collins, J.J. (2007) A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **130**: 797–810.

Kong, H.H., Oh, J., Deming, C., Conlan, S., Grice, E.A., Beatson, M.A., et al. (2012) Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res* **22**: 850–859.

Koosha, R.Z., Hosseini, H.M., Aghdam, E.M., Fooladi, A.A.I., and Tajandareh, S.G. (2016) Distribution of tsst-1 and mecA genes in *Staphylococcus aureus* isolated from clinical specimens. *Jundishapur J Microbiol* **9**: 29057.

Koskella, B. and Brockhurst, M.A. (2014) Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* **38**: 916–931.

Koskella, B. and Meaden, S. (2013) Understanding bacteriophage specificity in natural microbial communities. *Viruses* **5**: 806–823.

Kryazhimskiy, S. and Plotkin, J.B. (2008) The population genetics of dN/dS. *PLoS Genet* **4**: e1000304.

Kumar, P., Kretzschmar, B., Herold, S., Nau, R., Kreutzfeldt, M., Schütze, S., et al. (2015) Beneficial effect of chronic *Staphylococcus aureus* infection in a model of multiple sclerosis is mediated through the secretion of extracellular adherence protein. *J Neuroinflammation* **12**:.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., et al. (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* **357**: 1225–1240.

Kurt, K., Rasigade, J.P., Laurent, F., Goering, R. V., Žemličková, H., Machova, I., et al. (2013) Subpopulations of *Staphylococcus aureus* Clonal Complex 121 Are Associated with Distinct Clinical Entities. *PLoS One* **8**: e58155.

Kwong, J.C., Mccallum, N., Sintchenko, V., and Howden, B.P. (2015) Whole genome sequencing in clinical and public health microbiology. *Pathology* **47**: 199–210.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8**: 317–327.

Lan, R. and Reeves, P.R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol* **8**: 396–401.

Langan, S.M., Irvine, A.D., and Weidinger, S. (2020) Atopic dermatitis. *Lancet* **396**: 345–360.

Laouini, D., Kawamoto, S., Yalcindag, A., Bryce, P., Mizoguchi, E., Oettgen, H., and Geha, R.S. (2003) Epicutaneous sensitization with superantigen induces allergic skin inflammation. *J Allergy Clin Immunol* **112**: 981–987.

Laughter, M.R., Maymone, M.B.C., Mashayekhi, S., Arents, B.W.M., Karimkhani, C., Langan, S.M., et al. (2021) The global burden of atopic dermatitis: lessons from the Global Burden of Disease Study 1990–2017*. *Br J Dermatol* **184**: 304–309.

Lax, S.J., Harvey, J., Axon, E., Howells, L., Santer, M., Ridd, M.J., et al. (2022) Strategies for using topical corticosteroids in children and adults with eczema. *Cochrane Database Syst Rev* **2022**:.

Lee, A.S., De Lencastre, H., Garau, J., Kluytmans, J., Malhotra-Kumar, S., Peschel, A., and Harbarth, S. (2018) Methicillin-resistant *Staphylococcus aureus*. *Nat Rev Dis Prim* **4**: 1–23.

Leshem, T., Schnall, B.S., Azrad, M., Baum, M., Rokney, A., and Peretz, A. (2022) Incidence of biofilm formation among MRSA and MSSA clinical isolates from hospitalized patients in Israel. *J Appl Microbiol* **133**: 922–929.

Letunic, I. and Bork, P. (2021) Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**: W293–W296.

LEYDEN, J.J., MARPLES, R.R., and KLIGMAN, A.M. (1974) *Staphylococcus aureus* in the lesions of atopic dermatitis. *Br J Dermatol* **90**: 525–525.

Li, M., Du, X., Villaruz, A.E., Diep, B.A., Wang, D., Song, Y., et al. (2012) MRSA epidemic linked to a quickly spreading colonization and virulence determinant. *Nat Med* **18**: 816–819.

Lindsay, J.A. (2010) Genomic variation and evolution of *Staphylococcus aureus*. *Int J Med Microbiol* **300**: 98–103.

Lindsay, J.A. (2014) *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. *Int J Med Microbiol* **304**: 103–109.

Liu, B., Zheng, D., Zhou, S., Chen, L., and Yang, J. (2022) VFDB 2022: A general classification scheme for bacterial virulence factors. *Nucleic Acids Res* **50**: D912–D917.

Liu, H., Archer, N.K., Dillen, C.A., Wang, Y., Ashbaugh, A.G., Ortines, R. V., et al. (2017) *Staphylococcus aureus* Epicutaneous Exposure Drives Skin Inflammation via IL-36-Mediated T Cell Responses. *Cell Host Microbe* **22**: 653-666.e5.

Liu, N., Liu, D., Li, K., Hu, S., and He, Z. (2022) Pan-Genome Analysis of *Staphylococcus aureus* Reveals Key Factors Influencing Genomic Plasticity. *Microbiol Spectr* **10**:.

Liu, Y., Liu, X., Qu, Y., Wang, X., Li, L., and Zhao, X. (2012) Inhibitors of reactive oxygen species accumulation delay and/or reduce the lethality of several antistaphylococcal agents. *Antimicrob Agents Chemother* **56**: 6048–6050.

Lo, W.T., Wang, S.R., Tseng, M.H., Huang, C.F., Chen, S.J., and Wang, C.C. (2010) Comparative molecular analysis of meticillin-resistant *Staphylococcus aureus* isolates from children with atopic dermatitis and healthy subjects in Taiwan. *Br J Dermatol* **162**: 1110–1116.

Lobefaro, F., Gualdi, G., Di Nuzzo, S., and Amerio, P. (2022) Atopic Dermatitis: Clinical Aspects and Unmet Needs. *Biomedicines* **10**:.

Lomholt, H., Andersen, K.E., and Kilian, M. (2005) *Staphylococcus aureus* clonal dynamics and virulence factors in children with atopic dermatitis. *J Invest Dermatol* **125**: 977–982.

Lopez Carrera, Y.I., Al Hammadi, A., Huang, Y.H., Llamado, L.J., Mahgoub, E., and Tallman, A.M. (2019) Epidemiology, Diagnosis, and Treatment of Atopic Dermatitis in the Developing Countries of Asia, Africa, Latin America, and the Middle East: A Review. *Dermatol Ther (Heidelb)* **9**: 685–705.

Lowy, F.D. (1998) *Staphylococcus aureus* Infections. *N Engl J Med* **339**: 520–532.

Maiques, E., Úbeda, C., Campoy, S., Salvador, N., Lasa, Í., Novick, R.P., et al. (2006) β-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J Bacteriol* **188**: 2726–2729.

Malachowa, N. and Deleo, F.R. (2010) Mobile genetic elements of *Staphylococcus aureus*. *Cell Mol Life Sci* **67**: 3057.

Mandic-Mulec, I., Stefanic, P., and van Elsas, J.D. (2015) Ecology of Bacillaceae . *Microbiol Spectr* **3**:.

Marcos-Zambrano, L.J., Karaduzovic-Hadziabdic, K., Loncar Turukalo, T., Przymus, P., Trajkovik, V., Aasmets, O., et al. (2021) Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. *Front Microbiol* **12**: 313.

Maree, M., Thi Nguyen, L.T., Ohniwa, R.L., Higashide, M., Msadek, T., and Morikawa, K. (2022) Natural transformation allows transfer of SCCmec-mediated methicillin resistance in *Staphylococcus aureus* biofilms. *Nat Commun* **13**:.

Marsella, R., Olivry, T., and Carlotti, D.N. (2011) Current evidence of skin barrier dysfunction in human and canine atopic dermatitis. *Vet Dermatol* **22**: 239–248.

Martin, M.J., Estravís, M., García-Sánchez, A., Dávila, I., Isidoro-García, M., and Sanz, C. (2020) Genetics and epigenetics of atopic dermatitis: An updated systematic review. *Genes (Basel)* **11**: 442.

Martinez-Gutierrez, C.A., Aylward, F.O., and Lerat, E. (2019) Strong Purifying Selection Is Associated with Genome Streamlining in Epipelagic Marinimicrobia. *Genome Biol Evol* **11**: 2887–2894.

Mason, A., Foster, D., Bradley, P., Golubchik, T., Doumith, M., Gordon, N.C., et al. (2018) Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from *Staphylococcus aureus* whole-genome sequences. *J Clin Microbiol* **56**:.

Matsui, K., Nishikawa, A., Suto, H., Tsuboi, R., and Ogawa, H. (2000) Comparative Study of *Staphylococcus aureus* Isolated from Lesional and Non-Lesional Skin of Atopic Dermatitis Patients. *Microbiol Immunol* **44**: 945–947.

McAuliffe, O., Ross, R.P., and Hill, C. (2001) Lantibiotics: structure, biosynthesis and mode of action. *FEMS Microbiol Rev* **25**: 285–308.

McCarthy, A.J. and Lindsay, J.A. (2010) Genetic variation in *staphylococcus aureus* surface and immune evasion genes is lineage associated: Implications for vaccine design and host-pathogen interactions. *BMC Microbiol* **10**: 173.

McCarthy, A.J., Loeffler, A., Witney, A.A., Gould, K.A., Lloyd, D.H., and Lindsay, J.A. (2014) Extensive horizontal gene transfer during *staphylococcus aureus* co-colonization in vivo. *Genome Biol Evol* **6**: 2697–2708.

Meier-Kolthoff, J.P. and Göker, M. (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* **10**:.

Melles, D.C., Gorkink, R.F.J., Boelens, H.A.M., Snijders, S. V., Peeters, J.K., Moorhouse, M.J., et al. (2004) Natural population dynamics and expansion of pathogenic clones of *Staphylococcus aureus*. *J Clin Invest* **114**: 1732–1740.

Méric, G., Mageiros, L., Pensar, J., Laabei, M., Yahara, K., Pascoe, B., et al. (2018) Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nat Commun* **9**: 1–11.

Metzker, M.L. (2010) Sequencing technologies the next generation. *Nat Rev Genet* **11**: 31–46.

Michaelis, C. and Grohmann, E. (2023) Horizontal Gene Transfer of Antibiotic Resistance Genes in Biofilms. *Antibiotics* **12**:.

Monti, G., Tonetto, P., Mostert, M., and Oggero, R. (1996) *Staphylococcus aureus* skin colonization in infants with atopic dermatitis. *Dermatology* **193**: 83–87.

Moradigaravand, D., Palm, M., Farewell, A., Mustonen, V., Warringer, J., and Parts, L. (2018) Prediction of antibiotic resistance in *Escherichia coli* from large-scale pan-genome data. *PLoS Comput Biol* **14**: e1006258.

Moriwaki, M., Iwamoto, K., Niitsu, Y., Matsushima, A., Yanase, Y., Hisatsune, J., et al. (2019) *Staphylococcus aureus* from atopic dermatitis skin accumulates in the lysosomes of keratinocytes with induction of IL-1α secretion via TLR9. *Allergy Eur J Allergy Clin Immunol* **74**: 560–571.

Morse, M.L. (1959) Transduction By Staphylococcal Bacteriophage. *Proc Natl Acad Sci* **45**: 722–727.

Mu, Z. and Zhang, J. (2020) The Role of Genetics, the Environment, and Epigenetics in Atopic Dermatitis. *Adv Exp Med Biol* **1253**: 107–140.

Mulcahy, M.E., Geoghegan, J.A., Monk, I.R., O'Keeffe, K.M., Walsh, E.J., Foster, T.J., and McLoughlin, R.M. (2012) Nasal Colonisation by *Staphylococcus aureus* Depends upon Clumping Factor B Binding to the Squamous Epithelial Cell Envelope Protein Loricrin. *PLoS Pathog* **8**:.

Myles, I.A., Earland, N.J., Anderson, E.D., Moore, I.N., Kieh, M.D., Williams, K.W., et al. (2018) First-in-human topical microbiome transplantation with *Roseomonas mucosa* for atopic dermatitis. *JCI insight* **3**:.

Na, S.Y., Roh, J.Y., Kim, J.M., Tamang, M.D., and Lee, J.R. (2012) Analysis of colonization and genotyping of the exotoxins of *staphylococcus aureus* in patients with atopic dermatitis. *Ann Dermatol* **24**: 413–419.

Naik, S., Bouladoux, N., Wilhelm, C., Molloy, M.J., Salcedo, R., Kastenmuller, W., et al. (2012)

Compartmentalized control of skin immunity by resident commensals. *Science (80-)* **337**: 1115–1119.

Nakamura, Y., Oscherwitz, J., Cease, K.B., Chan, S.M., Muñoz-Planillo, R., Hasegawa, M., et al. (2013) *Staphylococcus* δ-toxin induces allergic skin disease by activating mast cells. *Nature* **503**: 397–401.

Nakamura, Y., Takahashi, H., Takaya, A., Inoue, Y., Katayama, Y., Kusuya, Y., et al. (2020) *Staphylococcus* Agr virulence is critical for epidermal colonization and associates with atopic dermatitis development. *Sci Transl Med* **12**: 4068.

Nakatsuji, T., Chen, T.H., Narala, S., Chun, K.A., Two, A.M., Yun, T., et al. (2017) Antimicrobials from human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis. *Sci Transl Med* **9**:.

Nakatsuji, T., Chen, T.H., Two, A.M., Chun, K.A., Narala, S., Geha, R.S., et al. (2016) *Staphylococcus aureus* Exploits Epidermal Barrier Defects in Atopic Dermatitis to Trigger Cytokine Expression. *J Invest Dermatol* **136**: 2192–2200.

Naorem, R.S., Goswami, G., Gyorgy, S., and Fekete, C. (2021) Comparative analysis of prophages carried by human and animal-associated *Staphylococcus aureus* strains spreading across the European regions. *Sci Reports 2021 111* **11**: 1–13.

Naureen, Z., Dautaj, A., Anpilogov, K., Camilleri, G., Dhuli, K., Tanzi, B., et al. (2020) Bacteriophages presence in nature and their role in the natural selection of bacterial populations. *Acta Biomed* **91**: 1–13.

Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N.C. (2020) CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol 2020 395* **39**: 578–585.

Nepal, R., Houtak, G., Shaghayegh, G., Bouras, G., Shearwin, K., Psaltis, A.J., et al. (2021) Prophages encoding human immune evasion cluster genes are enriched in *Staphylococcus aureus* isolated from chronic rhinosinusitis patients with nasal polyps. *Microb genomics* **7**: 726.

Nguyen, M.T., Hanzelmann, D., Härtner, T., Peschel, A., and Götz, F. (2015) Skinspecific unsaturated fatty acids boost the *Staphylococcus aureus* innate immune response. *Infect Immun* **84**: 205–215.

Novick, R.P. and Ram, G. (2017) Staphylococcal pathogenicity islands — movers and shakers in the genomic firmament. *Curr Opin Microbiol* **38**: 197–204.

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

O'Gara, J.P. (2017) Into the storm : Chasing the opportunistic pathogen *Staphylococcus aureus* from skin colonisation to life-threatening infections. *Environ Microbiol* **19**: 3823–3833.

Obata, S., Hisatsune, J., Kawasaki, H., Fukushima-Nomura, A., Ebihara, T., Arai, C., et al. (2023) Comprehensive Genomic Characterization of *Staphylococcus aureus* Isolated from Atopic Dermatitis Patients in Japan: Correlations with Disease Severity, Eruption Type, and Anatomical Site. *Microbiol Spectr* **11**: 16.

Ogonowska, P., Gilaberte, Y., Barańska-Rybak, W., and Nakonieczna, J. (2021) Colonization With *Staphylococcus aureus* in Atopic Dermatitis Patients: Attempts to Reveal the Unknown. *Front Microbiol* **11**: 3468.

Oh, J., Byrd, A.L., Deming, C., Conlan, S., Kong, H.H., Segre, J.A., et al. (2014) Biogeography and individuality shape function in the human skin metagenome. *Nature* **514**: 59–64.

Ono, H.K., Nishizawa, M., Yamamoto, Y., Hu, D.L., Nakane, A., Shinagawa, K., and Omoe, K. (2012) Submucosal mast cells in the gastrointestinal tract are a target of staphylococcal enterotoxin type A. *FEMS Immunol Med Microbiol* **64**: 392–402.

Ortega, E., Abriouel, H., Lucas, R., and Gálvez, A. (2010) Multiple Roles of *Staphylococcus aureus* Enterotoxins: Pathogenicity, Superantigenic Activity, and Correlation to Antibiotic Resistance. *Toxins (Basel)* **2**: 2117–2131.

Ou, C., Shang, D., Yang, J., Chen, B., Chang, J., Jin, F., and Shi, C. (2020) Prevalence of multidrug-resistant *Staphylococcus aureus* isolates with strong biofilm formation ability among animal-based food in Shanghai. *Food Control* **112**: 107106.

Oyserman, B.O., Moya, F., Lawson, C.E., Garcia, A.L., Vogt, M., Heffernen, M., et al. (2016) Ancestral genome reconstruction identifies the evolutionary basis for trait acquisition in polyphosphate accumulating bacteria. *ISME J* **10**: 2931–2945.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., et al. (2015) Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**: 3691–3693.

Paller, A.S., Kong, H.H., Seed, P., Naik, S., Scharschmidt, T.C., Gallo, R.L., et al. (2019) The microbiome in patients with atopic dermatitis. *J Allergy Clin Immunol* **143**: 26–35.

Park, H.Y., Kim, C.R., Huh, I.S., Jung, M.Y., Seo, E.Y., Park, J.H., et al. (2013) *Staphylococcus aureus* colonization in acute and chronic skin lesions of patients with atopic dermatitis. *Ann Dermatol* **25**: 410–416.

Park, J.M., Jo, J.H., Jin, H., Ko, H.C., Kim, M.B., Kim, J.M., et al. (2016) Change in antimicrobial susceptibility of skin-colonizing *Staphylococcus aureus* in Korean patients with atopic dermatitis during ten-year period. *Ann Dermatol* **28**: 470–478.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Parsons, R.J., Breitbart, M., Lomas, M.W., and Carlson, C.A. (2012) Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J* **6**: 273–284.

Pascolini, C., Sinagra, J., Pecetta, S., Bordignon, V., De Santis, A., Cilli, L., et al. (2011) Molecular and immunological characterization of *Staphylococcus aureus* in pediatric atopic dermatitis: Implications for prophylaxis and clinical management. *Clin Dev Immunol* **2011**:.

Paternoster, L., Standl, M., Waage, J., Baurecht, H., Hotze, M., Strachan, D.P., et al. (2015) Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat Genet* **47**: 1449–1456.

Peschel, A. and Otto, M. (2013) Phenol-soluble modulins and staphylococcal infection. *Nat Rev Microbiol* **11**: 667–673.

Piechota, M., Kot, B., Frankowska-Maciejewska, A., Grużewska, A., and Woźniak-Kosek, A. (2018) Biofilm Formation by Methicillin-Resistant and Methicillin-Sensitive *Staphylococcus aureus* Strains from Hospitalized Patients in Poland. *Biomed Res Int* **2018**: 1–7.

Pinchuk, I. V., Beswick, E.J., and Reyes, V.E. (2010) Staphylococcal Enterotoxins. *Toxins (Basel)* **2**: 2177–2197.

Pothmann, A., Illing, T., Wiegand, C., Hartmann, A.A., and Elsner, P. (2019) The Microbiome and Atopic Dermatitis: A Review. *Am J Clin Dermatol* **20**: 749–761.

Prestinaci, F., Pezzotti, P., and Pantosti, A. (2015) Antimicrobial resistance: A global multifaceted phenomenon. *Pathog Glob Health* **109**: 309–318.

Price, L.B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, P.S., et al. (2012) *Staphylococcus aureus* CC398: Host adaptation and emergence of methicillin resistance in livestock. *MBio* **3**: 1–6.

Qin, J., Li, Y., Cai, Z., Li, S., and Jianfeng Zhu, Fan Zhang, Suisha Liang, Wenwei Zhang, Yuanlin Guan, Dongqian Shen, Yangqing Peng, Dongya Zhang, Zhuye Jie, Wenxian Wu, Youwen Qin, Wenbin Xue, Junhua Li, Lingchuan Han, Donghui Lu, Peixian W, K.K.& J.W. (2012) A

metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**: 55–60.

Qin, L., McCausland, J.W., Cheung, G.Y.C., and Otto, M. (2016) PSM-Mec—A Virulence Determinant that Connects Transcriptional Regulation, Virulence, and Antibiotic Resistance in Staphylococci. *Front Microbiol* **7**:.

Quainoo, S., Coolen, J.P.M., van Hijum, S.A.F.T., Huynen, M.A., Melchers, W.J.G., van Schaik, W., and Wertheim, H.F.L. (2017) Whole-genome sequencing of bacterial pathogens: The future of nosocomial outbreak analysis. *Clin Microbiol Rev* **30**: 1015–1063.

Raivo Kolde (2019) Pretty Heatmaps. R package version 1.0.12. . *pheatmap https://CRANR-project.org/package=pheatmap*.

Rajka, G. (1974) Surface lipid estimation on the back of the hands in atopic dermatitis. *Arch Dermatol Forsch* **251**: 43–48.

Ramírez-Marín, H.A. and Silverberg, J.I. (2022) Differences between pediatric and adult atopic dermatitis. *Pediatr Dermatol* **39**: 345–353.

Rankin, D.J., Rocha, E.P.C., and Brown, S.P. (2011) What traits are carried on mobile genetic elements, and why. *Heredity (Edinb)* **106**: 1–10.

Rauer, L., Reiger, M., Bhattacharyya, M., Brunner, P.M., Krueger, J.G., Guttman-Yassky, E., et al. (2023) Skin microbiome and its association with host cofactors in determining atopic dermatitis severity. *J Eur Acad Dermatology Venereol* **37**: 772–782.

Rayner, C. and Munckhof, W.J. (2005) Antibiotics currently used in the treatment of infections caused by *Staphylococcus aureus*. *Intern Med J* **35**:.

Reynolds, M., Gorelick, J., and Bruno, M. (2020) Atopic dermatitis: A review of current diagnostic criteria and a proposed update to management. *J Drugs Dermatology* **19**: 244–248.

Rhoads, A. and Au, K.F. (2015) PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13**: 278–289.

Rhodes, N.J., Rohani, R., Yarnold, P.R., Pawlowski, A.E., Malczynski, M., Qi, C., et al. (2023) Machine Learning To Stratify Methicillin-Resistant *Staphylococcus aureus* Risk among Hospitalized Patients with Community-Acquired Pneumonia. *Antimicrob Agents Chemother* **67**:.

Rice, N.E., Patel, B.D., Lang, I.A., Kumari, M., Frayling, T.M., Murray, A., and Melzer, D. (2008) Filaggrin gene mutations are associated with asthma and eczema in later life. *J Allergy Clin Immunol* **122**: 834–836.

Richardson, E.J., Bacigalupe, R., Harrison, E.M., Weinert, L.A., Lycett, S., Vrieling, M., et al. (2018) Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat Ecol Evol* **2**: 1468–1478.

Riethmuller, C., McAleer, M.A., Koppes, S.A., Abdayem, R., Franz, J., Haftek, M., et al. (2015) Filaggrin breakdown products determine corneocyte conformation in patients with atopic dermatitis. *J Allergy Clin Immunol* **136**: 1573-1580.e2.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**: 77.

Rocha, E.P.C., Smith, J.M., Hurst, L.D., Holden, M.T.G., Cooper, J.E., Smith, N.H., and Feil, E.J. (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* **239**: 226–235.

Rodríguez, E., Baurecht, H., Herberich, E., Wagenpfeil, S., Brown, S.J., Cordell, H.J., et al. (2009) Meta-analysis of filaggrin polymorphisms in eczema and asthma: Robust risk factors in atopic disease. *J Allergy Clin Immunol* **123**: 1361-1370.e7.

Rojo, A., Aguinaga, A., Monecke, S., Yuste, J.R., Gastaminza, G., and España, A. (2014) *Staphylococcus aureus* genomic pattern and atopic dermatitis: May factors other than

superantigens be involved? *Eur J Clin Microbiol Infect Dis* **33**: 651–658.

Rosenberg, E. and Zilber-Rosenberg, I. (2018) The hologenome concept of evolution after 10 years. *Microbiome* **6**: 78.

Van Rossum, T., Ferretti, P., Maistrenko, O.M., and Bork, P. (2020) Diversity within species: interpreting strains in microbiomes. *Nat Rev Microbiol* **18**: 491–506.

Růžičková, V., Pantůček, R., Petráš, P., Machová, I., Kostýlková, K., and Doškař, J. (2012) Major clonal lineages in impetigo *Staphylococcus aureus* strains isolated in Czech and Slovak maternity hospitals. *Int J Med Microbiol* **302**: 237–241.

Saheb Kashaf, S., Harkins, C.P., Deming, C., Joglekar, P., Conlan, S., Holmes, C.J., et al. (2023) Staphylococcal diversity in atopic dermatitis from an individual to a global scale. *Cell Host Microbe* **31**: 578-592.e6.

Saier, M.H. and Reddy, B.L. (2015) Holins in bacteria, eukaryotes, and archaea: Multifunctional xenologues with potential biotechnological and biomedical applications. *J Bacteriol* **197**: 7–17.

Salimi, M., Barlow, J.L., Saunders, S.P., Xue, L., Gutowska-Owsiak, D., Wang, X., et al. (2013) A role for IL-25 and IL-33-driven type-2 innate lymphoid cells in atopic dermatitis. *J Exp Med* **210**: 2939–2950.

Salosensaari, A., Laitinen, V., Havulinna, A.S., Meric, G., Cheng, S., Perola, M., et al. (2021) Taxonomic signatures of cause-specific mortality risk in human gut microbiome. *Nat Commun 2021 121* **12**: 1–8.

Santiago-Rodriguez, T.M., Fornaciari, G., Luciani, S., Dowd, S.E., Toranzos, G.A., Marota, I., and Cano, R.J. (2015) Natural mummification of the human gut preserves bacteriophage DNA. *FEMS Microbiol Lett* **363**:.

Saunders, S.P., Moran, T., Floudas, A., Wurlod, F., Kaszlikowska, A., Salimi, M., et al. (2016) Spontaneous atopic dermatitis is mediated by innate immunity, with the secondary lung inflammation of the atopic March requiring adaptive immunity. *J Allergy Clin Immunol* **137**: 482–491.

Schaeffer, S.E. (2007) Graph clustering. *Comput Sci Rev* **1**: 27–64.

Schrader-Fischer, G. and Berger-Bächi, B. (2001) The AbcA transporter of *Staphylococcus aureus* affects cell autolysis. *Antimicrob Agents Chemother* **45**: 407–412.

Schulte, B., Bierbaum, G., Pohl, K., Goerke, C., and Wolz, C. (2013) Diversification of clonal complex 5 methicillin-resistant *Staphylococcus aureus* strains (Rhine-Hesse clone) within Germany. *J Clin Microbiol* **51**: 212–216.

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Sen, S., Sirobhushanam, S., Johnson, S.R., Song, Y., Tefft, R., Gatto, C., and Wilkinson, B.J. (2016) Growth-environment dependent modulation of *Staphylococcus aureus* branched-chain to straight-chain fatty acid ratio and incorporation of unsaturated fatty acids. *PLoS One* **11**: e0165300.

Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L. 0003 3527 8101, Solden, L.M., et al. (2020) DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* **48**: 8883–8900.

Shallcross, L.J., Fragaszy, E., Johnson, A.M., and Hayward, A.C. (2013) The role of the Panton-Valentine leucocidin toxin in staphylococcal disease: A systematic review and meta-analysis. *Lancet Infect Dis* **13**: 43–54.

Shepheard, M.A., Fleming, V.M., Connor, T.R., Corander, J., Feil, E.J., Fraser, C., and Hanage, W.P. (2013) Historical Zoonoses and Other Changes in Host Tropism of *Staphylococcus aureus*, Identified by Phylogenetic Analysis of a Population Dataset. *PLoS One* **8**: e62369.

Sherry, S., Xiao, C., Durbrow, K., Kimelman, M., Rodarmer, K., Shumway, M., and Yaschenko, E.

(2008) NCBI SRA Toolkit Technology for Next Generation Sequence Data.

Shi, B., Bangayan, N.J., Curd, E., Taylor, P.A., Gallo, R.L., Leung, D.Y.M., and Li, H. (2016) The skin microbiome is different in pediatric versus adult atopic dermatitis. *J Allergy Clin Immunol* **138**: 1233–1236.

Shi, B., Leung, D.Y.M., Taylor, P.A., and Li, H. (2018) Methicillin-Resistant *Staphylococcus aureus* Colonization Is Associated with Decreased Skin Commensal Bacteria in Atopic Dermatitis. *J Invest Dermatol* **138**: 1668–1671.

Shi, D., Higuchi, W., Takano, T., Saito, K., Ozaki, K., Takano, M., et al. (2011) Bullous impetigo in children infected with methicillin-resistant *Staphylococcus aureus* alone or in combination with methicillin-susceptible S. aureus: Analysis of genetic characteristics, including assessment of exfoliative toxin gene carriage. *J Clin Microbiol* **49**: 1972–1974.

Shimamori, Y., Mitsunaka, S., Yamashita, H., Suzuki, T., Kitao, T., Kubori, T., et al. (2020) Staphylococcal Phage in Combination with *Staphylococcus epidermidis* as a Potential Treatment for *Staphylococcus aureus*-Associated Atopic Dermatitis and Suppressor of Phage-Resistant Mutants. *Viruses* **13**: 7.

Sieprawska-Lupa, M., Mydel, P., Krawczyk, K., Wójcik, K., Puklo, M., Lupa, B., et al. (2004) Degradation of human antimicrobial peptide LL-37 by *Staphylococcus aureus*-derived proteinases. *Antimicrob Agents Chemother* **48**: 4673–4679.

Silhavy, T.J., Kahne, D., and Walker, S. (2010) The bacterial cell envelope. *Cold Spring Harb Perspect Biol* **2**:.

Silverberg, J.I. (2019) Comorbidities and the impact of atopic dermatitis. *Ann Allergy, Asthma Immunol* **123**: 144–151.

Simpson, E.L., Villarreal, M., Jepson, B., Rafaels, N., David, G., Hanifin, J., et al. (2018) Patients with Atopic Dermatitis Colonized with *Staphylococcus aureus* Have a Distinct Phenotype and Endotype. *J Invest Dermatol* **138**: 2224–2233.

Slatko, B.E., Gardner, A.F., and Ausubel, F.M. (2018) Overview of Next Generation Sequencing Technologies. *Curr Protoc Mol Biol* **122**: e59.

Sliz, E., Huilaja, L., Pasanen, A., Laisk, T., Reimann, E., Mägi, R., et al. (2022) Uniting biobank resources reveals novel genetic pathways modulating susceptibility for atopic dermatitis. *J Allergy Clin Immunol* **149**: 1105-1112.e9.

Smith, C.H. (2000) The Epidemiology, Causes and Prevention of Atopic Dermatitis. *Clin Exp Dermatol* **25**: 661–661.

Sonesson, A., Przybyszewska, K., Eriksson, S., Mörgelin, M., Kjellström, S., Davies, J., et al. (2017) Identification of bacterial biofilm and the *Staphylococcus aureus* derived protease, staphopain, on the skin surface of patients with atopic dermatitis. *Sci Rep* **7**: 8689.

Sospedra, I., Soriano, J.M., and Mañes, J. (2013) Enterotoxinomics: The omic sciences in the study of staphylococcal toxins analyzed in food matrices. *Food Res Int* **54**: 1052–1060.

Soucy, S.M., Huang, J., and Gogarten, J.P. (2015) Horizontal gene transfer: Building the web of life. *Nat Rev Genet* **16**: 472–482.

Spaulding, A.R., Salgado-Pabón, W., Kohler, P.L., Horswill, A.R., Leung, D.Y.M., and Schlievert, P.M. (2013) Staphylococcal and streptococcal superantigen exotoxins. *Clin Microbiol Rev* **26**: 422–447.

Spoor, L.E., McAdam, P.R., Weinert, L.A., Rambaut, A., Hasman, H., Aarestrup, F.M., et al. (2013) Livestock origin for a human pandemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *MBio* **4**:.

Stamatakis, A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stanczak-Mrozek, K.I., Laing, K.G., and Lindsay, J.A. (2017) Resistance gene transfer: Induction of transducing phage by sub-inhibitory concentrations of antimicrobials is not correlated to induction of lytic phage. *J Antimicrob Chemother* **72**: 1624–1631.

Starikova, E. V., Tikhonova, P.O., Prianichnikov, N.A., Rands, C.M., Zdobnov, E.M., Ilina, E.N., and Govorun, V.M. (2020) Phigaro: High-throughput prophage sequence annotation. *Bioinformatics* **36**: 3882–3884.

Stentzel, S., Teufelberger, A., Nordengrün, M., Kolata, J., Schmidt, F., van Crombruggen, K., et al. (2017) Staphylococcal serine protease–like proteins are pacemakers of allergic airway reactions to *Staphylococcus aureus*. *J Allergy Clin Immunol* **139**: 492-500.e8.

Strathdee, S.A., Hatfull, G.F., Mutalik, V.K., and Schooley, R.T. (2023) Phage therapy: From biological mechanisms to future directions. *Cell* **186**: 17–31.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: A genome comparison visualizer. *Bioinformatics* **27**: 1009–1010.

Sultan, A.R., Hoppenbrouwers, T., Lemmens-Den Toom, N.A., Snijders, S. V., Van Neck, J.W., Verbon, A., et al. (2019) During the early stages of *staphylococcus aureus* biofilm formation, induced neutrophil extracellular traps are degraded by autologous thermonuclease. *Infect Immun* **87**:.

Sun, Y., Wilkinson, B.J., Standiford, T.J., Akinbi, H.T., and O'Riordan, M.X.D. (2012) Fatty acids regulate stress resistance and virulence factor production for Listeria monocytogenes. *J Bacteriol* **194**: 5274–5284.

Suttle, C.A. (2007) Marine viruses - Major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.

Suttle, C.A. (1994) The significance of viruses to mortality in aquatic microbial communities. *Microb Ecol* **28**: 237–243.

Tattevin, P., Schwartz, B.S., Graber, C.J., Volinski, J., Bhukhen, Akta, Bhukhen, Arti, et al. (2012) Concurrent epidemics of skin and soft tissue infection and bloodstream infection due to community-associated methicillin-resistant *staphylococcus aureus*. *Clin Infect Dis* **55**: 781–788.

Tauber, M., Balica, S., Hsu, C.Y., Jean-Decoster, C., Lauze, C., Redoules, D., et al. (2016) *Staphylococcus aureus* density on lesional and nonlesional skin is strongly associated with disease severity in atopic dermatitis. *J Allergy Clin Immunol* **137**: 1272-1274.e3.

Teoh, W.P., Chen, X., Laczkovich, I., and Alonzo, F. (2021) *Staphylococcus aureus* adapts to the host nutritional landscape to overcome tissue-specific branched-chain fatty acid requirement. *Proc Natl Acad Sci U S A* **118**: e2022720118.

Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R.E., Mom, R., et al. (2021) PHROG: Families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics Bioinforma* **3**:.

Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J.A., and Chisholm, S.W. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci U S A* **108**:.

Thyssen, J.P., Corn, G., Wohlfahrt, J., Melbye, M., and Bager, P. (2019) Retrospective markers of paediatric atopic dermatitis persistence after hospital diagnosis: A nationwide cohort study. *Clin Exp Allergy* **49**: 1455–1463.

De Tomassi, A., Reiter, A., Reiger, M., Rauer, L., Rohayem, R., Traidl-Hoffmann, C., et al. (2023) Combining 16S Sequencing and qPCR Quantification Reveals *Staphylococcus aureus* Driven Bacterial Overgrowth in the Skin of Severe Atopic Dermatitis Patients. *Biomolecules* **13**: 1030.

Tomczak, H., Wróbel, J., Jenerowicz, D., Sadowska-Przytocka, A., Wachal, M., Adamski, Z., and Czarnecka-Operacz, M.M. (2019) The role of *Staphylococcus aureus* in atopic dermatitis: Microbiological and immunological implications. *Postep Dermatologii i Alergol* **36**: 485–491.

Tomi, N.S., Kränke, B., and Aberer, E. (2005) Staphylococcal toxins in patients with psoriasis, atopic dermatitis, and erythroderma, and in healthy control subjects. *J Am Acad Dermatol* **53**: 67–72.

Tong, S.Y.C., Davis, J.S., Eichenberger, E., Holland, T.L., and Fowler, V.G. (2015) *Staphylococcus aureus* infections: Epidemiology, pathophysiology, clinical manifestations, and management. *Clin Microbiol Rev* **28**: 603–661.

Torrelo, A. (2014) Atopic dermatitis in different skin types. What is to know? *J Eur Acad Dermatol Venereol* **28 Suppl 3**: 2–4.

Totté, J.E.E., van der Feltz, W.T., Hennekam, M., van Belkum, A., van Zuuren, E.J., and Pasmans, S.G.M.A. (2016) Prevalence and odds of *Staphylococcus aureus* carriage in atopic dermatitis: a systematic review and meta-analysis. *Br J Dermatol* **175**: 687–695.

Turner, D., Kropinski, A.M., and Adriaenssens, E.M. (2021) A Roadmap for Genome-Based Phage Taxonomy. *Viruses* **13**:.

Turner, N.A., Sharma-Kuinkel, B.K., Maskarinec, S.A., Eichenberger, E.M., Shah, P.P., Carugati, M., et al. (2019) Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nat Rev Microbiol* **17**: 203–218.

Ubelaker, M.H. and Rosenblum, E.D. (1978) Transduction of plasmid determinants in *Staphylococcus aureus* and Escherichia coli. *J Bacteriol* **133**: 699–707.

Uchiyama, J., Takemura-Uchiyama, I., Sakaguchi, Y., Gamoh, K., Kato, S.I., Daibata, M., et al. (2014) Intragenus generalized transduction in *Staphylococcus* spp. by a novel giant phage. *ISME J* **8**: 1949–1952.

Vakharia, P.P., Chopra, R., and Silverberg, J.I. (2018) Systematic Review of Diagnostic Criteria Used in Atopic Dermatitis Randomized Controlled Trials. *Am J Clin Dermatol* **19**: 15–22.

Vestergaard, M., Frees, D., and Ingmer, H. (2019) Antibiotic Resistance and the MRSA Problem. *Microbiol Spectr* **7**:.

Voigt, E., Rall, B.C., Chatzinotas, A., Brose, U., and Rosenbaum, B. (2021) Phage strategies facilitate bacterial coexistence under environmental variability. *PeerJ* **9**: e12194.

Vu, A.T., Baba, T., Chen, X., Le, T.A., Kinoshita, H., Xie, Y., et al. (2010) *Staphylococcus aureus* membrane and diacylated lipopeptide induce thymic stromal lymphopoietin in keratinocytes through the Toll-like receptor 2-Toll-like receptor 6 pathway. *J Allergy Clin Immunol* **126**:.

Van Wamel, W.J.B., Rooijakkers, S.H.M., Ruyken, M., Van Kessel, K.P.M., and Van Strijp, J.A.G. (2006) The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on β-hemolysin-converting bacteriophages. *J Bacteriol* **188**: 1310–1315.

Wang, I.N., Smith, D.L., and Young, R. (2000) Holins: The protein clocks of bacteriophage infections. *Annu Rev Microbiol* **54**: 799–825.

Wang, Q., Cai, L., Zhang, R., Wei, S., Li, F., Liu, Y., and Xu, Y. (2022) A Unique Set of Auxiliary Metabolic Genes Found in an Isolated Cyanophage Sheds New Light on Marine Phage-Host Interactions. *Microbiol Spectr* **10**:.

Wang, S., Zhao, C., Yin, Y., Chen, F., Chen, H., and Wang, H. (2022) A Practical Approach for Predicting Antimicrobial Phenotype Resistance in *Staphylococcus aureus* Through Machine Learning Analysis of Genome Data. *Front Microbiol* **13**: 605.

Wang, W., Baker, M., Hu, Y., Xu, J., Yang, D., Maciel-Guerra, A., et al. (2021) Whole-Genome Sequencing and Machine Learning Analysis of *Staphylococcus aureus* from Multiple Heterogeneous Sources in China Reveals Common Genetic Traits of Antimicrobial Resistance. *mSystems* **6**:.

Wang, Z., Hülpüsch, C., Schwierzeck, V., Alharbi, S.A., Reiger, M., Traidl-Hoffmann, C., et al. (2022) Complete and Draft Genome Sequences of 48 *Staphylococcus aureus* Isolates Obtained from Atopic Dermatitis Patients and Healthy Controls. *Microbiol Resour Announc* **11**:.

Weidenmaier, C., Kokai-Kun, J.F., Kristian, S.A., Chanturiya, T., Kalbacher, H., Gross, M., et al. (2004) Role of teichoic acids in *Staphylococcus aureus* nasal colonization, a major risk factor in nosocomial infections. *Nat Med* **10**: 243–245.

Weidinger, S. and Novak, N. (2016) Atopic dermatitis. *Lancet* **387**: 1109–1122.

Weidinger, S., Willis-Owen, S.A.G., Kamatani, Y., Baurecht, H., Morar, N., Liang, L., et al. (2013) A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. *Hum Mol Genet* **22**: 4841–4856.

Weinert, L.A., Welch, J.J., Suchard, M.A., Lemey, P., Rambaut, A., and Fitzgerald, J.R. (2012) Molecular dating of human-to-bovid host jumps by *Staphylococcus aureus* reveals an association with the spread of domestication. *Biol Lett* **8**: 829–832.

Wensel, C.R., Pluznick, J.L., Salzberg, S.L., and Sears, C.L. (2022) Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J Clin Invest* **132**:.

Werfel, T., Allam, J.-P., Biedermann, T., Eyerich, K., Gilles, S., Guttman-Yassky, E., et al. (2016) Cellular and molecular immunologic mechanisms in patients with atopic dermatitis. *J Allergy Clin Immunol* **138**: 336–349.

Wertheim, H.F., Melles, D.C., Vos, M.C., van Leeuwen, W., van Belkum, A., Verbrugh, H.A., and Nouwen, J.L. (2005) The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect Dis* **5**: 751–762.

Wickham, H. (2011) Ggplot2. *Wiley Interdiscip Rev Comput Stat* **3**: 180–185.

Williams, R.E.O. (1963) Healthy Carriage of *Staphylococcus Aureus*: Its Prevalence and Importance. *Bacteriol Rev* **27**: 56–71.

Winstel, V., Liang, C., Sanchez-Carballo, P., Steglich, M., Munar, M., Broker, B.M., et al. (2013) Wall teichoic acid structure governs horizontal gene transfer between major bacterial pathogens. *Nat Commun* **4**: 1–9.

Winter, C., Bouvier, T., Weinbauer, M.G., and Thingstad, T.F. (2010) Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the “Killing the Winner” Hypothesis Revisited. *Microbiol Mol Biol Rev* **74**: 42–57.

Xia, G. and Wolz, C. (2014) Phages of *Staphylococcus aureus* and their impact on host evolution. *Infect Genet Evol* **21**: 593–601.

Yamaguchi, T., Nishifuji, K., Sasaki, M., Fudaba, Y., Aepfelbacher, M., Takata, T., et al. (2002) Identification of the *Staphylococcus aureus* etd pathogenicity island which encodes a novel exfoliative toxin, ETD, and EDIN-B. *Infect Immun* **70**: 5835–5845.

Yang, J.J., Chang, T.W., Jiang, Y., Kao, H.J., Chiou, B.H., Kao, M.S., and Huang, C.M. (2018) Commensal *Staphylococcus aureus* provokes immunity to protect against skin infection of methicillin-resistant *staphylococcus aureus*. *Int J Mol Sci* **19**: 1290.

Yeung, M., Balma-Mena, A., Shear, N., Simor, A., Pope, E., Walsh, S., and McGavin, M.J. (2011) Identification of major clonal complexes and toxin producing strains among *Staphylococcus aureus* associated with atopic dermatitis. *Microbes Infect* **13**: 189–197.

Yin, H., Qiu, Z., Zhu, R., Wang, S., Gu, C., Yao, X., and Li, W. (2023) Dysregulated lipidome of sebum in patients with atopic dermatitis. *Allergy Eur J Allergy Clin Immunol* **78**: 1524–1537.

York, A. (2022) Phages to the rescue. *Nat Rev Microbiol* **20**: 703.

Young, R. and Bläsi, U. (1995) Holins: form and function in bacteriophage lysis. *FEMS Microbiol Rev* **17**: 191–205.

Yu, J., Jiang, F., Zhang, F., Hamushan, M., Du, J., Mao, Y., et al. (2021) Thermonucleases Contribute to *Staphylococcus aureus* Biofilm Formation in Implant-Associated Infections–A Redundant and Complementary Story. *Front Microbiol* **12**: 687888.

Zanger, P., Nurjadi, D., Schleucher, R., Scherbaum, H., Wolz, C., Kremsner, P.G., and Schulte, B. (2012) Import and spread of Panton-Valentine leukocidin-positive *staphylococcus aureus* through nasal carriage and skin infections in travelers returning from the tropics and subtropics. *Clin Infect Dis* **54**: 483–492.

Zapotoczna, M., Riboldi, G.P., Moustafa, A.M., Dickson, E., Narechania, A., Morrissey, J.A., et al. (2018) Mobile-genetic-element-encoded hypertolerance to copper protects *Staphylococcus aureus* from killing by host phagocytes. *MBio* **9**:.

Zeaki, N., Susilo, Y.B., Pregiel, A., Rådström, P., and Schelin, J. (2015) Prophage-encoded staphylococcal enterotoxin a: Regulation of production in *Staphylococcus aureus* strains representing different sea regions. *Toxins (Basel)* **7**: 5359–5376.

Zhang, M., Jiang, Z., Li, D., Jiang, D., Wu, Y., Ren, H., et al. (2015) Oral Antibiotic Treatment Induces Skin Microbiota Dysbiosis and Influences Wound Healing. *Microb Ecol* **69**: 415–421.

Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., et al. (2014) PanGP: A tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* **30**: 1297–1299.

Zhou, W., Spoto, M., Hardy, R., Guan, C., Fleming, E., Larson, P.J., et al. (2020) Host-Specific Evolutionary and Transmission Dynamics Shape the Functional Diversification of *Staphylococcus epidermidis* in Human Skin. *Cell* **180**: 454-470.e18.

Zhu, J., Li, H., Jing, Z.Z., Zheng, W., Luo, Y.R., Chen, S.X., and Guo, F. (2022) Robust host source tracking building on the divergent and non-stochastic assembly of gut microbiomes in wild and farmed large yellow croaker. *Microbiome* **10**: 1–15.

ZINDER, N.D. (1955) Bacterial transduction. *J Cell Physiol Suppl* **45**: 23–49.

Zipperer, A., Konnerth, M.C., Laux, C., Berscheid, A., Janek, D., Weidenmaier, C., et al. (2016) Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**: 511–516.

Zuo, G. (2021) CVTree: A Parallel Alignment-free Phylogeny and Taxonomy Tool Based on Composition Vectors of Genomes. *Genomics, Proteomics Bioinforma* **19**: 662–667.

van Zuuren, E.J., Fedorowicz, Z., Christensen, R., Lavrijsen, A., and Arents, B.W.M. (2017) Emollients and moisturisers for eczema. *Cochrane Database Syst Rev* **2017**:.

List of abbreviations

AD	Atopic dermatitis
AMG	Auxiliary metabolic gene
AMR	Antimicrobial resistance
ARG	Antibiotic resistance gene
AUC-ROC	Area under the curve - receiver operating characteristics
CC	Clonal complex
chp	Chemotaxis-inhibitory protein
CoNS	Coagulase-negative Staphylococcus
CWA	Cell wall anchored protein
dN/dS	Ratio of nonsynonymous and synonymous substitutions
eta	exfoliative toxins
FLG	Filaggrin
GBD	Global burden of disease
HE	Healthy individuals
HGT	Horizontal gene transfer
MGE	Mobile genetic element
ML	Machine learning
MLST	Multi-locus sequence typing
MRSA	Methicillin-resistant *S. aureus*
NGS	Next-generation sequencing
NMF	Natural moisturizing factor
ORF	Open reading frame
PCA	Principal Component analysis
PCoA	Principal Coordinates Analysis
PVL	Panton-Valentine leucocidin
RF	Random Forest
sak	Staphylokinase
scn	Staphylococcal complement inhibitor
Ses	Enterotoxins
SNP	Single nucleotide polymorphisms

SRA	Sequence Read Archive
SSSS	Staphylococcal scalded-skin syndrome
TGS	Third-generation sequencing
TSLP	Thymic stromal lymphopoietin
VF	Virulence factors
WGS	Whole genome sequencing

Acknowledgments

Nearly five years ago, with bags packed and goodbye said to my parents, I embarked on a solo journey to Germany. Reflecting in the departure lounge, I was overwhelmed with emotions: the mix of excitement and apprehension about the future filled me with boundless dreams and expectations. Now, as time has swiftly passed, I stand on the verge of earning my PhD. This journey, though arduous and fraught with challenges, has immensely strengthened me, both academically and personally, marking a period of significant growth and resilience.

Special thanks go to Prof. Dr. Michael Schloter, whose expertise and guidance significantly enriched my PhD experience, providing the support necessary from the outset to the successful conclusion of my thesis and paving a solid way for my future career. My gratitude also extends to Dr. Bärbel Foesel for her supervision in the initial three years of my PhD. Further deep appreciation is due to my thesis committee members, Dr. Matthias Reiger and Dr. Claudia Hülpüsch, for their invaluable mentoring and accessibility all the time through my PhD journey, which were crucial for developing my scientific taste. I express my profound gratitude to my collaborators Prof. Claudia Traidl-Hoffmann, Prof. Li Deng, Dr. Mohammadali Khan Mirzaei, and Xue Peng, for their engaging and fruitful collaborations on my papers, and also to Prof. Karin Pritsch for her kind and invaluable assistance and inspiration.

Of course, I cannot overlook the camaraderie and support of my dear friends and colleagues, Yuri, Sarah, Roberto, Juliette, Seun, Benoit, David, and all others (too many to count), who have provided stimulating conversations and much-needed joyfulness. Your presence made this challenging journey way more memorable and happier.

Lastly but not the least, I owe a debt of gratitude to my family for their forever unconditional support. To my parents, for their unwavering faith in me and constant encouragement; to my wife, Lili, for her love, understanding, and companionship; and to our daughter, Amo, born in January 2023, who has brought unparalleled happiness into our lives; you are the greatest paper of this journey.

This thesis embodies not just my efforts, but also the collective support and confidence that various individuals have generously extended to me. Thank you all, and to my diligent self.

Spring 2024

Munich, Germany

Appendix

Supplementary Tables:

Table S1. Metadata of the 300 *S. aureus* strains from public databases

Assembly	Health status	ST	BioSample	geographic location	Completeness	Contamination	Genome size (bp)	# scaffolds	N50 (scaffolds)	Longest scaffold (bp)	GC	# predicted genes
GCA_003112635.1	AD	1	SAMD00105921	Japan	99,51	0,08	2719131	21	238277	488984	32,7	2479
GCA_003113915.1	AD	1	SAMD00105922	Japan	99,51	0,08	2710485	56	80924	245435	32,7	2478
GCA_003113935.1	AD	1	SAMD00105923	Japan	99,32	0,08	2739227	23	208543	565691	32,7	2513
GCA_003113955.1	AD	5	SAMD00105924	Japan	98,53	0,82	2668542	225	20829	92976	32,9	2498
GCA_003113975.1	AD	5	SAMD00105925	Japan	99,51	0,11	2879122	21	273736	731396	32,6	2694
GCA_003113995.1	AD	5	SAMD00105926	Japan	98,01	0,13	2826326	172	28894	134659	32,7	2689
GCA_003114015.1	AD	5	SAMD00105927	Japan	99,51	0,08	2786638	23	200061	850683	32,7	2576
GCA_003114035.1	AD	5	SAMD00105928	Japan	99,34	0,08	2661369	33	227363	371517	32,7	2436
GCA_003114055.1	AD	5	SAMD00105929	Japan	99,51	0,08	2758253	26	178073	357119	32,8	2579
GCA_003114075.1	AD	5	SAMD00105930	Japan	99,51	0,08	2746442	48	107121	237608	32,7	2571
GCA_003114095.1	AD	5	SAMD00105931	Japan	99,23	0,08	2755584	80	70498	242145	32,7	2566
GCA_003114115.1	AD	5	SAMD00105932	Japan	99,44	0,11	2669492	96	49530	155988	32,7	2461
GCA_003116115.1	AD	5	SAMD00105933	Japan	99,41	0,11	2701085	54	83780	327035	32,7	2485
GCA_003114335.1	AD	5	SAMD00105934	Japan	99,51	0,08	2698604	40	114766	355132	32,7	2492
GCA_003116185.1	AD	5	SAMD00105935	Japan	99,51	0,11	2710866	102	45866	108687	32,6	2536
GCA_003114595.1	AD	5	SAMD00105936	Japan	98,84	0,08	2770385	74	83545	320907	32,7	2599
GCA_003114615.1	AD	5	SAMD00105937	Japan	99,51	0,08	2791276	29	153316	580661	32,7	2602
GCA_003114635.1	AD	5	SAMD00105938	Japan	99,51	0,64	2795806	34	147777	528218	32,7	2607
GCA_003114655.1	AD	5	SAMD00105939	Japan	99,51	0,11	2825769	50	121256	210479	32,7	2644
GCA_003114675.1	AD	5	SAMD00105940	Japan	99,51	0,08	2707117	26	204006	391946	32,7	2523
GCA_002925605.1	AD	1	SAMN08456629	Russia	99,51	0,13	2718298	15	674576	1350013	32,7	2510
GCA_003017835.1	AD	1	SAMN08456633	Russia	99,51	0,13	2666594	86	124869	249739	32,7	2450
GCA_002925615.1	AD	1	SAMN08456648	Russia	99,51	0,08	2742926	64	187578	317884	32,7	2544
GCA_003017755.1	AD	1	SAMN08456651	Russia	99,51	0,08	2754974	59	114549	330595	32,7	2566
GCA_003017735.1	AD	1	SAMN08456662	Russia	99,51	0,18	2713020	176	51999	171744	32,7	2544
GCA_015714625.1	AD	582	SAMN10689345	USA	99,51	2,33	2801767	19	342008	662204	32,9	2558
GCA_015714615.1	AD	582	SAMN10689346	USA	99,51	0,25	2880750	36	152386	324611	32,9	2689
GCA_015714605.1	AD	582	SAMN10689347	USA	99,51	0,08	2800401	19	916832	1198293	32,9	2548
GCA_015714575.1	AD	582	SAMN10689349	USA	99,51	0,64	2943889	46	202345	354358	32,8	2772
GCA_015714565.1	AD	582	SAMN10689350	USA	99,51	0,08	2766366	21	650910	747645	32,9	2539
GCA_015714545.1	AD	464	SAMN10689351	USA	99,51	0,64	2955838	46	159209	382339	32,8	2776
GCA_015714515.1	AD	398	SAMN10689354	USA	99,51	0,08	2807311	27	454157	633813	32,9	2545

GCA_015714505.1	AD	398	SAMN10689355	USA	99,51	0,08	2748058	21	454147	632461	32,9	2481
GCA_015714445.1	AD	188	SAMN10689356	USA	99,51	0,08	2744233	22	375520	570485	32,9	2474
GCA_015714455.1	AD	188	SAMN10689357	USA	99,51	0,08	2787218	11	971708	1341435	33	2556
GCA_012274715.1	AD	8	SAMN10689358	USA	99,51	0,08	2756502	12	865560	1296930	32,9	2491
GCA_015714475.1	AD	188	SAMN10689359	USA	99,51	0,08	2941236	36	185113	514850	32,9	2725
GCA_015714405.1	AD	109	SAMN10689360	USA	99,51	0,08	2837863	21	541458	966145	32,9	2602
GCA_015714415.1	AD	188	SAMN10689361	USA	99,51	0,64	2961729	39	320885	594907	32,8	2781
GCA_012274685.1	AD	8	SAMN10689363	USA	99,51	0,22	2910141	59	131488	290753	32,8	2746
GCA_015714365.1	AD	101	SAMN10689366	USA	99,51	0,08	2829081	23	870954	992495	32,9	2599
GCA_016107245.1	AD	672	SAMN10689367	USA	99,37	0,08	2898379	43	304416	509627	32,8	2731
GCA_012274665.1	AD	8	SAMN10689370	USA	99,51	0,08	2793877	26	454010	633897	32,9	2532
GCA_015714355.1	AD	101	SAMN10689371	USA	99,51	0,08	2795620	25	454010	633897	32,9	2530
GCA_015714345.1	AD	101	SAMN10689372	USA	99,51	0,08	2773401	19	446464	763080	32,8	2517
GCA_015714315.1	AD	97	SAMN10689374	USA	99,51	0,13	2738309	14	606771	1101688	33	2497
GCA_015714305.1	AD	97	SAMN10689375	USA	99,51	0,08	2805890	21	688039	820585	33	2565
GCA_015714285.1	AD	97	SAMN10689376	USA	99,51	0,22	2790674	30	176722	363345	33	2531
GCA_015714255.1	AD	97	SAMN10689377	USA	99,51	0,08	2759119	19	633232	858694	33	2493
GCA_015714245.1	AD	97	SAMN10689378	USA	99,51	0,08	2755934	19	633232	858694	33	2490
GCA_015714225.1	AD	97	SAMN10689379	USA	99,51	0,08	2747182	22	446525	760980	32,8	2474
GCA_012274605.1	AD	6	SAMN10689380	USA	99,51	1,79	2829911	15	685041	964546	33	2613
GCA_015714205.1	AD	97	SAMN10689382	USA	99,51	0,22	2879514	41	174680	392829	33	2663
GCA_012274765.1	AD	8	SAMN10689383	USA	99,51	0,64	2776899	45	187598	586611	32,7	2552
GCA_015714165.1	AD	87	SAMN10689384	USA	99,51	0,08	2803368	11	982742	1357496	32,9	2548
GCA_012274615.1	AD	7	SAMN10689385	USA	99,51	0,08	2871584	22	437593	561066	32,9	2653
GCA_015714175.1	AD	97	SAMN10689386	USA	99,51	0,22	2955660	49	139846	324207	32,9	2826
GCA_015714145.1	AD	81	SAMN10689387	USA	99,51	0,08	2897362	31	180596	394792	33	2692
GCA_015714115.1	AD	81	SAMN10689388	USA	99,51	0,11	2946423	28	237225	597275	32,8	2716
GCA_012274575.1	AD	6	SAMN10689389	USA	99,51	0,11	2935170	48	187776	386701	32,8	2732
GCA_015714065.1	AD	81	SAMN10689391	USA	99,51	0,22	2892686	40	170371	331518	32,9	2715
GCA_015714075.1	AD	81	SAMN10689392	USA	99,51	0,22	2746492	31	230406	343620	33	2464
GCA_015714045.1	AD	81	SAMN10689393	USA	99,51	0,22	2748869	31	230406	343397	33	2463
GCA_015714025.1	AD	72	SAMN10689394	USA	99,51	0,1	2757680	13	490174	1358069	33	2504
GCA_015713965.1	AD	72	SAMN10689395	USA	99,41	0,08	2760944	12	543458	1313386	32,9	2478
GCA_015713945.1	AD	59	SAMN10689396	USA	99,51	0,64	2951775	36	202345	512390	32,8	2775
GCA_015713975.1	AD	72	SAMN10689398	USA	99,51	0,08	2747152	27	275025	538888	32,9	2494
GCA_015713955.1	AD	72	SAMN10689399	USA	99,51	0,08	2760082	34	170235	550590	32,8	2559
GCA_015713915.1	AD	59	SAMN10689400	USA	99,51	0,1	2822363	38	241832	550274	32,7	2631
GCA_015713905.1	AD	59	SAMN10689401	USA	99,51	0,64	2959449	42	345890	454582	32,7	2791
GCA_016107265.1	AD	2276	SAMN10689403	USA	99,51	0,08	2813979	17	598493	1170101	32,8	2564
GCA_015713855.1	AD	45	SAMN10689404	USA	99,51	0,64	2815953	39	287165	517078	32,7	2586
GCA_015713865.1	AD	45	SAMN10689405	USA	99,51	0,64	2865121	34	346006	834970	32,7	2651
GCA_015713845.1	AD	30	SAMN10689406	USA	99,51	0,64	2856737	35	305124	582172	32,6	2650
GCA_012274705.1	AD	8	SAMN10689408	USA	99,51	0,1	2759367	16	509508	660879	33	2479
GCA_016107225.1	AD	672	SAMN10689409	USA	99,23	0,08	2775911	19	566023	883770	32,9	2534
GCA_015713805.1	AD	30	SAMN10689411	USA	99,51	0,64	2869160	27	610731	860022	32,7	2672
GCA_015713765.1	AD	30	SAMN10689413	USA	99,51	0,64	2905691	41	176253	804317	32,7	2712
GCA_015713745.1	AD	30	SAMN10689416	USA	99,51	0,08	2842001	21	597517	1184049	32,9	2592
GCA_015713715.1	AD	30	SAMN10689417	USA	99,51	0,08	2738725	18	439504	613190	33	2491

GCA_015713705.1	AD	30	SAMN10689418	USA	99,51	0,08	2912916	36	331721	958026	33	2724
GCA_015713665.1	AD	30	SAMN10689419	USA	99,51	0,22	2861852	32	267374	363402	33	2633
GCA_015713655.1	AD	22	SAMN10689421	USA	99,51	0,08	2796422	26	632315	858691	32,9	2568
GCA_015713645.1	AD	22	SAMN10689423	USA	99,4	0,13	2752643	16	869623	874496	33	2528
GCA_015713605.1	AD	15	SAMN10689425	USA	99,51	0,08	2793112	20	338213	763436	32,9	2540
GCA_015713555.1	AD	15	SAMN10689426	USA	99,49	0,22	2778790	34	176624	307700	32,9	2538
GCA_015713575.1	AD	15	SAMN10689427	USA	99,51	0,08	2782144	13	971474	1337347	33	2556
GCA_015713545.1	AD	15	SAMN10689428	USA	99,51	0,64	2881184	36	213968	582538	32,8	2675
GCA_015713515.1	AD	15	SAMN10689430	USA	99,51	0,64	2926032	35	346005	898262	32,7	2731
GCA_015713505.1	AD	15	SAMN10689432	USA	99,51	0,64	2925858	38	194939	807882	32,7	2729
GCA_015713485.1	AD	15	SAMN10689434	USA	99,51	0,64	2898634	240	33573	138004	32,7	2758
GCA_015713455.1	AD	15	SAMN10689435	USA	99,51	0,64	2924183	43	346538	808266	32,7	2735
GCA_015713445.1	AD	15	SAMN10689436	USA	99,37	0,08	2820480	124	49481	138019	32,8	2642
GCA_015713395.1	AD	8	SAMN10689437	USA	99,51	0,64	2909857	46	160734	551300	32,7	2733
GCA_015713385.1	AD	8	SAMN10689438	USA	99,51	0,08	2821461	20	331263	702916	32,9	2602
GCA_015713425.1	AD	15	SAMN10689439	USA	99,51	0,64	2965478	45	220277	537350	32,8	2788
GCA_015713365.1	AD	8	SAMN10689443	USA	99,51	0,11	2818164	24	316964	999797	32,7	2582
GCA_015713325.1	AD	8	SAMN10689445	USA	99,43	0,11	2814983	27	255218	491584	32,7	2582
GCA_015713335.1	AD	8	SAMN10689446	USA	99,51	0,11	2819575	21	320633	1314845	32,7	2578
GCA_015713295.1	AD	8	SAMN10689447	USA	99,51	0,22	2861921	31	194463	409780	32,9	2619
GCA_015713275.1	AD	8	SAMN10689448	USA	99,51	0,08	2856012	26	252414	519030	32,9	2615
GCA_015713265.1	AD	8	SAMN10689449	USA	99,51	0,22	2931419	39	172893	394310	32,9	2785
GCA_015713245.1	AD	8	SAMN10689450	USA	99,51	0,08	2762510	19	438770	654415	33	2514
GCA_015713195.1	AD	8	SAMN10689451	USA	99,51	0,13	2761598	16	302547	984245	32,9	2523
GCA_015713185.1	AD	8	SAMN10689452	USA	99,41	0,08	2767418	17	497670	573359	32,9	2523
GCA_015713205.1	AD	8	SAMN10689453	USA	99,51	0,64	2948029	45	180818	453704	32,8	2788
GCA_015713105.1	AD	8	SAMN10689455	USA	99,51	0,08	2894889	38	124931	304770	32,9	2678
GCA_015713165.1	AD	8	SAMN10689456	USA	99,51	0,08	2758347	21	411416	716606	32,9	2498
GCA_015713145.1	AD	8	SAMN10689457	USA	98,95	0,1	2798417	10	1027069	1200029	32,9	2573
GCA_015713115.1	AD	8	SAMN10689458	USA	99,41	0,08	2755153	32	204978	494745	32,7	2503
GCA_015713085.1	AD	8	SAMN10689459	USA	99,41	0,08	2737591	30	270487	494824	32,8	2481
GCA_015713055.1	AD	8	SAMN10689460	USA	99,41	0,08	2734350	29	205191	494745	32,8	2484
GCA_015713045.1	AD	8	SAMN10689461	USA	99,41	0,08	2740742	36	205109	495143	32,8	2488
GCA_015713015.1	AD	8	SAMN10689462	USA	99,51	0,64	2904481	47	202205	582040	32,7	2714
GCA_015712965.1	AD	8	SAMN10689463	USA	99,51	0,08	2793680	26	336942	633896	32,9	2534
GCA_015713005.1	AD	8	SAMN10689464	USA	99,51	0,08	2777978	21	381390	627565	32,9	2506
GCA_015712985.1	AD	8	SAMN10689465	USA	99,51	0,1	2832208	13	1037039	1057715	32,9	2612
GCA_015712925.1	AD	8	SAMN10689466	USA	99,51	0,08	2827843	28	275134	542340	32,9	2619
GCA_015712905.1	AD	8	SAMN10689467	USA	99,51	0,08	2781684	24	276624	763412	32,8	2515
GCA_015712945.1	AD	8	SAMN10689468	USA	99,51	0,08	2695781	21	409495	612892	32,9	2432
GCA_015712885.1	AD	8	SAMN10689469	USA	99,51	0,36	2831456	41	152807	331532	32,9	2623
GCA_015712815.1	AD	8	SAMN10689470	USA	99,51	0,08	2806879	21	713014	813137	32,9	2558
GCA_015712865.1	AD	8	SAMN10689471	USA	99,4	0,08	2780537	23	673783	743278	32,9	2553
GCA_015712805.1	AD	8	SAMN10689472	USA	99,4	0,08	2778553	23	426289	743211	32,9	2560
GCA_015712835.1	AD	8	SAMN10689473	USA	99,51	0,08	2811985	23	272234	592853	32,8	2545
GCA_015712775.1	AD	8	SAMN10689474	USA	99,41	0,08	2886229	16	694071	1340388	32,7	2647
GCA_015712715.1	AD	8	SAMN10689475	USA	99,51	0,08	2823461	16	929288	1171194	32,9	2593
GCA_015712705.1	AD	8	SAMN10689476	USA	99,51	0,13	2783847	14	677005	1306291	33	2536

GCA_011319695.1	AD	5	SAMN10875734	USA	99,51	0,08	2799774	53	148454	530975	32,7	2625
GCA_011319675.1	AD	5	SAMN10875735	USA	99,51	0,08	2715043	49	120737	278830	32,7	2491
GCA_011319685.1	AD	5	SAMN10875736	USA	99,51	0,22	2831910	57	115199	297043	32,7	2689
GCA_011319615.1	AD	5	SAMN10875737	USA	99,51	0,22	2697599	416	12677	48857	32,8	2563
GCA_011319595.1	AD	5	SAMN10875738	USA	99,51	0,08	2782338	61	103463	199360	32,7	2574
GCA_011319655.1	AD	5	SAMN10875739	USA	99,51	0,08	2798418	56	105785	274302	32,7	2653
GCA_011319625.1	AD	5	SAMN10875740	USA	99,51	0,1	2765828	73	115385	321639	32,7	2599
GCA_011319565.1	AD	5	SAMN10875741	USA	99,51	0,11	2885565	71	85007	249138	32,6	2710
GCA_011319535.1	AD	5	SAMN10875742	USA	99,51	0,64	2739705	71	87016	217158	32,6	2522
GCA_011319515.1	AD	5	SAMN10875743	USA	99,51	0,1	2723095	67	78032	182226	32,7	2509
GCA_005774645.1	AD	5	SAMN11585614	Russia	99,51	0,16	2821661	341	56036	220849	32,9	2856
GCA_020907025.1	AD	-	SAMN17729798	USA	98,95	0,08	2745773	58	112807	247713	32,7	2557
GCA_020906965.1	AD	-	SAMN17729799	USA	98,95	0,19	2703168	24	231214	463220	32,8	2486
GCA_020907005.1	AD	-	SAMN17729800	USA	98,83	0,08	2721982	70	90716	277570	32,7	2515
GCA_020906975.1	AD	-	SAMN17729804	USA	98,95	0,08	2739068	78	78642	158694	32,7	2542
GCA_020907065.1	AD	-	SAMN17729810	USA	98,95	0,08	2744177	18	363737	764272	32,7	2531
GCA_020907085.1	AD	-	SAMN17729811	USA	98,95	0,08	2747473	58	103296	209078	32,7	2564
GCA_020907125.1	AD	-	SAMN17729812	USA	98,87	0,19	2748284	79	91325	185177	32,7	2576
GCA_020907035.1	AD	-	SAMN17729813	USA	98,95	0,64	2746351	68	77910	185727	32,7	2570
GCA_020907145.1	AD	-	SAMN17729826	USA	99,51	0,25	2832934	48	150429	419457	32,8	2698
GCA_020906905.1	AD	-	SAMN17729830	USA	99,51	0,64	2917496	26	631715	1241310	32,6	2735
GCA_020906865.1	AD	-	SAMN17729831	USA	99,51	0,68	2915150	34	336411	466183	32,6	2736
ERS1638080	HE	1	SAMEA103948912	UK	98,95	0,13	2705401	53	196370	498328	32,8	2504
ERS1638081	HE	1	SAMEA103948913	UK	98,95	0,13	2705785	42	393918	1280983	32,8	2501
ERS1638082	HE	1	SAMEA103948914	UK	98,95	0,13	2706989	49	260435	857234	32,8	2506
ERS1638083	HE	1	SAMEA103948915	UK	98,95	0,13	2705509	42	260320	1280899	32,8	2502
ERS1638084	HE	5	SAMEA103948916	UK	98,95	0,13	2705524	40	393917	1280983	32,8	2501
ERS1638095	HE	5	SAMEA103948927	UK	99,51	0,08	2700950	50	315349	657199	32,8	2492
ERS1638096	HE	5	SAMEA103948928	UK	99,51	0,08	2702813	54	315349	570525	32,8	2498
ERS1638097	HE	5	SAMEA103948929	UK	99,51	0,08	2700057	49	438752	657172	32,7	2490
ERS1638098	HE	5	SAMEA103948930	UK	99,51	0,08	2700734	48	437254	657126	32,7	2495
ERS1638100	HE	5	SAMEA103948932	UK	99,41	0,08	2686409	65	252291	520809	32,7	2474
ERS1638103	HE	5	SAMEA103948935	UK	99,41	0,08	2686618	60	252269	520930	32,7	2470
ERS1638112	HE	5	SAMEA103948944	UK	99,51	0,22	2814292	82	174964	393496	32,7	2675
ERS1638115	HE	5	SAMEA103948947	UK	99,51	0,08	2772372	62	274240	489369	32,7	2579
ERS1638116	HE	5	SAMEA103948948	UK	99,51	0,27	2767260	38	427443	769642	32,7	2562
ERS1638117	HE	5	SAMEA103948949	UK	99,51	0,08	2763997	35	427674	769642	32,7	2560
ERS1638118	HE	5	SAMEA103948950	UK	99,51	0,26	2769783	42	428370	769642	32,8	2569
ERS1638120	HE	5	SAMEA103948952	UK	99,51	0,22	2811502	71	174964	417908	32,7	2659
ERS1638121	HE	5	SAMEA103948953	UK	99,51	0,22	2812999	78	174965	417908	32,7	2667
ERS1638122	HE	5	SAMEA103948954	UK	99,51	0,22	2812502	76	174965	417908	32,7	2666
ERS1638123	HE	5	SAMEA103948955	UK	99,51	0,22	2813899	78	209117	417908	32,7	2668
ERS1638124	HE	5	SAMEA103948956	UK	99,51	0,22	2812732	74	174965	417908	32,7	2663
ERS1638125	HE	5	SAMEA103948957	UK	98,95	0,22	2778681	65	174831	352261	32,8	2592
ERS1638126	HE	5	SAMEA103948958	UK	98,95	0,22	2773234	62	174829	352261	32,8	2587
ERS1638127	HE	5	SAMEA103948959	UK	98,95	0,22	2773269	60	174829	352261	32,8	2590
ERS1638128	HE	5	SAMEA103948960	UK	98,95	0,22	2775440	60	175059	352261	32,8	2590
ERS1638129	HE	5	SAMEA103948961	UK	98,95	0,22	2775564	65	174828	352262	32,8	2598

ERS1638130	HE	5	SAMEA103948962	UK	99,43	0,22	2780875	63	196536	447372	32,7	2563
ERS1638131	HE	5	SAMEA103948963	UK	99,43	0,22	2779194	59	176734	326295	32,7	2562
ERS1638132	HE	5	SAMEA103948964	UK	99,43	0,22	2780819	68	176740	447372	32,7	2567
ERS1638133	HE	5	SAMEA103948965	UK	99,43	0,22	2783037	73	216624	447371	32,7	2567
ERS1638134	HE	5	SAMEA103948966	UK	99,43	0,22	2781966	78	191342	326295	32,7	2572
ERS1638149	HE	5	SAMEA103948981	UK	99,41	0,08	2685433	67	243859	506078	32,7	2481
ERS1638150	HE	5	SAMEA103948982	UK	99,51	0,22	2815533	85	174985	426871	32,7	2673
ERS1638152	HE	5	SAMEA103948984	UK	99,51	0,22	2815574	79	174987	426871	32,7	2670
ERS1638153	HE	5	SAMEA103948985	UK	99,51	0,22	2815158	82	175446	424672	32,7	2673
ERS1638154	HE	5	SAMEA103948986	UK	99,51	0,22	2815474	84	175446	424859	32,7	2674
ERS1638160	HE	5	SAMEA103948992	UK	99,48	0,22	2851463	61	208292	756063	32,7	2700
ERS1638161	HE	5	SAMEA103948993	UK	99,48	0,22	2851444	65	208759	489289	32,7	2702
ERS1638162	HE	5	SAMEA103948994	UK	99,51	0,08	2716237	40	448187	1021166	32,7	2486
ERS1638163	HE	5	SAMEA103948995	UK	99,48	0,22	2852877	72	171343	434486	32,7	2701
ERS1638164	HE	5	SAMEA103948996	UK	99,48	0,22	2852516	69	208759	755262	32,7	2708
ERS1638165	HE	5	SAMEA103948997	UK	99,51	0,08	2811090	40	617779	918038	32,7	2596
ERS1638167	HE	5	SAMEA103948999	UK	99,51	0,08	2808117	29	790264	1002062	32,7	2583
ERS1638168	HE	5	SAMEA103949000	UK	99,51	0,08	2808469	34	702652	757385	32,7	2584
ERS1638169	HE	5	SAMEA103949001	UK	99,51	0,08	2807754	30	1002603	1339124	32,7	2588
ERS1034193	HE	1	SAMEA3727044	Gambia	99,51	0,15	2748275	85	252819	652776	32,7	2567
ERS1034196	HE	1	SAMEA3727047	Gambia	99,51	0,22	2812202	63	170887	300428	32,8	2641
ERS1034197	HE	1	SAMEA3727048	Gambia	99,51	0,13	2830241	46	989952	1025681	32,8	2679
ERS1034199	HE	1	SAMEA3727050	Gambia	99,23	0,08	2737386	36	243933	594599	32,7	2513
ERS1034201	HE	1	SAMEA3727052	Gambia	99,51	0,08	2822925	53	320759	503327	32,7	2653
ERS1034203	HE	1	SAMEA3727054	Gambia	99,23	0,08	2789306	80	244050	594232	32,8	2614
ERS5272144	HE	5	SAMEA7515657	Australia	99,51	0,08	2848331	41	317625	651926	32,7	2682
ERS5272146	HE	5	SAMEA7515659	Australia	99,51	0,08	2848633	42	583613	1189910	32,7	2684
ERS5272149	HE	8	SAMEA7515662	Australia	99,51	0,08	2802792	43	272618	651705	32,7	2610
ERS5272150	HE	8	SAMEA7515663	Australia	99,51	0,08	2849139	51	1465870	1465870	32,7	2693
ERS5272151	HE	8	SAMEA7515664	Australia	99,51	0,08	2849009	44	570533	1189902	32,7	2683
ERS5272152	HE	8	SAMEA7515665	Australia	99,51	0,08	2845558	35	1465312	1465312	32,7	2678
ERS5272154	HE	8	SAMEA7515667	Australia	99,51	0,08	2849203	41	279728	991850	32,7	2686
ERS5272155	HE	12	SAMEA7515668	Australia	99,51	0,08	2813454	39	583201	1189908	32,8	2636
ERS5272156	HE	12	SAMEA7515669	Australia	99,51	0,08	2848150	40	583614	1189847	32,7	2683
ERS5272157	HE	12	SAMEA7515670	Australia	99,51	0,08	2848375	45	583613	1192934	32,7	2681
ERS5272159	HE	12	SAMEA7515672	Australia	99,51	0,08	2849494	47	583714	844657	32,7	2694
ERS5272160	HE	12	SAMEA7515673	Australia	99,51	0,11	2884004	42	583613	1129000	32,7	2728
ERS5272161	HE	12	SAMEA7515674	Australia	99,51	0,11	2883230	50	584885	1189892	32,7	2727
ERS5272163	HE	15	SAMEA7515676	Australia	99,51	0,08	2848596	41	584492	1189910	32,7	2684
ERS5272164	HE	15	SAMEA7515677	Australia	99,51	0,08	2847749	38	583554	1189911	32,7	2683
ERS5272165	HE	15	SAMEA7515678	Australia	99,51	0,08	2850594	38	583615	1189851	32,7	2689
ERS5272166	HE	15	SAMEA7515679	Australia	99,51	0,08	2848147	39	583613	1189797	32,7	2684
ERS5272167	HE	15	SAMEA7515680	Australia	99,51	0,08	2848044	43	583615	1189907	32,7	2684
ERS5272168	HE	15	SAMEA7515681	Australia	99,51	0,08	2849069	50	583732	904195	32,7	2691
ERS5272169	HE	15	SAMEA7515682	Australia	99,51	0,08	2848054	45	279728	1084448	32,7	2686
ERS5272170	HE	22	SAMEA7515683	Australia	99,51	0,08	2849450	49	583620	1189922	32,7	2696
ERS5272171	HE	22	SAMEA7515684	Australia	99,51	0,11	2884157	42	584904	1189909	32,7	2729
ERS5272172	HE	22	SAMEA7515685	Australia	99,51	0,08	2762065	40	317625	626351	32,7	2553

ERS5272173	HE	22	SAMEA7515686	Australia	99,51	0,08	2849114	44	583613	1189911	32,7	2686
ERS5272174	HE	22	SAMEA7515687	Australia	99,51	0,08	2848668	43	583614	1189844	32,7	2685
ERS5272175	HE	30	SAMEA7515688	Australia	99,51	0,08	2876648	38	583615	1189634	32,7	2713
ERS5272176	HE	30	SAMEA7515689	Australia	99,51	0,08	2848387	41	583615	1189908	32,7	2685
ERS5312759	HE	30	SAMEA7556117	Australia	99,51	0,08	2777697	50	244109	708330	32,7	2581
ERS5312760	HE	30	SAMEA7556118	Australia	99,51	0,08	2717570	39	244095	639691	32,7	2500
ERS5312763	HE	30	SAMEA7556121	Australia	99,51	0,08	2739697	54	230399	350862	32,8	2577
ERS5312764	HE	30	SAMEA7556122	Australia	99,51	0,13	2795860	44	327509	1208296	32,7	2624
ERS5312766	HE	30	SAMEA7556124	Australia	99,51	0,08	2771873	60	183356	550768	32,7	2622
ERS5312767	HE	34	SAMEA7556125	Australia	99,51	0,08	2774395	47	183425	413942	32,7	2615
ERS5312769	HE	39	SAMEA7556127	Australia	99,51	0,08	2760082	43	180174	616242	32,6	2541
ERS5312774	HE	39	SAMEA7556132	Australia	99,51	0,08	2810113	55	243722	641997	32,7	2627
ERS5312775	HE	39	SAMEA7556133	Australia	99,51	0,08	2788290	35	374913	703128	32,7	2590
ERS5312777	HE	39	SAMEA7556135	Australia	99,51	0,22	2844917	56	305557	621939	32,7	2698
ERS5312778	HE	39	SAMEA7556136	Australia	99,51	0,08	2848688	42	318041	871978	32,7	2689
ERS5312781	HE	45	SAMEA7556139	Australia	99,51	0,13	2832595	32	353579	822925	32,7	2675
ERS5312782	HE	45	SAMEA7556140	Australia	99,51	0,13	2837308	34	318972	460509	32,7	2681
ERS5312785	HE	45	SAMEA7556143	Australia	99,32	0,08	2855835	38	425438	552442	32,6	2658
ERS5312787	HE	45	SAMEA7556145	Australia	99,51	0,08	2738500	38	669111	797612	32,7	2547
ERS5312789	HE	45	SAMEA7556147	Australia	99,51	0,22	2777516	48	202385	350397	32,7	2581
ERS5312790	HE	45	SAMEA7556148	Australia	99,51	0,08	2752162	38	208454	626118	32,7	2546
ERS5312791	HE	45	SAMEA7556149	Australia	99,51	0,08	2811418	49	220059	855403	32,7	2625
ERS5312796	HE	45	SAMEA7556154	Australia	99,51	0,08	2775683	55	243724	419423	32,7	2589
ERS5312797	HE	72	SAMEA7556155	Australia	99,51	0,08	2744831	63	183846	274945	32,7	2561
ERS5312801	HE	72	SAMEA7556159	Australia	99,51	0,08	2718970	39	430072	797385	32,7	2531
ERS5312804	HE	72	SAMEA7556162	Australia	99,51	0,08	2780671	28	1074519	1351692	32,7	2537
ERS5312805	HE	72	SAMEA7556163	Australia	99,42	0,08	2770986	46	243732	730784	32,7	2573
ERS5312807	HE	88	SAMEA7556165	Australia	99,51	0,08	2776254	57	221842	341828	32,7	2586
ERS5312810	HE	88	SAMEA7556168	Australia	99,51	0,08	2789517	63	232120	475479	32,7	2613
ERS5312811	HE	88	SAMEA7556169	Australia	99,51	0,08	2749438	36	764296	855028	32,8	2544
ERS5312812	HE	93	SAMEA7556170	Australia	99,51	0,08	2729628	43	469681	960313	32,7	2520
ERS5312813	HE	93	SAMEA7556171	Australia	99,51	0,1	2733678	51	975648	1299926	32,7	2537
ERS5312815	HE	93	SAMEA7556173	Australia	99,51	0,08	2729157	36	975467	1299926	32,7	2519
ERS5312817	HE	93	SAMEA7556175	Australia	99,51	0,08	2728658	36	638877	975548	32,7	2525
ERS5312818	HE	93	SAMEA7556176	Australia	99,51	0,08	2722051	30	313861	596758	32,7	2524
ERS5312820	HE	97	SAMEA7556178	Australia	99,51	0,08	2728721	41	1023815	1348561	32,7	2516
ERS5312827	HE	101	SAMEA7556185	Australia	99,51	0,08	2732014	35	215329	430603	32,8	2530
ERS5312830	HE	101	SAMEA7556188	Australia	99,51	0,11	2784501	53	222767	320642	32,7	2603
ERS5312832	HE	101	SAMEA7556190	Australia	99,51	0,22	2849891	45	191563	379960	32,7	2697
ERS5312836	HE	182	SAMEA7556194	Australia	99,51	0,22	2810634	67	644835	820846	32,7	2621
ERS5312837	HE	188	SAMEA7556195	Australia	99,51	0,08	2774763	31	290124	545786	32,7	2626
ERS5312842	HE	239	SAMEA7556200	Australia	99,51	0,36	2769711	54	619583	1270293	32,3	2575
ERS5312844	HE	239	SAMEA7556202	Australia	99,51	0,08	2786099	41	604449	1292048	32,7	2551
ERS5312847	HE	398	SAMEA7556205	Australia	99,65	0,36	2744542	24	1812559	1812559	32,2	2539
ERS5312849	HE	508	SAMEA7556207	Australia	99,51	0,22	2809489	76	644816	820729	32,7	2628
ERS5312854	HE	582	SAMEA7556212	Australia	99,51	0,08	2717842	46	216851	325839	32,7	2520
ERS5312858	HE	582	SAMEA7556216	Australia	99,32	0,08	2767030	23	617984	985787	32,7	2542
ERS5312862	HE	582	SAMEA7556220	Australia	99,51	0,08	2720386	40	233442	609698	32,7	2508

ERS5312863	HE	730	SAMEA7556221	Australia	99,51	0,08	2730783	56	211241	337220	32,7	2531
ERS5312865	HE	730	SAMEA7556223	Australia	99,51	0,08	2721282	43	210930	401234	32,7	2511
ERS5312874	HE	762	SAMEA7556232	Australia	99,51	0,08	2753460	34	272800	540864	32,7	2530
ERS5312875	HE	762	SAMEA7556233	Australia	99,51	0,08	2728282	28	328744	618475	32,7	2495
ERS5312876	HE	762	SAMEA7556234	Australia	99,32	0,08	2767690	41	255225	617208	32,7	2552
ERS5312878	HE	789	SAMEA7556236	Australia	99,51	0,08	2808235	32	184852	711594	32,7	2619
ERS5312880	HE	953	SAMEA7556238	Australia	99,51	0,08	2738747	43	214772	342203	32,7	2533
ERS5312884	HE	953	SAMEA7556242	Australia	99,51	0,11	2773920	44	205945	559020	32,6	2564
ERS5312886	HE	1159	SAMEA7556244	Australia	99,51	0,11	2828016	36	238839	344833	32,7	2661
ERS5312887	HE	1223	SAMEA7556245	Australia	99,51	0,13	2711899	30	998531	1023462	32,8	2511
ERS5312889	HE	1850	SAMEA7556247	Australia	99,51	0,22	2850978	52	262042	479790	32,7	2699
ERS5312890	HE	2889	SAMEA7556248	Australia	99,51	0,08	2692800	38	702135	885263	32,8	2465
GCA_000149015.1	HE	2889	SAMN00001489	not determined	99,51	0,08	2818925	5	2786578	2786578	32,7	2590
GCA_000159555.1	HE	3474	SAMN00001490	not determined	99,41	0,35	2779997	248	418208	690422	33	2615
GCA_000153665.1	HE	2889	SAMN00001491	not determined	98,58	0,32	2845480	174	1018247	1398786	32,9	2691
GCA_000159535.2	HE	2889	SAMN00002240	not determined	99,51	0,25	2827166	2	2802675	2802675	32,8	2614
GCA_000187165.1	HE	4618	SAMN00189157	not determined	99,51	0,08	2826042	47	143482	312871	32,6	2645
GCA_003017355.1	HE	5189	SAMN06835230	USA	99,51	0,69	2747755	3	2700728	2700728	32,8	2552
GCA_003017535.1	HE	5189	SAMN06835231	USA	99,51	0,25	2943092	2	2927335	2927335	32,8	2780
GCA_003017495.1	HE	5189	SAMN06835232	USA	99,51	0,1	2830166	2	2800572	2800572	32,8	2627
GCA_003017305.1	HE	5110	SAMN06835234	USA	99,13	1,04	2979281	2	2965341	2965341	32,8	2828
GCA_004012035.1	HE	-	SAMN10394659	South Korea	99,51	0,19	2827757	74	121200	331117	32,8	2647
GCA_003989655.1	HE	5189	SAMN10639319	Malaysia	99,51	0,22	3107513	114	152788	586667	32,8	3021
GCA_003989815.1	HE	-	SAMN10639402	Malaysia	99,51	0,08	3035495	131	142911	567060	32,7	2989
GCA_004026165.1	HE	-	SAMN10659762	South Korea	99,13	0,1	2815599	2	2796411	2796411	32,9	2586
GCA_014436675.1	HE	-	SAMN15660211	France	99,51	0,13	2694498	25	607734	980109	32,8	2496
GCA_022226975.1	HE	-	SAMN25144898	USA	99,51	0,13	2700946	1	2700946	2700946	32,9	2479
GCA_022226995.1	HE	-	SAMN25144899	USA	99,51	0,64	2833317	3	2817030	2817030	32,8	2615

Table S2. Metadata of the 48 *S. aureus* genome assemblies from Augsburg, Germany

Strain	Assembly accession	BioSample accession	Genome size (bp)	# of scaffolds	N50	GC %	# coding genes	Chromosome	Contamination %	# realigned subreads	Average subread length (bp)	Skin location	Health status[a]
194	CP077932	SAMN19550019	2754399	1	2754399	33	2562	Complete	0.13	120834	4943	Skin	AD
195	CP077931	SAMN19550020	2754414	1	2754414	33	2556	Complete	0.13	120639	4884		
196	CP077928 CP077929 CP077930	SAMN19550021	2828046	3	2754404	33	2642	Complete	0.97	98907	5164		
197	CP077926 CP077927	SAMN19550022	2788328	2	2754402	33	2573	Complete	0.13	135207	4965	Nose	
198	CP077925	SAMN19550023	2754401	1	2754401	33	2564	Complete	0.13	138422	4638		
199	CP077939 CP077940	SAMN19550024	2793031	2	2764397	33	2585	Draft	1.54	179644	5099		
200	CP077924	SAMN19550025	2719166	1	2719166	33	2500	Complete	0.08	169836	4607	Nose	HE
201	CP077923	SAMN19550026	2719235	1	2719235	33	2502	Complete	0.08	206880	5115		
202	CP077922	SAMN19550027	2719233	1	2719233	33	2505	Complete	0.08	196190	5243		
276	CP077921	SAMN19550028	2730804	1	2730804	32.9	2537	Complete	0.08	281434	5163	Skin	AD
277	CP077920	SAMN19550029	2749120	1	2749120	32.9	2546	Complete	0.11	181163	4818		
278	CP077918 CP077919	SAMN19550030	2785715	2	2732781	32.8	2585	Complete	1.77	178032	5263		
280	CP077917	SAMN19550031	2748827	1	2748827	32.9	2548	Complete	0.11	95473	5487	Nose	
281	CP077916	SAMN19550032	2748821	1	2748821	32.9	2545	Complete	0.11	200607	5157		
282	CP077915	SAMN19550033	2732780	1	2732780	32.9	2539	Complete	0.08	152840	4691		
317	CP077913 CP077914	SAMN19550034	2955403	2	2911834	32.8	2767	Draft	0.25	232973	5074	Nose	HE
318	CP077911 CP077912	SAMN19550035	2952667	2	2919985	32.8	2740	Complete	2.5	209420	5129		
319	CP077936 CP077937 CP077938	SAMN19550036	2952363	3	2890321	32.8	2754	Complete	0.62	259002	5765		
322	CP077908 CP077909 CP077910	SAMN19550037	2831656	3	2764234	32.8	2646	Complete	0.13	316777	4789	Nose	HE
323	CP077906 CP077907	SAMN19550038	2873158	2	2828011	32.9	2644	Draft	3.25	497238	5126		
324	CP077899 CP077900 CP077901 CP077902 CP077903 CP077904 CP077905	SAMN19550039	3123465	7	2816989	32.8	2818	Draft	4.15	342044	5027		

327	CP077898	SAMN19550040	2764236	1	2764236	33	2594	Complete	0.13	169770	5122		
328	CP077897	SAMN19550041	2764234	1	2764234	33	2594	Complete	0.13	243189	5188		
329	CP077894 CP077895 CP077896	SAMN19550042	2877025	3	2795787	32.9	2659	Complete	0.78	158893	5223	Nose	HE
332	CP077893	SAMN19550043	2745575	1	2745575	32.8	2579	Complete	0.08	126671	5338		
333	CP077889 CP077890 CP077891 CP077892	SAMN19550044	2840121	4	2745576	32.8	2621	Complete	3.45	195715	4983	Nose	HE
334	CP077888	SAMN19550045	2745573	1	2745573	32.8	2584	Complete	0.08	156913	5237		
355	CP077885 CP077886 CP077887	SAMN19550046	2912691	3	2855354	32.7	2781	Complete	0.17	147539	4626		
356	CP077880 CP077881 CP077882 CP077883 CP077884	SAMN19550047	2930327	5	2855358	32.7	2799	Complete	0.16	169612	3707	Nose	
357	CP077877 CP077878 CP077879	SAMN19550048	2898286	3	2855353	32.7	2770	Complete	0.12		4959		AD
358	CP077872 CP077873 CP077874 CP077875 CP077876	SAMN19550049	2926152	5	2855863	32.7	2790	Complete	0.16	497238	4670		
359	CP077933 CP077934 CP077935	SAMN19550050	2884333	3	2599728	32.7	2751	Complete	0.12	95473	4369	Skin	
360	CP077870 CP077871	SAMN19550051	2902129	2	2855872	32.8	2768	Complete	0.08	120834	4904		
365	CP077869	SAMN19550052	2757494	1	2757494	32.9	2553	Complete	0.22	120639	4256		
366	CP077868	SAMN19550053	2757281	1	2757281	32.9	2552	Complete	0.22	98907	4611	Nose	
367	CP077867	SAMN19550054	2786390	1	2786390	32.9	2571	Complete	0.78	135207	5389		
370	CP077865 CP077866	SAMN19550055	2838117	2	2756732	32.9	2609	Complete	4.05	138422	5100		AD
371	CP077863 CP077864	SAMN19550056	2793848	2	2757339	32.9	2566	Complete	0.22	179644	5129	Skin	
372	CP077861 CP077862	SAMN19550057	2798879	2	2757024	32.9	2576	Complete	1.91	169836	5166		
AC4	CP077860	SAMN19550058	2766976	1	2766976	32.9	2599	Complete	0.08	206880	5165		
AC5	CP077859	SAMN19550059	2766976	1	2766976	32.9	2599	Complete	0.08	196190	5092		
AC6	CP077858	SAMN19550060	2766990	1	2766990	32.9	2597	Complete	0.08	281434	4691	Non-lesional skin	AD

L4	CP077857	SAMN19550061	2766975	1	2766975	32.9	2599	Complete	0.08	181163	4842	Lesional skin	AD
L5	CP077856	SAMN19550062	2766975	1	2766975	32.9	2597	Complete	0.08	178032	4834		
L6	CP077855	SAMN19550063	2766978	1	2766978	32.9	2597	Complete	0.08	95473	5045		
N4	CP077854	SAMN19550064	2766976	1	2766976	32.9	2598	Complete	0.08	200607	4866	Nose	
N5	CP077853	SAMN19550065	2766976	1	2766976	32.9	2599	Complete	0.08	152840	5114		
N6	CP077852	SAMN19550066	2766978	1	2766978	32.9	2599	Complete	0.08	232973	4993		

Table S3. Metadata of *S. aureus* strains from animals

Assembly Accession	CheckM Completeness (%)	CheckM Contamination (%)	Genome size (bp)	# contigs	N50 (contigs)	GC (%)	# predicted genes	Host name
GCA_000276625.1	99.45	0.08	2790393	28	503158	32.8	2599	Cow
GCA_000401595.1	96.17	0.22	2772919	78	83839	32.8	2622	Cow
GCA_000409145.1	99.51	0.08	2844315	118	167283	32.8	2733	Cow
GCA_000409165.1	99.51	0.08	2808519	80	130864	32.7	2675	Cow
GCA_000636875.1	99.51	0.08	2704926	46	105744	32.7	2524	Cow
GCA_000636955.1	0.00	0.00	2798615	54	105192	32.9	2601	Cow
GCA_000637035.1	99.41	0.08	2747543	59	96681	32.7	2526	Cow
GCA_000637335.1	99.51	0.17	2806953	49	138708	32.9	2627	Cow
GCA_000637375.1	0.00	0.00	2867623	58	153428	32.9	2678	Cow
GCA_000637435.1	95.98	0.22	2792847	52	135022	32.9	2615	Cow
GCA_000637495.1	96.06	0.22	2741455	37	150042	32.9	2521	Cow
GCA_000637515.1	0.00	0.00	2747143	37	186015	32.9	2528	Cow
GCA_000637535.1	0.00	0.00	2742959	38	137793	32.9	2530	Cow
GCA_000637555.1	94.95	0.78	2744844	38	152418	32.9	2527	Cow
GCA_000637575.1	99.26	0.08	2741716	31	149804	32.9	2518	Cow
GCA_000637595.1	0.00	0.00	2750424	43	151678	32.9	2529	Cow
GCA_000637615.1	0.00	0.00	2743229	33	178030	32.9	2519	Cow
GCA_000637995.1	99.51	0.19	2851162	60	152568	32.7	2650	Cow
GCA_000638015.1	0.00	0.00	2775755	36	217984	32.8	2543	Cow
GCA_000638035.1	96.33	0.41	2733651	37	188509	32.9	2494	Cow
GCA_000638055.1	99.51	0.11	2804875	48	135022	32.8	2576	Cow
GCA_000638075.1	0.00	0.00	2824119	35	194453	32.9	2611	Cow
GCA_000638095.1	0.00	0.00	2809674	39	240221	32.9	2610	Cow
GCA_000638275.1	0.00	0.00	2769356	33	201379	32.9	2561	Cow
GCA_000638295.1	96.95	0.30	2744372	48	113289	32.9	2542	Cow
GCA_000638315.1	0.00	0.00	2774241	45	137312	32.8	2537	Cow
GCA_000638335.1	0.00	0.00	2768702	39	147726	32.9	2549	Cow
GCA_000638355.1	0.00	0.00	2789939	33	235979	32.9	2586	Cow
GCA_000639015.1	96.17	0.25	2757818	39	181186	32.9	2526	Cow
GCA_000639035.1	99.51	0.08	2801584	38	180089	32.9	2585	Cow
GCA_000639055.1	0.00	0.00	2796572	45	170841	32.9	2576	Cow
GCA_000639155.1	89.73	0.85	2755919	33	208532	32.8	2539	Cow
GCA_000639175.1	0.00	0.00	2799615	38	202448	32.8	2600	Cow
GCA_000639195.1	93.81	0.81	2772669	44	170130	32.9	2567	Cow
GCA_000639275.1	99.51	0.13	2770137	25	299646	32.8	2573	Cow
GCA_000639315.1	0.00	0.00	2711411	43	148109	32.8	2486	Cow
GCA_000639335.1	86.03	0.89	2772811	39	215792	32.9	2557	Cow
GCA_000639755.1	82.50	0.39	2756995	45	106127	32.8	2546	Cow
GCA_000639775.1	0.00	0.00	2773355	31	201407	32.8	2570	Cow
GCA_000639795.1	87.90	0.78	2798345	33	203887	32.9	2588	Cow
GCA_000639815.1	0.00	0.00	2873820	48	179467	32.9	2713	Cow
GCA_000639835.1	0.00	0.00	2777201	32	215050	32.8	2578	Cow

GCA_000639855.1	84.00	0.58	2854709	34	210455	32.9	2683	Cow
GCA_000639875.1	99.51	0.25	2877449	51	152284	32.9	2694	Cow
GCA_000684415.1	99.51	0.08	2774273	25	270232	32.8	2569	Cow
GCA_000684635.1	99.51	0.08	2806388	42	393161	32.9	2625	Cow
GCA_000949555.1	98.83	0.25	2650629	26	250418	32.6	2506	Cow
GCA_000949565.1	99.46	0.08	2699501	31	184423	32.7	2475	Cow
GCA_000949575.1	98.83	0.22	2660305	25	344782	32.6	2517	Cow
GCA_000949585.1	98.83	0.22	2646563	29	165590	32.6	2499	Cow
GCA_001051265.1	99.45	3.66	2838583	87	644842	32.5	2864	Cow
GCA_001196615.1	99.41	0.24	2752300	24	288781	32.7	2526	Cow
GCA_001278865.1	92.44	11.72	3129233	514	21192	32.9	5237	Cow
GCA_001298325.1	99.41	0.35	2805554	56	194079	32.7	2637	Cow
GCA_001469015.1	98.45	0.08	2682322	92	48414	32.7	2544	Cow
GCA_001469065.1	95.17	1.52	2509504	287	17721	32.7	2544	Cow
GCA_001469085.1	96.49	0.75	2579334	568	8606	32.9	2660	Cow
GCA_001469105.1	95.54	0.91	2622746	194	33040	32.8	2597	Cow
GCA_001636535.1	99.51	0.08	2692988	133	44880	32.6	2483	Cow
GCA_001639755.1	99.12	0.08	2720879	144	40637	32.8	2575	Cow
GCA_001639785.1	98.32	0.22	2648896	104	51869	32.7	2520	Cow
GCA_001639795.1	97.31	0.22	2551190	213	23221	33.0	2475	Cow
GCA_001639825.1	98.13	0.29	2642470	76	79315	32.8	2479	Cow
GCA_001639835.1	98.95	0.08	2604095	107	43187	32.8	2413	Cow
GCA_001639865.1	99.51	0.08	2779913	70	91068	32.6	2583	Cow
GCA_001639885.1	99.31	0.13	2729672	92	52360	32.7	2564	Cow
GCA_001680925.1	97.80	0.15	2699299	84	64854	32.7	2525	Cow
GCA_001680935.1	0.00	0.00	2677396	23	175785	32.7	2439	Cow
GCA_001680995.1	88.44	0.78	2710669	35	192514	32.7	2503	Cow
GCA_001681005.1	77.12	0.76	2738044	35	202483	32.7	2526	Cow
GCA_001681075.1	73.73	0.76	2696221	28	294781	32.7	2569	Cow
GCA_001681085.1	60.02	2.22	2693244	161	32786	32.7	2534	Cow
GCA_001681155.1	27.96	4.34	2765013	32	130009	32.8	2568	Cow
GCA_001681175.1	98.83	1.06	2699779	171	31291	32.7	2605	Cow
GCA_001855505.1	99.41	0.08	2731946	44	73607	32.7	2541	Cow
GCA_001982965.1	99.51	0.08	2767852	48	509023	32.8	2600	Cow
GCA_001982975.1	99.51	0.08	2779589	67	666472	32.8	2610	Cow
GCA_001983025.1	99.51	0.08	2678585	97	63940	32.8	2496	Cow
GCA_001983065.1	99.51	0.72	2703795	152	48646	32.8	2549	Cow
GCA_001983085.1	99.51	0.27	2694043	154	52605	32.8	2539	Cow
GCA_001983095.1	99.51	0.11	2734678	202	30007	32.8	2596	Cow
GCA_001983125.1	99.51	0.08	2754418	37	668039	32.7	2584	Cow
GCA_002099175.1	99.51	0.08	2762421	145	34448	32.8	2595	Cow
GCA_002374075.1	98.85	0.19	2789721	29	254979	32.8	2656	Cow
GCA_002761175.1	98.24	0.08	2732925	56	73034	32.7	2906	Cow
GCA_002810375.1	96.51	0.92	2714821	53	75406	32.7	3150	Cow
GCA_002810415.1	97.94	0.17	2719190	51	90795	32.8	2948	Cow
GCA_002835445.1	98.60	0.10	2691752	54	92538	32.7	2789	Cow
GCA_002849265.1	97.11	0.21	2709990	59	92539	32.7	3070	Cow
GCA_002952015.1	99.40	0.08	2718638	1	2718638	32.7	2533	Cow

GCA_003030225.1	99.51	0.64	2719423	1	2719423	32.9	2463	Cow
GCA_003131045.1	99.36	8.09	3029009	23	500103	33.0	2812	Cow
GCA_003131055.1	99.36	0.22	2713523	18	424315	32.7	2549	Cow
GCA_003131115.1	98.83	0.22	2754640	46	251215	32.7	2643	Cow
GCA_003131125.1	99.36	0.22	2714159	18	424276	32.7	2548	Cow
GCA_003131165.1	98.95	0.08	2707119	37	209358	32.8	2528	Cow
GCA_003131225.1	98.83	0.22	2695752	29	288500	32.7	2546	Cow
GCA_003133785.1	98.83	0.22	2697903	34	288552	32.7	2557	Cow
GCA_003186105.1	99.51	0.08	2808798	1	2808798	32.8	2588	Cow
GCA_003186125.1	98.95	0.08	2876138	2	2844513	32.8	2703	Cow
GCA_003721115.1	60.59	2.44	2846305	70	121591	32.7	2672	Cow
GCA_003850765.1	30.30	4.51	2612145	24	232672	32.7	2506	Cow
GCA_003952625.1	99.46	1.02	2876590	1	2876590	32.9	2660	Cow
GCA_003957315.1	99.46	1.20	2775972	1	2775972	32.9	2530	Cow
GCA_004803435.1	98.85	0.08	2706407	35	372386	32.8	2525	Cow
GCA_005153305.1	99.51	0.08	2736781	18	293330	32.7	2542	Cow
GCA_006386445.1	99.51	0.08	2764278	24	553168	32.6	2527	Cow
GCA_006386475.1	99.51	0.08	2811282	25	737913	32.6	2605	Cow
GCA_006511305.1	98.83	0.22	2692063	29	332638	32.7	2557	Cow
GCA_006511315.1	99.37	0.08	2767107	36	199379	32.6	2538	Cow
GCA_006511345.1	99.51	0.08	2764300	26	553168	32.6	2529	Cow
GCA_006511365.1	98.83	0.22	2683714	24	222787	32.7	2554	Cow
GCA_006511395.1	99.51	0.08	2871522	99	149345	32.7	2685	Cow
GCA_006511405.1	98.83	0.22	2727334	28	152898	32.7	2614	Cow
GCA_006511435.1	99.51	0.22	2741642	24	504217	32.6	2572	Cow
GCA_006511445.1	99.51	0.40	2761938	172	56216	32.7	2674	Cow
GCA_006511475.1	99.49	0.08	2772469	32	276209	32.8	2592	Cow
GCA_006511485.1	99.51	0.08	2778733	26	697818	32.6	2545	Cow
GCA_006511515.1	99.23	0.08	2703316	39	651423	32.6	2500	Cow
GCA_006511525.1	99.23	0.11	2795859	23	549659	32.6	2584	Cow
GCA_006511575.1	99.51	0.11	2797370	24	319673	32.7	2604	Cow
GCA_006511585.1	98.95	0.08	2735858	28	411256	32.7	2552	Cow
GCA_006511605.1	99.51	0.50	2753887	78	194119	32.9	2548	Cow
GCA_006511625.1	99.51	0.10	2776122	65	122268	32.7	2588	Cow
GCA_006511635.1	99.48	0.57	2665745	96	70541	32.8	2468	Cow
GCA_006511665.1	98.83	0.61	2712902	88	83639	32.7	2624	Cow
GCA_006511685.1	99.51	0.08	2764300	26	553168	32.6	2529	Cow
GCA_006511695.1	99.51	0.08	2766245	24	698581	32.6	2532	Cow
GCA_006511735.1	81.39	0.60	2697325	26	231196	32.7	2571	Cow
GCA_006511755.1	49.52	2.93	2741651	22	504299	32.6	2571	Cow
GCA_006511765.1	22.52	0.54	2797370	24	319673	32.7	2604	Cow
GCA_006511775.1	99.51	0.44	2845788	99	144788	32.8	2666	Cow
GCA_006511805.1	98.95	0.35	2737662	65	143063	32.8	2633	Cow
GCA_009696725.1	99.51	0.08	2744089	34	203653	32.7	2552	Cow
GCA_009696735.1	99.51	0.08	2756061	40	366440	32.7	2563	Cow
GCA_009696765.1	99.51	0.08	2759556	37	366440	32.7	2564	Cow
GCA_009696775.1	99.56	0.22	2713317	34	232744	32.7	2559	Cow
GCA_009696785.1	99.51	0.22	2722478	49	148482	32.7	2516	Cow

GCA_009696825.1	99.51	0.08	2742906	37	236144	32.7	2534	Cow
GCA_009696835.1	99.51	0.22	2722483	52	148482	32.7	2516	Cow
GCA_009696865.1	99.51	0.08	2724026	75	104687	32.8	2543	Cow
GCA_009696875.1	99.51	0.08	2671502	43	200216	32.7	2459	Cow
GCA_009696905.1	99.51	0.08	2745631	69	89575	32.8	2577	Cow
GCA_009696915.1	99.51	0.08	2670471	42	200214	32.7	2459	Cow
GCA_009696935.1	99.51	0.08	2747437	53	123287	32.7	2567	Cow
GCA_009696965.1	99.51	0.08	2747294	48	123103	32.7	2567	Cow
GCA_009696985.1	99.51	0.08	2748054	47	123287	32.7	2568	Cow
GCA_009697005.1	99.51	0.08	2746928	50	123287	32.7	2566	Cow
GCA_009697015.1	99.51	0.08	2751168	53	124078	32.7	2574	Cow
GCA_009697035.1	99.51	0.08	2746472	48	123425	32.7	2564	Cow
GCA_009697055.1	99.51	0.08	2748168	47	127893	32.7	2567	Cow
GCA_009697085.1	99.51	0.08	2820150	55	154453	32.7	2648	Cow
GCA_009763195.1	98.83	0.81	2793567	37	644499	32.8	2652	Cow
GCA_010994155.1	99.51	0.27	2783281	44	346241	32.8	2593	Cow
GCA_010994215.1	99.51	0.27	2783872	45	346241	32.8	2602	Cow
GCA_011007075.1	99.51	1.20	2832820	54	101417	32.6	2684	Cow
GCA_011007115.1	99.15	0.08	2881931	68	84977	32.7	2720	Cow
GCA_011007135.1	69.97	1.71	2912832	79	93111	32.8	2768	Cow
GCA_011007155.1	69.38	1.87	2809528	73	83453	32.7	2616	Cow
GCA_011007165.1	50.08	3.10	3066775	103	80817	32.7	2960	Cow
GCA_011007185.1	99.51	0.27	2885413	71	80043	32.7	2733	Cow
GCA_011007255.1	99.51	0.22	2766323	49	120057	32.7	2547	Cow
GCA_011007275.1	99.51	0.22	2798561	48	138211	32.8	2608	Cow
GCA_011032785.1	99.51	0.24	2777910	2	2752113	32.9	2557	Cow
GCA_012044215.1	98.91	0.63	2741492	1	2741492	32.9	2501	Cow
GCA_012044475.1	99.51	0.45	2744920	1	2744920	32.9	2489	Cow
GCA_012277065.1	99.51	0.08	2720045	44	182134	32.6	2573	Cow
GCA_013303105.1	98.85	0.64	2719833	20	361239	32.7	2527	Cow
GCA_013413435.1	99.49	0.11	2755333	36	285081	32.8	2549	Cow
GCA_013413445.1	99.49	0.11	2754752	35	241126	32.8	2549	Cow
GCA_015224275.1	99.51	0.08	2750508	8	688037	32.7	2563	Cow
GCA_015277615.1	99.51	0.27	2612544	5	1079443	32.8	2494	Cow
GCA_015277655.1	99.51	0.08	2799428	22	433254	32.6	2595	Cow
GCA_015277665.1	99.51	0.08	2801855	18	433254	32.6	2602	Cow
GCA_015277695.1	99.51	0.08	2735791	16	436719	32.7	2538	Cow
GCA_016617525.1	98.28	17.24	3107556	76	98412	33.0	3004	Cow
GCA_016888225.1	99.51	0.11	2739072	66	144567	32.7	2538	Cow
GCA_016888235.1	99.51	0.22	2852604	72	130749	32.7	2709	Cow
GCA_017328685.1	99.51	0.08	2769458	1	2769358	32.6	2556	Cow
GCA_017328705.1	99.51	0.08	2754973	1	2754873	32.6	2546	Cow
GCA_017328725.1	99.51	0.08	2751525	1	2751425	32.6	2538	Cow
GCA_017328745.1	99.51	0.08	2681229	1	2681129	32.6	2474	Cow
GCA_017328765.1	99.51	0.08	2768002	1	2767902	32.6	2557	Cow
GCA_017328785.1	99.51	0.08	2730444	1	2730344	32.6	2494	Cow
GCA_017328805.1	99.51	0.08	2681263	1	2681163	32.6	2474	Cow
GCA_017328825.1	99.37	0.08	2732917	1	2732817	32.6	2504	Cow

GCA_017328845.1	99.23	0.08	2789839	1	2789739	32.8	2550	Cow
GCA_017328865.1	99.37	0.08	2737381	1	2737281	32.6	2506	Cow
GCA_017328885.1	99.51	0.08	2794579	1	2794479	32.7	2604	Cow
GCA_017328905.1	26.23	0.27	2793667	1	2793567	32.7	2606	Cow
GCA_017328925.1	99.51	0.08	2678982	1	2678882	32.6	2467	Cow
GCA_017328945.1	99.51	0.08	2746167	1	2746067	32.6	2547	Cow
GCA_017328965.1	99.51	0.08	2730577	1	2730477	32.6	2498	Cow
GCA_017328985.1	99.51	0.08	2798511	1	2798411	32.7	2602	Cow
GCA_017329005.1	99.51	0.08	2806436	1	2806336	32.7	2606	Cow
GCA_017329025.1	99.51	0.08	2717720	1	2717620	32.6	2487	Cow
GCA_017329165.1	99.51	0.08	2762786	1	2762686	32.6	2557	Cow
GCA_017329185.1	99.51	0.08	2713113	1	2713013	32.6	2487	Cow
GCA_017329205.1	99.01	0.56	2758836	1	2758736	32.7	2605	Cow
GCA_017329225.1	99.43	0.08	2712686	1	2712586	32.6	2483	Cow
GCA_017329245.1	99.51	0.08	2769120	1	2769020	32.6	2559	Cow
GCA_017329265.1	99.51	0.08	2770194	1	2770094	32.6	2557	Cow
GCA_017329285.1	99.51	0.08	2763396	1	2763296	32.6	2549	Cow
GCA_017329305.1	99.51	0.08	2769460	1	2769360	32.6	2557	Cow
GCA_017329325.1	99.51	0.08	2772765	1	2772665	32.6	2566	Cow
GCA_017329345.1	99.51	0.08	2754316	1	2754216	32.6	2543	Cow
GCA_017329365.1	99.51	0.08	2769333	1	2769233	32.6	2561	Cow
GCA_017329385.1	99.13	0.08	2758059	1	2757959	32.7	2606	Cow
GCA_017329405.1	99.47	0.08	2758027	1	2757927	32.7	2550	Cow
GCA_017329425.1	99.51	0.08	2713974	1	2713874	32.6	2490	Cow
GCA_017329445.1	99.18	0.08	2755378	1	2755278	32.7	2606	Cow
GCA_017329465.1	99.51	0.08	2770739	1	2770639	32.6	2556	Cow
GCA_017329485.1	99.51	0.08	2730510	1	2730410	32.6	2501	Cow
GCA_017329505.1	99.51	0.08	2732448	1	2732348	32.7	2537	Cow
GCA_017329525.1	99.51	0.11	2847634	1	2847534	32.7	2703	Cow
GCA_018408225.1	74.66	1.18	2719662	20	356410	32.8	2556	Cow
GCA_018967665.1	81.82	0.39	2650102	50	106859	32.6	2430	Cow
GCA_018967685.1	51.21	2.31	2708026	35	184122	32.7	2534	Cow
GCA_018967725.1	26.23	0.35	2709149	31	202951	32.7	2539	Cow
GCA_018967745.1	99.49	0.08	2706216	36	131975	32.7	2536	Cow
GCA_018967765.1	99.51	0.08	2736240	50	137632	32.6	2552	Cow
GCA_018967775.1	99.51	0.08	2700972	48	132641	32.7	2521	Cow
GCA_018967805.1	99.51	0.08	2752523	47	126106	32.7	2583	Cow
GCA_018967885.1	99.51	0.08	2705109	61	83578	32.7	2537	Cow
GCA_018967905.1	99.51	0.08	2712956	45	309172	32.7	2543	Cow
GCA_018967965.1	99.51	0.25	2732921	27	203353	32.7	2565	Cow
GCA_018967985.1	99.51	0.25	2761668	46	249508	32.8	2582	Cow
GCA_018968005.1	99.51	0.08	2661209	24	326361	32.7	2449	Cow
GCA_018968025.1	99.51	0.08	2751002	37	149626	32.7	2581	Cow
GCA_018986155.1	99.40	0.08	2793678	60	151257	32.7	2606	Cow
GCA_018986195.1	99.51	0.08	2816200	32	696284	32.8	2659	Cow
GCA_018986215.1	99.51	0.08	2762043	33	1385792	32.7	2560	Cow
GCA_018986275.1	99.51	0.08	2714012	35	307265	32.7	2509	Cow
GCA_018986295.1	99.51	0.08	2804606	47	314623	32.7	2606	Cow

GCA_018986315.1	99.51	0.08	2804671	50	254400	32.7	2614	Cow
GCA_018986555.1	99.51	0.08	2720711	54	107753	32.7	2563	Cow
GCA_018986575.1	99.40	0.08	2588690	53	108692	32.8	2373	Cow
GCA_018986585.1	99.51	0.08	2660311	42	140532	32.7	2442	Cow
GCA_018986645.1	99.40	0.08	2588690	53	108692	32.8	2373	Cow
GCA_018995185.1	99.51	0.22	2732334	39	170521	32.7	2540	Cow
GCA_018996965.1	99.51	0.08	2744933	35	247934	32.7	2572	Cow
GCA_018997005.1	99.51	0.64	2701845	91	57736	32.8	2554	Cow
GCA_018998425.1	99.51	0.08	2741306	56	143011	32.7	2580	Cow
GCA_018998485.1	99.51	0.08	2691062	35	234367	32.6	2491	Cow
GCA_018998505.1	75.99	0.57	2678282	62	84215	32.7	2502	Cow
GCA_018998525.1	75.34	0.76	2743501	36	269582	32.7	2564	Cow
GCA_018998585.1	72.96	0.67	2667869	59	75363	32.7	2510	Cow
GCA_018998685.1	72.27	1.88	2729986	47	187923	32.7	2566	Cow
GCA_018998695.1	99.51	0.08	2681455	45	121052	32.7	2491	Cow
GCA_019049915.1	98.83	0.22	2666410	38	251194	32.6	2514	Cow
GCA_019049955.1	98.83	0.22	2655900	30	288395	32.6	2498	Cow
GCA_019049965.1	99.51	0.08	2738660	28	438062	32.7	2540	Cow
GCA_019049995.1	99.51	0.08	2703234	36	229535	32.8	2522	Cow
GCA_019050015.1	99.51	0.08	2786984	33	213256	32.8	2619	Cow
GCA_019050035.1	99.51	0.08	2787153	34	213257	32.8	2622	Cow
GCA_019050055.1	98.83	0.22	2666423	39	345031	32.6	2518	Cow
GCA_019050065.1	99.51	0.08	2702838	38	234335	32.8	2521	Cow
GCA_019050095.1	98.83	0.22	2653199	28	345032	32.6	2491	Cow
GCA_019050115.1	99.51	0.22	2710124	43	279675	32.8	2512	Cow
GCA_019050145.1	98.83	0.22	2656985	35	251118	32.6	2505	Cow
GCA_019050155.1	99.41	0.22	2717344	17	537442	32.7	2541	Cow
GCA_019050185.1	98.65	0.22	2652518	29	299711	32.6	2485	Cow
GCA_019050205.1	99.51	0.08	2705237	32	290801	32.8	2520	Cow
GCA_019050225.1	99.51	0.08	2703458	37	231903	32.8	2496	Cow
GCA_019050245.1	98.95	0.08	2705272	33	229383	32.8	2520	Cow
GCA_019050255.1	98.65	0.22	2660487	40	288688	32.6	2510	Cow
GCA_019050285.1	99.51	0.22	2754248	38	414161	32.8	2576	Cow
GCA_019050305.1	99.32	0.08	2806664	39	727671	32.7	2633	Cow
GCA_019050345.1	99.51	0.08	2707016	36	299165	32.8	2514	Cow
GCA_019100485.1	99.26	1.12	2701093	87	51617	32.8	2558	Cow
GCA_019149025.1	99.51	0.09	2710702	132	41070	32.7	2585	Cow
GCA_019149105.1	88.68	0.78	2654333	40	198056	32.7	2434	Cow
GCA_019149125.1	0.00	0.00	2717042	33	301135	32.7	2534	Cow
GCA_019149145.1	82.25	0.37	2758596	65	106493	32.7	2598	Cow
GCA_019149185.1	99.51	0.08	2688864	50	112510	32.7	2506	Cow
GCA_019149225.1	61.94	2.36	2588690	53	108692	32.8	2373	Cow
GCA_019279675.1	53.50	2.28	2785076	14	434905	32.6	2574	Cow
GCA_019279695.1	99.51	0.08	2796432	17	433098	32.6	2599	Cow
GCA_019279705.1	99.51	0.08	2776137	14	433253	32.6	2573	Cow
GCA_019279725.1	99.51	0.08	2784339	14	511155	32.6	2577	Cow
GCA_019279755.1	99.51	0.08	2750208	6	1376894	32.7	2561	Cow
GCA_019349095.1	99.14	0.11	2882591	178	66767	32.9	2915	Cow

GCA_019349155.1	99.40	0.08	2821440	128	91501	32.9	2753	Cow
GCA_019784245.1	99.51	0.08	2796432	17	433098	32.6	2599	Cow
GCA_020072335.1	99.51	0.08	2717144	35	285392	32.7	2490	Cow
GCA_020073845.1	98.95	0.08	2705171	132	67369	32.7	2537	Cow
GCA_020073865.1	99.34	0.13	2757106	86	128765	32.7	2545	Cow
GCA_020073905.1	97.53	0.18	2792873	171	41506	32.5	2654	Cow
GCA_020073925.1	99.39	1.25	2887226	397	22505	32.7	2894	Cow
GCA_020073945.1	98.69	0.08	2667401	107	91045	32.7	2481	Cow
GCA_020073965.1	99.49	0.08	2791271	37	296698	32.6	2584	Cow
GCA_020073985.1	99.51	0.11	2764982	39	255387	32.7	2553	Cow
GCA_020073995.1	99.51	0.08	2747973	54	148230	32.6	2534	Cow
GCA_020074025.1	99.51	0.08	2778012	45	218556	32.6	2596	Cow
GCA_020074045.1	99.51	0.08	2772442	52	230966	32.7	2575	Cow
GCA_020074065.1	99.51	0.08	2699539	31	283740	32.6	2493	Cow
GCA_020074085.1	97.30	0.08	2706169	150	57502	32.7	2534	Cow
GCA_020074105.1	99.51	0.08	2739451	72	81678	32.9	2535	Cow
GCA_020074125.1	98.20	0.08	2659977	36	352566	32.7	2459	Cow
GCA_020074135.1	99.51	0.08	2728869	90	81280	32.9	2531	Cow
GCA_020074165.1	98.95	0.10	2792587	104	83551	32.7	2605	Cow
GCA_020074285.1	99.45	0.08	2880454	99	74085	32.8	2741	Cow
GCA_020074345.1	99.51	0.08	2786580	72	137024	32.8	2578	Cow
GCA_020074385.1	95.13	0.80	2805966	100	72297	32.7	2612	Cow
GCA_020074685.1	70.64	1.71	2741926	31	338236	32.9	2526	Cow
GCA_020074705.1	15.68	0.91	2660125	23	318518	32.7	2427	Cow
GCA_020074775.1	67.11	1.84	2775568	95	99620	32.6	2585	Cow
GCA_021384725.1	51.69	2.32	2789085	70	149833	32.8	2615	Cow
GCA_021384745.1	99.51	0.11	2804148	51	306642	32.7	2625	Cow
GCA_021384765.1	99.51	0.11	2832934	54	190473	32.7	2660	Cow
GCA_021384795.1	99.51	0.11	2881474	63	315742	32.7	2725	Cow
GCA_021384825.1	99.51	0.32	2945916	80	156740	32.7	2812	Cow
GCA_023383235.1	98.83	0.22	2744728	27	196932	32.7	2635	Cow
GCA_023383245.1	98.83	0.22	2704323	28	388036	32.7	2575	Cow
GCA_023914165.1	99.37	0.08	2676917	98	64289	32.7	2516	Cow
GCA_024178305.1	99.51	0.08	2777647	15	622344	32.6	2589	Cow
GCA_024178405.1	99.51	0.08	2737569	16	386823	32.7	2545	Cow
GCA_025559405.1	99.28	0.30	2796785	2	2757383	32.8	2630	Cow
GCA_025559465.1	99.30	0.30	2804001	2	2800218	32.8	2625	Cow
GCA_025559585.1	99.40	0.30	2792079	3	2783822	32.8	2610	Cow
GCA_025559645.1	99.40	0.30	2777142	1	2777142	32.8	2604	Cow
GCA_900097995.1	99.51	61.42	4502717	971	55007	32.7	4724	Cow
GCA_900098065.1	99.51	0.08	2746327	72	65181	32.6	2592	Cow
GCA_900098185.1	88.64	43.43	3410398	7974	466	32.7	8936	Cow
GCA_900098415.1	98.83	0.22	2729205	30	151382	32.7	2612	Cow
GCA_900098425.1	98.83	0.22	2756145	51	151087	32.7	2664	Cow
GCA_900098505.1	98.55	0.22	2727819	41	122776	32.7	2622	Cow
GCA_900098555.1	98.95	0.22	2748139	70	100758	32.8	2599	Cow
GCA_900457865.1	99.49	0.08	2771350	2	2768606	32.8	2532	Cow
GCA_900458085.1	99.45	0.13	2864503	3	2803632	32.8	2709	Cow

GCA_900458115.1	98.83	0.22	2701690	2	2693200	32.7	2519	Cow
GCA_900458235.1	99.51	0.18	2923343	2	2902232	32.9	2749	Cow
GCA_900458415.1	99.49	0.13	2802960	2	2799580	32.8	2584	Cow
GCA_900474535.1	98.80	0.22	2702202	1	2702202	32.8	2517	Cow
GCA_918417985.1	99.51	0.19	2827640	182	94357	32.9	2686	Cow
GCA_918418275.1	99.51	0.22	2809701	113	186239	32.8	2656	Cow
GCA_918429155.1	99.51	0.78	2799710	178	124221	32.6	2678	Cow
GCF_000637215.1	0.00	0.00	2770434	45	178917	32.9	2578	Dog
GCF_000637655.1	0.00	0.00	2737321	36	137848	32.9	2518	Dog
GCF_000638735.1	0.00	0.00	2765817	46	272019	32.9	2591	Dog
GCF_000638875.1	0.00	0.00	2709272	28	327440	32.8	2494	Dog
GCF_001191855.1	99.51	0.22	2816780	49	170565	32.7	2638	Dog
GCF_001192075.1	99.51	0.30	2826671	53	150854	32.7	2651	Dog
GCF_001192175.1	99.51	0.22	2776080	36	174144	32.7	2576	Dog
GCF_001193355.1	0.00	0.00	2758555	31	311125	32.7	2546	Dog
GCF_001194725.1	85.47	0.30	2827184	41	173961	32.7	2643	Dog
GCF_001195105.1	0.00	0.00	2763909	46	163552	32.7	2564	Dog
GCF_001195205.1	0.00	0.00	2800734	33	255498	32.7	2620	Dog
GCF_001196555.1	0.00	0.00	2753701	50	87739	32.7	2566	Dog
GCF_001196655.1	0.00	0.00	2798418	35	176891	32.7	2607	Dog
GCF_001196795.1	0.00	0.00	2800993	51	143517	32.7	2602	Dog
GCF_001197715.1	99.51	0.22	2800628	39	211660	32.7	2610	Dog
GCF_001207435.1	0.00	0.00	2818726	51	135327	32.7	2641	Dog
GCF_001208245.1	92.45	0.30	2788322	60	86955	32.7	2613	Dog
GCF_001208665.1	99.51	0.22	2807472	52	167715	32.7	2632	Dog
GCF_001208785.1	0.00	0.00	3077144	1312	11085	32.6	3781	Dog
GCF_001209365.1	99.51	0.22	2777305	36	175705	32.7	2594	Dog
GCF_001210805.1	0.00	0.00	2787947	40	150860	32.7	2586	Dog
GCF_001210925.1	99.51	0.22	2771159	52	108242	32.7	2593	Dog
GCF_001211025.1	92.45	0.30	2837327	60	174123	32.7	2687	Dog
GCF_001212545.1	0.00	0.00	2786744	53	109212	32.7	2588	Dog
GCF_001349375.1	99.51	0.22	2832318	45	170553	32.7	2654	Dog
GCF_001349395.1	99.32	0.22	2798977	39	158081	32.7	2618	Dog
GCF_001349415.1	0.00	0.00	2830456	52	146517	32.7	2657	Dog
GCF_001349435.1	99.51	0.22	2788595	51	170564	32.7	2591	Dog
GCF_001349495.1	99.51	0.22	2750553	39	174166	32.7	2550	Dog
GCF_001349515.1	99.51	0.22	2802267	34	170556	32.7	2620	Dog
GCF_001349535.1	99.51	0.22	2798111	39	134214	32.7	2611	Dog
GCF_001349555.1	0.00	0.00	2837836	46	211686	32.7	2658	Dog
GCF_001349575.1	0.00	0.00	2799976	40	125471	32.7	2617	Dog
GCF_001349595.1	0.00	0.00	2790851	41	170564	32.7	2583	Dog
GCF_001349615.1	0.00	0.00	2835501	58	170547	32.7	2670	Dog
GCF_001349655.1	0.00	0.00	2782849	52	119337	32.7	2601	Dog
GCF_001349675.1	81.71	0.34	2829369	44	147433	32.7	2650	Dog
GCF_001349695.1	99.51	0.22	2802913	33	172663	32.7	2608	Dog
GCF_001349715.1	0.00	0.00	2744133	43	148577	32.7	2544	Dog
GCF_001349735.1	89.32	0.30	2826833	52	146517	32.7	2648	Dog
GCF_001349755.1	0.00	0.00	2832904	43	174003	32.7	2657	Dog

GCF_001349815.1	0.00	0.00	2787004	44	127897	32.7	2610	Dog
GCF_001349855.1	80.60	0.76	2816127	60	77128	32.7	2649	Dog
GCF_001349915.1	0.00	0.00	2799420	59	83742	32.7	2639	Dog
GCF_001350015.1	0.00	0.00	2783329	32	174166	32.7	2568	Dog
GCF_002097265.1	0.00	0.00	2733550	223	37955	32.7	2592	Dog
GCF_002097305.1	92.81	0.30	2704944	157	27696	32.7	2529	Dog
GCF_002097385.1	95.46	0.36	2651292	314	14928	32.9	2556	Dog
GCF_002097525.1	85.27	0.08	2658526	361	12860	32.9	2576	Dog
GCF_002097615.1	80.04	0.86	2739784	150	36566	32.7	2611	Dog
GCF_002097655.1	0.00	0.00	2727529	175	30046	32.8	2573	Dog
GCF_002104815.1	83.30	0.43	2779668	216	51253	32.8	2668	Dog
GCF_003029665.1	0.00	0.00	2982183	4	1026003	33.0	2802	Dog
GCF_015209525.1	0.00	0.00	2803184	47	214221	32.8	2616	Dog
GCF_015209565.1	0.00	0.00	2804428	50	174229	32.8	2623	Dog
GCF_016092305.1	99.41	0.11	2731466	57	154077	32.7	2527	Dog
GCF_016092375.1	99.51	0.13	2748546	36	235895	32.8	2573	Dog
GCA_000638775.1	0.00	0.00	2796211	36	138500	32.9	2601	Horse
GCA_000638855.1	0.00	0.00	2782117	40	173263	32.9	2566	Horse
GCA_000684495.1	0.00	0.00	2779070	53	103620	32.9	2575	Horse
GCA_002097325.1	97.28	0.00	2740304	271	20783	32.8	2637	Horse
GCA_002097345.1	93.22	1.21	2571408	551	7404	33.0	2598	Horse
GCA_002097425.1	95.57	0.95	2618523	496	8575	32.9	2636	Horse
GCA_002097625.1	89.53	0.22	2571120	512	7747	33.0	2585	Horse
GCA_002097645.1	99.49	0.33	2735749	177	30597	32.8	2573	Horse
GCA_003001175.1	0.00	0.00	2797201	36	249644	32.8	2593	Horse
GCA_003001185.1	0.00	0.00	2841587	36	149998	32.9	2657	Horse
GCA_003001215.1	0.00	0.00	2842635	36	326322	32.9	2661	Horse
GCA_003001225.1	0.00	0.00	2797663	38	249644	32.8	2596	Horse
GCA_003001255.1	0.00	0.00	2797064	37	249644	32.8	2594	Horse
GCA_003001275.1	0.00	0.00	2797999	38	249713	32.8	2594	Horse
GCA_003001295.1	0.00	0.00	2856018	61	148498	32.9	2689	Horse
GCA_003001315.1	0.00	0.00	2852794	52	148498	32.9	2687	Horse
GCA_003001335.1	0.00	0.00	2855011	51	253843	32.9	2692	Horse
GCA_003001345.1	99.51	0.08	2865535	42	268003	32.9	2713	Horse
GCA_003001375.1	0.00	0.00	2849169	43	196999	32.8	2679	Horse
GCA_003001395.1	93.70	0.63	2818166	40	251448	32.9	2644	Horse
GCA_003001415.1	99.51	0.08	2849614	47	196999	32.8	2685	Horse
GCA_003001435.1	99.51	0.08	2849775	46	196999	32.8	2684	Horse
GCA_003001455.1	0.00	0.00	2850120	45	196999	32.8	2682	Horse
GCA_003001475.1	0.00	0.00	2843900	38	326322	32.9	2661	Horse
GCA_003001485.1	0.00	0.00	2794575	41	253302	32.9	2593	Horse
GCA_003057555.1	93.93	0.63	2814183	63	95080	32.9	2654	Horse
GCA_003327885.1	0.00	0.00	2816634	47	171625	32.8	2623	Horse
GCA_003328335.1	99.51	0.08	2837688	36	326321	32.9	2652	Horse
GCA_003328355.1	0.00	0.00	2816939	39	251449	32.9	2646	Horse
GCA_003328375.1	99.51	0.08	2837815	36	326321	32.9	2653	Horse
GCA_003328385.1	99.51	0.08	2792319	37	253302	32.8	2592	Horse
GCA_003328395.1	99.40	0.08	2837840	38	326322	32.9	2656	Horse

GCA_003328405.1	99.51	0.08	2793761	38	253302	32.9	2591	Horse
GCA_007185035.1	99.51	0.41	2957206	2	2905582	32.8	2751	Horse
GCA_008079925.1	98.81	0.08	2870030	105	87148	32.7	2743	Horse
GCA_008079965.1	98.81	0.08	2860551	134	62211	32.7	2746	Horse
GCA_008079975.1	98.81	0.08	2859175	148	47544	32.7	2756	Horse
GCA_008080005.1	98.81	0.08	2873555	105	101863	32.7	2750	Horse
GCA_008080015.1	98.81	0.08	2873960	138	83121	32.7	2764	Horse
GCA_008080045.1	98.81	0.08	2871605	111	68000	32.7	2744	Horse
GCA_008080055.1	98.73	0.08	2858730	169	43706	32.7	2767	Horse
GCA_008080075.1	96.15	0.52	2596831	633	7349	33.3	2722	Horse
GCA_008080105.1	98.81	0.08	2870998	105	87464	32.7	2740	Horse
GCA_008080115.1	98.81	0.08	2857725	181	49494	32.7	2777	Horse
GCA_008080135.1	98.81	0.08	2901472	125	85039	32.6	2784	Horse
GCA_008080175.1	98.81	0.08	2865129	139	53024	32.7	2764	Horse
GCA_008080185.1	98.62	0.08	2865927	132	55928	32.7	2748	Horse
GCA_008080215.1	90.94	1.77	2394797	914	4088	33.7	2659	Horse
GCA_008080225.1	74.78	0.91	2869436	125	69513	32.7	2756	Horse
GCA_008080235.1	0.00	0.00	2717550	461	11619	33.0	2749	Horse
GCA_008080255.1	77.97	0.93	2843707	182	42723	32.7	2744	Horse
GCA_008080265.1	0.00	0.00	2870253	109	87463	32.7	2742	Horse
GCA_008080305.1	0.00	0.00	2875554	104	86375	32.7	2748	Horse
GCA_008080325.1	0.00	0.00	2866589	139	63346	32.7	2758	Horse
GCA_008080335.1	79.77	2.02	2482827	774	5458	33.5	2670	Horse
GCA_008080355.1	0.00	0.00	2837661	230	32542	32.8	2775	Horse
GCA_008080375.1	98.22	0.27	2871407	118	82975	32.7	2753	Horse
GCA_008123345.1	95.35	0.73	2832264	125	49236	32.7	2670	Horse
GCA_008123355.1	0.00	0.00	2838369	89	79184	32.6	2675	Horse
GCA_008123575.1	0.00	0.00	2831919	76	77100	32.6	2663	Horse
GCA_008123595.1	0.00	0.00	2834377	103	51572	32.6	2662	Horse
GCA_008123625.1	79.17	1.12	2832886	102	101064	32.6	2659	Horse
GCA_008123635.1	0.00	0.00	2826315	119	58798	32.7	2669	Horse
GCA_008123655.1	73.86	1.08	2834074	103	49984	32.6	2662	Horse
GCA_008123665.1	31.67	2.09	2836155	89	80498	32.6	2657	Horse
GCA_008123695.1	0.00	0.00	2835256	54	141790	32.6	2644	Horse
GCA_008123715.1	0.00	0.00	2866870	129	73723	32.8	2746	Horse
GCA_008123735.1	0.00	0.00	2831380	92	64316	32.6	2653	Horse
GCA_008123765.1	0.00	0.00	2831522	66	111743	32.7	2654	Horse
GCA_008123775.1	0.00	0.00	2827287	93	65509	32.7	2661	Horse
GCA_008123805.1	99.51	0.38	2844326	110	70275	32.7	2686	Horse
GCA_008123815.1	99.51	0.22	2828043	119	51856	32.7	2663	Horse
GCA_008123845.1	99.51	0.25	2838053	69	121077	32.6	2666	Horse
GCA_008123865.1	99.51	0.27	2831484	129	40948	32.7	2670	Horse
GCA_008123885.1	0.00	0.00	2856063	238	25829	32.8	2778	Horse
GCA_008123905.1	0.00	0.00	2870228	115	99667	32.8	2753	Horse
GCA_010363625.1	0.00	0.00	2853262	147	49482	32.8	2700	Horse
GCA_010363635.1	0.00	0.00	2855351	94	77264	32.6	2678	Horse
GCA_010363645.1	83.01	0.79	2744471	98	58766	32.9	2575	Horse
GCA_010363655.1	0.00	0.00	2849042	190	32358	32.8	2734	Horse

GCA_010363705.1	0.00	0.00	2760572	108	55004	32.8	2637	Horse
GCA_010363725.1	0.00	0.00	2868901	93	113827	32.6	2693	Horse
GCA_010363735.1	0.00	0.00	2797926	153	37574	32.8	2629	Horse
GCA_010363755.1	0.00	0.00	2810428	154	44308	32.7	2669	Horse
GCA_010363805.1	99.51	0.08	2681767	70	69455	32.9	2489	Horse
GCA_010363815.1	99.51	0.22	2751939	82	78380	32.8	2590	Horse
GCA_010363825.1	0.00	0.00	2761915	152	32881	32.9	2606	Horse
GCA_010363835.1	0.00	0.00	2757840	26	305130	32.8	2567	Horse
GCA_010363845.1	0.00	0.00	2654663	38	153401	32.7	2408	Horse
GCA_010363915.1	41.99	2.64	2686040	135	34809	32.9	2553	Horse
GCA_010363925.1	0.00	0.00	2897728	218	24239	32.8	2793	Horse
GCA_010363935.1	0.00	0.00	2646430	154	30586	33.0	2481	Horse
GCA_010363945.1	0.00	0.00	2768475	88	76423	32.8	2617	Horse
GCA_010364005.1	0.00	0.00	2824324	170	41461	32.8	2680	Horse
GCA_010364015.1	0.00	0.00	2723519	318	18740	32.9	2653	Horse
GCA_010364035.1	6.77	0.28	2782914	146	68055	32.9	2647	Horse
GCA_010364065.1	6.22	0.07	2886810	183	37039	32.8	2768	Horse
GCA_010364075.1	0.00	0.00	2816791	105	68002	32.8	2662	Horse
GCA_010364105.1	30.31	1.82	2717987	41	143495	32.7	2514	Horse
GCA_010364125.1	3.58	0.00	2792876	55	250821	32.8	2624	Horse
GCA_010364145.1	91.58	0.42	2765262	163	28436	32.8	2611	Horse
GCA_010364165.1	0.00	0.00	2856728	89	103927	32.8	2688	Horse
GCA_010364195.1	0.00	0.00	2845085	83	89316	32.8	2681	Horse
GCA_010364225.1	0.00	0.00	2743061	76	121543	32.8	2615	Horse
GCA_010364235.1	0.00	0.00	2762761	107	60503	32.8	2590	Horse
GCA_010364265.1	0.00	0.00	2929105	189	30800	32.7	2814	Horse
GCA_010364285.1	0.00	0.00	2868740	186	36888	32.8	2771	Horse
GCA_010364295.1	80.66	1.04	2676557	109	49040	32.9	2477	Horse
GCA_010364325.1	77.45	0.96	2858885	91	114357	32.8	2697	Horse
GCA_010364355.1	0.00	0.00	2701412	249	20236	33.0	2573	Horse
GCA_010364385.1	0.00	0.00	2755605	337	15252	33.0	2644	Horse
GCA_010364405.1	0.00	0.00	2734117	294	16050	33.0	2648	Horse
GCA_010364415.1	0.00	0.00	2695757	220	20842	32.9	2521	Horse
GCA_010364445.1	0.00	0.00	2789855	185	33357	32.9	2637	Horse
GCA_010364455.1	62.51	1.64	2811629	192	27490	32.8	2699	Horse
GCA_010364505.1	0.00	0.00	2904812	195	26592	32.7	2778	Horse
GCA_010364525.1	81.68	1.12	2662088	165	29808	32.9	2508	Horse
GCA_010364545.1	81.78	1.00	2729336	217	23523	32.9	2609	Horse
GCA_010364555.1	51.48	2.50	2799122	217	24346	32.9	2693	Horse
GCA_010364605.1	0.00	0.00	2733120	182	29338	32.9	2568	Horse
GCA_010364615.1	51.40	2.55	2757971	183	28228	32.8	2611	Horse
GCA_010364645.1	0.00	0.00	2905925	109	79482	32.7	2777	Horse
GCA_010364655.1	0.00	0.00	2761268	103	68907	32.8	2580	Horse
GCA_010364665.1	0.00	0.00	2801370	156	41618	32.8	2708	Horse
GCA_010364705.1	99.51	0.22	2739512	103	46693	32.8	2572	Horse
GCA_010364725.1	98.18	0.12	2665792	302	14948	33.0	2506	Horse
GCA_010364745.1	0.00	0.00	2854188	98	79144	32.8	2689	Horse
GCA_010364755.1	0.00	0.00	2866245	102	110284	32.8	2708	Horse

GCA_010364775.1	0.00	0.00	2823829	139	44177	32.9	2690	Horse
GCA_010364805.1	0.00	0.00	2772619	134	46690	32.7	2586	Horse
GCA_010364825.1	0.00	0.00	2635427	151	29513	33.0	2441	Horse
GCA_010364845.1	0.00	0.00	2762367	193	24611	32.9	2601	Horse
GCA_010364865.1	0.00	0.00	2754353	131	52110	32.8	2601	Horse
GCA_010364895.1	87.61	0.93	2820945	151	49010	32.9	2692	Horse
GCA_010364925.1	0.00	0.00	2781277	110	54918	32.8	2610	Horse
GCA_010364965.1	0.00	0.00	2777941	66	75151	32.8	2609	Horse
GCA_010364985.1	84.82	0.84	2681423	182	21534	33.0	2539	Horse
GCA_010364995.1	99.51	0.08	2748750	108	55095	32.8	2613	Horse
GCA_010365025.1	0.00	0.00	2722254	152	33190	32.9	2545	Horse
GCA_010365035.1	99.51	0.38	2756290	127	45490	32.8	2633	Horse
GCA_010365075.1	99.51	0.22	2735482	127	38192	32.8	2600	Horse
GCA_019426135.1	0.00	0.00	2934241	57	145101	32.7	2813	Horse
GCA_019426145.1	0.00	0.00	2797187	30	182678	32.7	2592	Horse
GCA_019426175.1	0.00	0.00	2881336	37	133873	32.7	2756	Horse
GCA_019426185.1	0.00	0.00	2872592	53	178770	32.7	2716	Horse
GCA_000443265.1	0.00	0.00	2700799	49	152039	32.8	2475	Pig
GCA_000577675.1	0.00	0.00	2760400	134	38726	32.8	2660	Pig
GCA_000636155.1	0.00	0.00	2697691	47	94942	32.7	2476	Pig
GCA_000636755.1	0.00	0.00	2768179	35	170792	32.9	2555	Pig
GCA_000636775.1	0.00	0.00	2828214	49	137424	32.9	2640	Pig
GCA_000637115.1	0.00	0.00	2749299	46	210838	32.9	2516	Pig
GCA_000637155.1	0.00	0.00	2759747	31	410172	32.8	2520	Pig
GCA_000637175.1	0.00	0.00	2790968	40	201072	32.9	2566	Pig
GCA_000637195.1	8.51	1.14	2855966	66	74313	32.7	2677	Pig
GCA_000637635.1	0.00	0.00	2778077	40	136452	32.9	2566	Pig
GCA_000637675.1	0.00	0.00	2778143	39	164135	32.9	2584	Pig
GCA_000637695.1	0.00	0.00	2767294	36	215005	32.9	2572	Pig
GCA_000637735.1	0.00	0.00	2750337	41	185735	32.9	2510	Pig
GCA_000637775.1	0.00	0.00	2728857	33	180185	32.9	2527	Pig
GCA_000637795.1	21.14	0.80	2752151	32	234095	32.9	2531	Pig
GCA_000637835.1	0.00	0.00	2765581	37	188841	32.8	2565	Pig
GCA_000637855.1	0.00	0.00	2753729	49	137848	32.9	2572	Pig
GCA_000637875.1	0.00	0.00	2732193	23	430511	32.9	2509	Pig
GCA_000637895.1	15.48	0.57	2734096	30	218190	32.9	2515	Pig
GCA_000637935.1	0.00	0.00	2789906	39	162875	32.9	2572	Pig
GCA_000637955.1	0.00	0.00	2772729	40	137847	32.9	2559	Pig
GCA_000637975.1	0.00	0.00	2794969	40	206867	32.9	2574	Pig
GCA_000638135.1	0.00	0.00	2791366	33	212059	32.9	2568	Pig
GCA_000638155.1	0.00	0.00	2819305	47	210840	32.9	2609	Pig
GCA_000638175.1	0.00	0.00	2738012	40	206471	32.9	2497	Pig
GCA_000638195.1	0.00	0.00	2784062	69	75491	32.7	2606	Pig
GCA_000638235.1	0.00	0.00	2806340	35	201512	32.9	2617	Pig
GCA_000638255.1	0.00	0.00	2770167	41	170685	32.9	2544	Pig
GCA_000638375.1	0.00	0.00	2826130	48	218173	32.9	2602	Pig
GCA_000638395.1	0.00	0.00	2783098	46	137850	32.9	2563	Pig
GCA_000638415.1	0.00	0.00	2737686	33	276047	32.8	2501	Pig

GCA_000638515.1	35.39	0.72	2795217	33	179945	32.9	2578	Pig
GCA_000638555.1	31.82	0.80	2863895	51	222579	32.9	2679	Pig
GCA_000638655.1	28.89	0.92	2831096	40	140948	32.9	2639	Pig
GCA_000639075.1	0.00	0.00	2809499	36	148328	32.8	2597	Pig
GCA_000639095.1	0.00	0.00	2761478	35	137848	32.9	2560	Pig
GCA_000639135.1	0.00	0.00	2772448	43	185717	32.8	2561	Pig
GCA_000639555.1	0.00	0.00	2811208	38	137732	32.9	2617	Pig
GCA_000640475.1	33.57	0.80	2786275	45	140578	32.8	2587	Pig
GCA_000647295.1	0.00	0.00	2858892	42	171883	32.9	2688	Pig
GCA_000647315.1	0.00	0.00	2717945	32	238857	32.9	2509	Pig
GCA_000647355.1	0.00	0.00	2761095	36	197020	32.8	2560	Pig
GCA_000647375.1	0.00	0.00	2775018	44	103513	32.9	2567	Pig
GCA_000647415.1	31.70	0.82	2780289	40	137361	32.9	2601	Pig
GCA_000647435.1	0.00	0.00	2816101	46	173454	32.9	2636	Pig
GCA_000682815.1	44.75	0.71	2728435	57	135021	32.8	2549	Pig
GCA_000682835.1	0.00	0.00	2845182	51	137848	32.9	2696	Pig
GCA_000684735.1	36.01	0.80	2754298	39	137402	32.9	2518	Pig
GCA_000684985.1	0.00	0.00	2698033	60	90387	32.7	2483	Pig
GCA_000827155.1	0.00	0.00	2880108	44	164393	32.9	2710	Pig
GCA_001281605.1	0.00	0.00	2775603	51	137629	32.8	2601	Pig
GCA_001455955.1	31.54	0.82	2805326	20	466156	32.9	2588	Pig
GCA_002025125.1	86.14	1.75	2802697	1	2802697	32.9	2551	Pig
GCA_002204555.1	29.99	0.73	2861508	1	2861508	32.9	2621	Pig
GCA_002204575.1	0.00	0.00	2833430	1	2833430	33.0	2614	Pig
GCA_002250135.1	0.00	0.00	2703727	115	52295	32.7	2514	Pig
GCA_002272985.1	31.23	1.19	2829512	157	37433	32.8	2716	Pig
GCA_002272995.1	8.52	1.12	2842047	157	27805	32.8	2734	Pig
GCA_002273065.1	0.00	0.00	2786516	111	55651	32.8	2645	Pig
GCA_002273075.1	7.48	1.12	2785372	69	71421	32.8	2621	Pig
GCA_002273085.1	0.00	0.00	2810766	91	59660	32.8	2652	Pig
GCA_002273165.1	38.65	1.18	2662194	243	18862	33.0	2564	Pig
GCA_002273225.1	0.00	0.00	2871315	110	39584	32.7	2731	Pig
GCA_002273245.1	32.51	1.83	2795064	174	32478	32.8	2716	Pig
GCA_002273255.1	0.00	0.00	2866860	107	56210	32.8	2744	Pig
GCA_002273265.1	30.41	1.05	2885429	168	31200	32.8	2811	Pig
GCA_002273305.1	20.69	1.01	2950113	173	27600	32.7	2880	Pig
GCA_002273335.1	0.00	0.00	2825420	102	41123	32.8	2683	Pig
GCA_002273395.1	30.63	0.97	2743121	157	30009	33.0	2656	Pig
GCA_002273445.1	35.60	1.69	2481871	131	42091	33.0	2392	Pig
GCA_002273455.1	0.00	0.00	2838465	180	23515	32.9	2757	Pig
GCA_002273475.1	28.47	1.24	2803286	118	52807	32.9	2687	Pig
GCA_002273495.1	0.00	0.00	2825424	113	38992	32.9	2696	Pig
GCA_002273545.1	0.00	0.00	2809487	71	80945	32.9	2648	Pig
GCA_002273585.1	33.57	1.67	2807756	140	36828	32.9	2676	Pig
GCA_002273615.1	26.95	1.07	2768283	192	23810	32.9	2630	Pig
GCA_002273665.1	0.00	0.00	2700832	154	29645	32.9	2562	Pig
GCA_002273705.1	0.00	0.00	2692396	134	40074	33.0	2543	Pig
GCA_002273785.1	0.00	0.00	2760960	126	51830	32.9	2616	Pig

GCA_002273825.1	43.78	1.02	2766929	153	33457	32.9	2661	Pig
GCA_002273835.1	26.53	0.92	2794778	115	40332	32.8	2625	Pig
GCA_002273945.1	0.00	0.00	2742057	172	26182	32.8	2629	Pig
GCA_002273955.1	19.89	1.55	2709301	187	28044	32.9	2597	Pig
GCA_002273995.1	20.97	1.21	2549148	261	17141	33.1	2475	Pig
GCA_002274015.1	0.00	0.00	2793837	188	29895	32.8	2707	Pig
GCA_002274035.1	5.85	0.61	2704845	176	31472	32.9	2583	Pig
GCA_002274115.1	0.00	0.00	2524485	259	18598	33.0	2469	Pig
GCA_002275035.1	0.00	0.00	2913297	107	47316	32.7	2783	Pig
GCA_002275045.1	1.98	0.03	2868987	137	36846	32.8	2754	Pig
GCA_002275055.1	87.45	1.39	2416237	175	30960	32.9	2384	Pig
GCA_002275105.1	0.00	0.00	2862876	198	25310	32.8	2812	Pig
GCA_002275115.1	0.00	0.00	2849429	141	37936	32.8	2712	Pig
GCA_002275125.1	0.00	0.00	2923998	142	37620	32.8	2812	Pig
GCA_002275205.1	0.00	0.00	2813875	105	40280	32.9	2687	Pig
GCA_002275225.1	0.00	0.00	2728652	116	38113	33.0	2604	Pig
GCA_002275285.1	0.00	0.00	2730852	163	30360	32.9	2602	Pig
GCA_002275295.1	0.00	0.00	2794320	105	39262	32.8	2652	Pig
GCA_002275365.1	0.00	0.00	2781883	198	23616	32.9	2706	Pig
GCA_002275385.1	0.00	0.00	2754227	131	33834	32.9	2618	Pig
GCA_002275425.1	0.00	0.00	2666570	155	36169	32.9	2543	Pig
GCA_002275435.1	0.00	0.00	2792989	91	51240	32.8	2625	Pig
GCA_002303755.1	0.00	0.00	2441373	201	21279	33.1	2424	Pig
GCA_002304615.1	0.00	0.00	2742118	114	43137	32.9	2588	Pig
GCA_002843775.1	0.00	0.00	2779248	392	12961	32.9	2753	Pig
GCA_002843805.1	0.00	0.00	2781942	142	45900	32.8	2634	Pig
GCA_002843825.1	0.00	0.00	2720017	523	10274	32.9	2742	Pig
GCA_002843845.1	0.00	0.00	2700476	550	9088	32.9	2731	Pig
GCA_002843855.1	0.00	0.00	2745687	333	16738	32.8	2674	Pig
GCA_002934885.1	98.28	50.76	3039519	1137	12092	33.4	3376	Pig
GCA_002934925.1	0.00	0.00	2852402	224	47804	32.9	2699	Pig
GCA_002934945.1	0.00	0.00	2867714	234	47056	33.0	2718	Pig
GCA_002935015.1	0.00	0.00	2848442	248	39482	32.9	2702	Pig
GCA_002935045.1	0.00	0.00	2812349	216	61764	33.2	2658	Pig
GCA_003111745.1	0.00	0.00	2916774	5	2902681	33.0	2653	Pig
GCA_003301075.1	0.00	0.00	2784452	52	110871	32.7	2563	Pig
GCA_003301115.1	11.08	1.19	2825910	64	77283	32.7	2649	Pig
GCA_003309045.1	0.00	0.00	2738754	51	125572	32.7	2518	Pig
GCA_003309085.1	0.00	0.00	2812056	73	101703	32.7	2635	Pig
GCA_003309105.1	0.00	0.00	2752262	46	118352	32.7	2542	Pig
GCA_003309125.1	99.41	0.35	2762131	62	108364	32.8	2561	Pig
GCA_003309145.1	99.41	0.35	2877000	51	158346	32.6	2694	Pig
GCA_003309165.1	98.85	0.35	2744280	48	129868	32.7	2542	Pig
GCA_003309175.1	99.41	0.35	2855301	115	65346	32.6	2709	Pig
GCA_003309205.1	99.51	0.59	3047217	185	142609	32.6	2929	Pig
GCA_003309235.1	99.51	0.57	2949017	119	167036	32.6	2799	Pig
GCA_003309265.1	99.36	0.40	2808349	75	130387	32.7	2650	Pig
GCA_003309275.1	0.00	0.00	3167773	454	12543	32.6	3194	Pig

GCA_003309285.1	0.00	0.00	2955687	135	124804	32.6	2806	Pig
GCA_003309325.1	92.22	1.21	2754312	36	207299	32.8	2536	Pig
GCA_003309355.1	94.75	0.89	2758182	30	429393	32.8	2560	Pig
GCA_003309385.1	95.48	1.34	2808934	58	108474	32.7	2637	Pig
GCA_003309405.1	99.41	0.35	2863415	104	75466	32.6	2695	Pig
GCA_003309445.1	93.30	0.91	2753168	35	228135	32.8	2547	Pig
GCA_003309465.1	0.00	0.00	2893756	118	134855	32.7	2746	Pig
GCA_003309745.1	72.61	24.83	3034893	587	5894	32.6	3161	Pig
GCA_003309755.1	88.80	2.80	3071924	147	126029	32.8	3004	Pig
GCA_003309805.1	0.00	0.00	2971994	116	166324	32.7	2840	Pig
GCA_003309815.1	0.00	0.00	2766549	60	75576	32.7	2583	Pig
GCA_003309845.1	0.00	0.00	2733888	36	200559	32.8	2516	Pig
GCA_003309865.1	99.51	0.16	2786387	54	136496	32.8	2573	Pig
GCA_003309885.1	0.00	0.00	2785156	43	203659	32.8	2573	Pig
GCA_003309905.1	0.00	0.00	2770352	34	208902	32.8	2549	Pig
GCA_003309985.1	0.00	0.00	2815781	81	75457	32.7	2653	Pig
GCA_003310025.1	0.00	0.00	2761230	60	124106	32.7	2553	Pig
GCA_003310045.1	88.70	1.95	2795473	57	86867	32.7	2620	Pig
GCA_003310055.1	99.41	0.35	2795890	61	85115	32.7	2620	Pig
GCA_003310085.1	99.41	0.35	2778903	60	83743	32.7	2596	Pig
GCA_003310105.1	0.00	0.00	2793884	65	101207	32.7	2613	Pig
GCA_003310115.1	99.41	0.35	2794259	60	87020	32.7	2620	Pig
GCA_003310145.1	99.41	0.35	2793212	58	86868	32.7	2617	Pig
GCA_003310155.1	99.41	0.35	2793662	56	86868	32.7	2619	Pig
GCA_003310175.1	99.51	0.35	2857869	58	108395	32.7	2697	Pig
GCA_003310205.1	99.41	0.35	2793760	62	86943	32.7	2618	Pig
GCA_003310215.1	99.41	0.35	2794341	58	87152	32.7	2619	Pig
GCA_003310235.1	99.41	0.38	2823743	63	74014	32.6	2652	Pig
GCA_003310265.1	99.41	0.35	2795789	57	82747	32.7	2617	Pig
GCA_003310275.1	99.41	0.35	2794032	55	86867	32.7	2619	Pig
GCA_003310305.1	99.41	0.35	2793684	51	127764	32.7	2617	Pig
GCA_003310465.1	99.41	0.35	2793647	52	173454	32.7	2616	Pig
GCA_003310505.1	99.41	0.35	2807325	64	106663	32.7	2642	Pig
GCA_003310515.1	99.41	0.71	2911881	143	86905	32.6	2802	Pig
GCA_003310525.1	99.41	0.35	2865133	114	124167	32.6	2715	Pig
GCA_003310565.1	99.41	0.35	2804622	81	62757	32.7	2623	Pig
GCA_003310575.1	99.41	0.35	2780960	54	145039	32.7	2598	Pig
GCA_003310585.1	99.41	0.35	2793918	55	125979	32.7	2616	Pig
GCA_003310625.1	99.41	0.35	2825691	84	108501	32.7	2654	Pig
GCA_003310635.1	99.41	0.35	2793706	61	73100	32.7	2620	Pig
GCA_003310645.1	99.41	0.35	2810209	60	90757	32.6	2632	Pig
GCA_003310685.1	99.41	0.35	2826544	56	108466	32.6	2648	Pig
GCA_003310705.1	99.41	0.38	2824479	62	104996	32.6	2657	Pig
GCA_003310715.1	99.41	0.35	2794542	57	96752	32.7	2620	Pig
GCA_003310745.1	99.51	0.35	2809010	59	99633	32.6	2622	Pig
GCA_003310765.1	99.41	0.35	2807517	73	124914	32.7	2639	Pig
GCA_003310985.1	91.53	1.36	2794310	58	86924	32.7	2620	Pig
GCA_003311005.1	93.99	1.34	2792130	57	104996	32.7	2621	Pig

GCA_003311025.1	84.33	2.82	2793701	51	167250	32.7	2615	Pig
GCA_003311045.1	84.80	2.71	2794344	54	100440	32.7	2621	Pig
GCA_003311055.1	49.93	0.75	2785170	55	130576	32.7	2613	Pig
GCA_003311085.1	15.10	0.87	2760955	60	83170	32.7	2560	Pig
GCA_003311105.1	99.41	0.35	2757955	68	71366	32.7	2557	Pig
GCA_003311115.1	99.51	0.35	2849524	59	124938	32.6	2693	Pig
GCA_003311145.1	99.41	0.35	2751877	61	86989	32.7	2549	Pig
GCA_003311165.1	99.41	0.35	2785626	54	125976	32.7	2604	Pig
GCA_003311175.1	99.41	0.27	2741657	59	108387	32.7	2538	Pig
GCA_003311205.1	99.41	0.35	2779304	55	181629	32.7	2597	Pig
GCA_003311215.1	99.41	0.35	2792250	56	94005	32.7	2619	Pig
GCA_003311245.1	99.41	0.35	2772003	55	86946	32.7	2566	Pig
GCA_003311265.1	99.41	0.35	2794258	54	130578	32.7	2614	Pig
GCA_003311465.1	99.41	0.35	2795017	59	107049	32.7	2618	Pig
GCA_003311475.1	99.41	0.35	2794471	55	86868	32.7	2619	Pig
GCA_003311515.1	99.41	0.35	2796021	59	87152	32.7	2623	Pig
GCA_003311525.1	99.51	0.35	2852748	59	86921	32.6	2693	Pig
GCA_003311545.1	99.41	0.35	2794864	47	197174	32.7	2623	Pig
GCA_003311575.1	99.41	0.35	2797467	57	101128	32.7	2618	Pig
GCA_003311585.1	99.41	0.35	2780832	47	130614	32.7	2595	Pig
GCA_003327935.1	99.51	0.38	2915983	156	104123	32.9	2840	Pig
GCA_003327995.1	99.51	0.38	2916901	68	281923	32.9	2803	Pig
GCA_003432345.1	99.41	0.38	2842229	2	2839540	32.9	2625	Pig
GCA_003432365.1	99.41	0.35	2855865	1	2855865	32.8	2637	Pig
GCA_003665115.1	99.41	0.48	2812608	120	94486	33.0	2634	Pig
GCA_003665125.1	99.41	1.58	2792024	77	145840	32.7	2589	Pig
GCA_004011165.1	99.51	0.08	2812862	63	134645	32.8	2623	Pig
GCA_004026185.1	99.13	0.08	2775616	2	2756431	32.9	2516	Pig
GCA_004136235.1	98.38	0.08	2786527	1	2786527	32.9	2565	Pig
GCA_004136255.1	99.51	0.08	3050015	1	3050015	32.8	2914	Pig
GCA_004136655.1	91.27	1.92	2850509	1	2850509	32.8	2642	Pig
GCA_004681195.1	79.00	2.63	2895108	57	137296	32.8	2721	Pig
GCA_005153985.1	23.56	0.71	2785796	2	2755273	32.8	2540	Pig
GCA_009765015.1	12.92	0.50	2910372	75	166505	32.9	2839	Pig
GCA_011007035.1	99.51	0.08	2918828	77	83370	32.7	2783	Pig
GCA_011007045.1	99.51	0.08	2890947	72	93111	32.7	2740	Pig
GCA_011007055.1	99.51	0.08	2869570	66	93140	32.7	2701	Pig
GCA_011007265.1	99.51	0.22	2762086	54	126662	32.7	2548	Pig
GCA_011007315.1	99.41	0.35	2820198	61	108216	32.7	2669	Pig
GCA_011007335.1	99.41	0.35	2809362	72	130692	32.7	2639	Pig
GCA_011007355.1	99.41	0.45	2800026	57	99380	32.7	2620	Pig
GCA_011290635.1	99.41	0.38	2821677	58	161789	32.7	2659	Pig
GCA_011290765.1	99.41	0.35	2814988	62	123971	32.7	2659	Pig
GCA_011290785.1	99.41	0.35	2791546	61	142306	32.7	2600	Pig
GCA_011290835.1	99.41	0.41	2796862	67	107802	32.7	2629	Pig
GCA_011290845.1	99.41	0.35	2789863	53	253813	32.7	2620	Pig
GCA_011290855.1	99.41	0.35	2747383	57	151022	32.7	2548	Pig
GCA_011290865.1	99.41	0.35	2765955	56	208918	32.7	2564	Pig

GCA_011290875.1	99.41	0.35	2814180	62	208918	32.7	2647	Pig
GCA_011290935.1	99.41	0.35	2837141	129	129050	32.7	2724	Pig
GCA_011290955.1	99.41	0.35	2812255	78	116365	32.7	2675	Pig
GCA_011290975.1	99.41	0.35	2803050	62	140831	32.7	2634	Pig
GCA_011290995.1	99.41	0.35	2798087	56	147924	32.7	2626	Pig
GCA_011291015.1	99.41	0.35	2794601	65	150932	32.7	2628	Pig
GCA_011291025.1	99.41	0.35	2788116	59	136121	32.7	2624	Pig
GCA_011291045.1	98.85	0.35	2806051	58	386527	32.7	2636	Pig
GCA_011291055.1	98.54	0.91	2765301	61	151022	32.7	2562	Pig
GCA_011291095.1	87.87	1.98	2759750	60	147924	32.7	2559	Pig
GCA_011291105.1	0.00	0.00	2798076	63	112272	32.7	2632	Pig
GCA_011291135.1	87.03	2.49	2798168	57	141718	32.7	2626	Pig
GCA_011291145.1	90.90	1.92	2812801	62	123972	32.7	2655	Pig
GCA_011291155.1	0.00	0.00	2802272	54	201822	32.7	2624	Pig
GCA_011291195.1	58.18	3.07	2786216	55	233107	32.7	2615	Pig
GCA_011291215.1	0.00	0.00	2802947	51	201821	32.7	2623	Pig
GCA_011291235.1	99.41	0.38	2785893	52	233107	32.7	2617	Pig
GCA_011291245.1	99.41	0.35	2804515	51	233106	32.7	2627	Pig
GCA_011291255.1	99.41	0.35	2787869	56	150512	32.7	2620	Pig
GCA_011291265.1	99.41	0.35	2800270	52	237949	32.7	2624	Pig
GCA_011291315.1	99.41	0.35	2820109	67	159265	32.7	2662	Pig
GCA_011291335.1	99.41	0.35	2788875	52	254587	32.7	2614	Pig
GCA_011291345.1	99.41	0.35	2784486	56	202772	32.7	2610	Pig
GCA_011291365.1	99.41	0.35	2780673	60	153459	32.7	2613	Pig
GCA_011291385.1	99.41	0.35	2735437	68	122801	32.7	2538	Pig
GCA_011291415.1	99.41	0.35	2801917	91	194446	32.7	2666	Pig
GCA_011291425.1	99.41	0.35	2800858	59	151022	32.7	2626	Pig
GCA_011291455.1	99.27	0.35	2801527	50	250714	32.7	2634	Pig
GCA_011291475.1	99.41	0.35	2790872	55	286021	32.7	2614	Pig
GCA_011291485.1	99.41	0.35	2785938	57	286020	32.7	2608	Pig
GCA_011291515.1	99.41	0.35	2762404	68	120949	32.7	2565	Pig
GCA_011291525.1	99.41	0.35	2736680	54	191710	32.7	2540	Pig
GCA_011291555.1	99.41	0.35	2801858	55	201821	32.7	2623	Pig
GCA_011291575.1	99.41	0.35	2804546	52	233106	32.7	2627	Pig
GCA_011291585.1	99.41	0.35	2781849	57	124131	32.7	2587	Pig
GCA_011291595.1	99.41	0.35	2732797	54	124357	32.7	2533	Pig
GCA_011291635.1	98.63	0.35	2741050	69	120964	32.7	2542	Pig
GCA_011291645.1	99.41	0.35	2780305	55	129931	32.7	2586	Pig
GCA_011291675.1	99.41	0.35	2826941	61	233106	32.7	2669	Pig
GCA_011291685.1	85.70	2.85	2792275	60	151022	32.7	2624	Pig
GCA_011291715.1	86.21	2.49	2798145	56	147924	32.7	2627	Pig
GCA_011291735.1	93.20	1.36	2755560	59	145164	32.7	2552	Pig
GCA_011291755.1	77.03	3.61	2814159	70	124131	32.7	2651	Pig
GCA_011291775.1	71.98	3.64	2798427	57	147924	32.7	2628	Pig
GCA_011291815.1	60.98	3.13	2759407	60	151022	32.7	2560	Pig
GCA_011291825.1	52.81	1.95	2757398	63	154087	32.7	2564	Pig
GCA_011291875.1	99.41	0.35	2794440	55	186570	32.7	2627	Pig
GCA_011291885.1	99.41	0.35	2811128	84	113835	32.7	2662	Pig

GCA_011291915.1	99.41	0.35	2801191	59	151022	32.7	2628	Pig
GCA_011291935.1	99.41	0.35	2794403	56	186570	32.7	2630	Pig
GCA_011291945.1	99.41	0.35	2854779	71	86875	32.8	2714	Pig
GCA_011291975.1	99.41	0.35	2815237	60	138952	32.7	2654	Pig
GCA_011291995.1	99.41	0.35	2854562	81	122921	32.8	2725	Pig
GCA_011292005.1	99.41	0.35	2761868	51	233167	32.7	2560	Pig
GCA_011292035.1	99.41	0.35	2778695	63	126316	32.8	2620	Pig
GCA_011292045.1	99.41	0.35	2775440	53	306785	32.7	2597	Pig
GCA_011292075.1	99.38	0.35	2784258	47	267139	32.7	2617	Pig
GCA_011292085.1	99.39	0.35	2765426	59	138951	32.7	2596	Pig
GCA_011292095.1	99.41	0.35	2754177	64	144536	32.7	2553	Pig
GCA_011292105.1	99.41	0.35	2796885	52	197086	32.7	2623	Pig
GCA_011292115.1	99.41	0.35	2775226	52	151040	32.7	2591	Pig
GCA_011292175.1	99.41	0.35	2766293	61	174033	32.7	2569	Pig
GCA_011292185.1	99.41	0.35	2797379	62	150932	32.7	2633	Pig
GCA_011292195.1	99.41	0.35	2767968	65	87425	32.8	2595	Pig
GCA_011292205.1	99.41	0.35	2773291	53	189407	32.7	2595	Pig
GCA_011292215.1	99.41	0.35	2794016	54	189408	32.7	2617	Pig
GCA_011292275.1	99.41	0.39	2824351	59	201054	32.7	2656	Pig
GCA_011292295.1	99.41	0.35	2783121	59	150932	32.7	2614	Pig
GCA_011292305.1	99.41	0.35	2801894	57	189409	32.7	2627	Pig
GCA_011292315.1	99.41	0.35	2815486	61	128005	32.7	2656	Pig
GCA_011292325.1	99.41	0.35	2815102	65	123304	32.7	2660	Pig
GCA_011292375.1	99.41	0.35	2771102	60	186353	32.7	2572	Pig
GCA_011292395.1	95.26	1.34	2806800	56	179996	32.7	2630	Pig
GCA_011292405.1	67.96	3.64	2819604	85	122830	32.7	2690	Pig
GCA_011292415.1	33.50	1.00	2798824	55	233047	32.7	2624	Pig
GCA_011292425.1	15.52	1.04	2813161	66	142247	32.7	2642	Pig
GCA_011292435.1	99.41	0.35	2765122	54	181109	32.7	2568	Pig
GCA_011292495.1	99.41	0.35	2813786	67	128045	32.7	2650	Pig
GCA_011292515.1	99.41	0.35	2786468	59	147924	32.7	2587	Pig
GCA_011292525.1	99.41	0.35	2794217	60	101984	32.7	2616	Pig
GCA_011292555.1	99.51	0.56	2881960	110	99241	32.7	2768	Pig
GCA_011292565.1	99.41	0.35	2783762	52	233106	32.7	2602	Pig
GCA_011292595.1	99.41	0.35	2791581	53	197508	32.7	2620	Pig
GCA_011292615.1	99.41	0.35	2800646	57	147924	32.7	2627	Pig
GCA_011292625.1	99.41	0.35	2867039	66	165578	32.7	2724	Pig
GCA_013366385.1	99.51	0.08	2754640	38	205642	32.8	2524	Pig
GCA_013371405.1	99.65	1.99	3033408	71	159182	33.0	2824	Pig
GCA_013372685.1	99.51	0.16	2825174	77	109384	32.8	2622	Pig
GCA_013372705.1	99.51	0.08	2736144	35	206342	32.8	2505	Pig
GCA_013602305.1	99.51	0.21	2850443	29	158238	32.7	2643	Pig
GCA_015767475.1	99.41	0.08	2755848	3	2752938	32.8	2495	Pig
GCA_015767495.1	99.51	0.08	2885697	5	2839435	33.0	2680	Pig
GCA_016920425.1	99.41	0.08	2811983	46	138166	32.6	2636	Pig
GCA_016920445.1	99.41	0.08	2768557	32	224462	32.7	2575	Pig
GCA_016920495.1	99.41	0.22	2749925	34	154222	32.7	2556	Pig
GCA_016920525.1	99.41	0.08	2817800	42	150025	32.6	2637	Pig

GCA_016920535.1	99.18	1.04	2739847	85	68329	32.7	2585	Pig
GCA_016920545.1	99.41	0.81	2745162	96	55871	32.7	2590	Pig
GCA_016920585.1	99.41	0.08	2738907	28	203063	32.7	2540	Pig
GCA_016920595.1	99.41	0.08	2737728	33	154603	32.7	2539	Pig
GCA_016920625.1	98.09	0.08	2710877	41	107115	32.7	2511	Pig
GCA_016920635.1	99.41	0.08	2742640	32	231352	32.7	2543	Pig
GCA_018141385.1	73.35	2.91	2744548	1	2744548	33.0	2496	Pig
GCA_018682235.1	44.80	0.76	2865201	269	108676	33.1	2768	Pig
GCA_018682265.1	48.27	0.67	2828584	140	113968	32.9	2703	Pig
GCA_018987325.1	40.46	0.83	2832304	4	2809746	32.9	2591	Pig
GCA_018987345.1	99.32	0.08	2804228	3	2796421	33.0	2604	Pig
GCA_019039135.1	99.51	0.38	2932563	18	1508659	33.0	2754	Pig
GCA_019585335.1	98.28	6.21	2834040	86	107840	33.2	2660	Pig
GCA_020702585.2	99.51	0.08	2809169	2	2806809	32.9	2554	Pig
GCA_022549815.1	99.51	0.08	2905917	37	253758	32.7	2729	Pig
GCA_022549825.1	99.51	0.08	2650587	48	256631	32.9	2424	Pig
GCA_022549905.1	99.51	0.08	2882301	37	225053	32.7	2706	Pig
GCA_022549975.1	99.51	0.08	2785549	32	353278	32.8	2573	Pig
GCA_022549985.1	99.51	0.08	2796183	34	423973	32.8	2593	Pig
GCA_022549995.1	99.51	0.08	2794262	31	423108	32.8	2585	Pig
GCA_024515555.1	99.51	0.11	2779386	50	468047	32.8	2578	Pig
GCA_024515565.1	99.51	0.08	2812873	52	424039	32.8	2617	Pig
GCA_024515595.1	99.51	0.08	2836891	70	213507	32.8	2670	Pig
GCA_024515615.1	99.51	0.08	2786185	54	241417	32.8	2585	Pig
GCA_024515635.1	99.51	0.08	2765575	47	410023	32.8	2557	Pig
GCA_024515655.1	99.41	0.08	2744292	80	626486	32.8	2524	Pig
GCA_024515675.1	99.41	0.08	2763648	35	395132	32.7	2549	Pig
GCA_024515695.1	99.41	0.08	2731642	40	395072	32.7	2504	Pig
GCA_024515705.1	99.41	0.08	2728509	34	394945	32.7	2502	Pig
GCA_024515755.1	99.41	0.08	2731155	33	626493	32.7	2507	Pig
GCA_024515775.1	99.51	0.08	2809202	41	173178	32.9	2623	Pig
GCA_024515795.1	99.51	0.08	2809798	40	287205	32.9	2626	Pig
GCA_024515805.1	99.51	0.08	2786308	38	423992	32.8	2584	Pig
GCA_024515835.1	99.51	0.16	2870771	61	150259	32.8	2658	Pig
GCA_024515855.1	99.51	0.08	2855614	39	241812	32.9	2683	Pig
GCA_024515875.1	95.26	1.34	2865133	114	124167	32.6	2715	Pig
GCA_024515895.1	90.51	2.20	2787696	50	101642	32.7	2607	Pig
GCA_024515915.1	0.00	0.00	2737798	53	106149	32.7	2525	Pig
GCA_024515935.1	99.41	0.35	2829213	59	169522	32.7	2660	Pig
GCA_024515955.1	79.17	2.94	2816902	64	85453	32.7	2649	Pig
GCA_024580375.1	0.00	0.00	2828676	89	545094	32.8	2692	Pig
GCA_024580615.1	42.94	0.73	2828128	84	545439	32.8	2692	Pig
GCA_024580815.1	29.84	1.83	2831487	88	544928	32.8	2696	Pig
GCA_024581045.1	99.51	0.15	2825941	76	545554	32.8	2682	Pig
GCA_024588215.1	99.51	0.16	2815325	97	149706	32.8	2584	Pig
GCA_024588235.1	99.51	0.69	2857703	57	141053	32.9	2676	Pig
GCA_024588305.1	99.51	0.21	2795450	69	122237	32.8	2574	Pig
GCA_024588315.1	99.51	0.16	2812109	221	122868	32.4	2549	Pig

GCA_024588355.1	99.51	0.21	2793607	82	102032	32.8	2571	Pig
GCA_024588375.1	99.51	29.34	3628111	1837	86896	32.7	4642	Pig
GCA_024588385.1	99.51	0.18	2861179	58	276640	32.9	2696	Pig
GCA_024588415.1	99.51	0.16	2768348	80	149708	32.8	2530	Pig
GCA_024588425.1	99.51	0.16	2780541	87	149874	32.8	2550	Pig
GCA_024588435.1	99.51	0.21	2798704	75	111707	32.8	2581	Pig
GCA_024588475.1	99.51	0.15	2769808	67	301985	32.8	2544	Pig
GCA_024588495.1	99.51	0.13	2854839	57	201660	32.9	2678	Pig
GCA_024588535.1	99.51	0.16	2888998	89	122963	32.8	2702	Pig
GCA_024588575.1	99.51	0.16	2835665	86	137923	32.8	2623	Pig
GCA_024588595.1	99.51	0.13	2875591	88	239444	32.9	2699	Pig
GCA_024588615.1	99.51	0.13	2866357	50	423879	32.8	2679	Pig
GCA_024588625.1	99.51	0.08	2843232	232	135091	32.8	2724	Pig
GCA_024588675.1	99.51	0.21	2866661	94	149643	32.8	2655	Pig
GCA_024588695.1	99.51	0.08	2844158	68	197232	32.8	2670	Pig
GCA_024588715.1	99.51	0.21	2819501	79	149708	32.8	2592	Pig
GCA_024588735.1	99.51	0.20	2866042	49	307329	32.8	2682	Pig
GCA_024588755.1	99.51	0.08	2862275	140	172499	32.9	2717	Pig
GCA_024588765.1	99.51	0.16	2816445	88	122860	32.8	2601	Pig
GCA_024588795.1	90.99	1.80	2872285	63	200769	32.9	2702	Pig
GCA_024588815.1	85.21	2.12	2842190	66	200901	32.8	2666	Pig
GCA_024588825.1	99.51	0.13	2798646	39	224313	32.8	2587	Pig
GCA_024588855.1	0.00	0.00	2843668	59	197234	32.8	2663	Pig
GCA_024588895.1	53.87	1.59	2813835	89	122865	32.8	2602	Pig
GCA_024588915.1	37.41	0.94	2867374	55	301810	32.9	2674	Pig
GCA_024588925.1	99.51	0.08	2804586	76	423769	32.9	2612	Pig
GCA_024588945.1	99.51	0.13	2870569	62	169239	32.9	2702	Pig
GCA_024588975.1	99.51	0.21	2803182	92	144046	32.8	2573	Pig
GCA_024588995.1	99.51	0.69	2876758	68	285965	32.9	2716	Pig
GCA_024589015.1	99.51	0.13	2764335	51	206361	32.8	2539	Pig
GCA_024589055.1	99.51	0.08	2781446	54	150109	32.8	2563	Pig
GCA_900088495.1	99.28	0.08	2795084	26	340700	32.8	2592	Pig
GCA_900088505.1	99.51	0.08	2799161	29	340700	32.8	2600	Pig
GCA_900088515.1	99.51	0.08	2816773	30	278432	32.9	2644	Pig
GCA_943789715.1	99.51	0.08	2772477	124	130999	32.8	2589	Pig
GCA_000024585.1	99.51	0.08	2847542	4	2824404	32.8	2643	Poultry
GCA_000636795.1	99.41	0.08	2738992	54	99761	32.7	2527	Poultry
GCA_000636835.1	99.41	0.11	2778536	51	116073	32.6	2565	Poultry
GCA_000636915.1	99.51	0.08	2811754	44	186088	32.9	2586	Poultry
GCA_000636935.1	99.41	0.08	2720043	49	82862	32.7	2503	Poultry
GCA_000637255.1	0.00	0.00	2829675	43	191379	32.8	2637	Poultry
GCA_000637275.1	99.51	0.17	2935355	60	137433	32.8	2783	Poultry
GCA_000637315.1	0.00	0.00	2779581	33	204321	32.8	2582	Poultry
GCA_000637355.1	0.00	0.00	2859902	50	167796	32.9	2685	Poultry
GCA_000637415.1	0.00	0.00	2861204	72	152443	32.9	2668	Poultry
GCA_000637455.1	0.00	0.00	2782796	41	165294	32.9	2582	Poultry
GCA_000637475.1	0.00	0.00	2878030	53	169749	32.8	2704	Poultry
GCA_000637755.1	0.00	0.00	2827682	58	139899	32.9	2637	Poultry

GCA_000638455.1	99.51	0.12	2804405	33	191318	32.8	2605	Poultry
GCA_000638475.1	0.00	0.00	2806630	31	230166	32.8	2605	Poultry
GCA_000638495.1	99.45	2.39	2864511	89	69475	32.5	2812	Poultry
GCA_000639355.1	97.09	0.08	2786909	35	185843	32.9	2588	Poultry
GCA_000639455.1	0.00	0.00	2829441	29	324106	32.8	2662	Poultry
GCA_000639475.1	0.00	0.00	2847825	49	185156	32.8	2695	Poultry
GCA_000639495.1	95.20	0.30	2768173	41	195084	32.9	2583	Poultry
GCA_000639895.1	95.78	0.32	2851226	69	91001	32.8	2682	Poultry
GCA_000639915.1	95.31	0.35	2866674	53	201477	32.9	2715	Poultry
GCA_000639935.1	95.31	0.24	2809736	59	122386	32.8	2625	Poultry
GCA_000684355.1	99.51	0.08	2842068	60	135022	32.8	2668	Poultry
GCA_000684375.1	99.51	0.08	2816363	51	179074	32.8	2635	Poultry
GCA_000684395.1	0.00	0.00	2871870	64	137848	32.9	2712	Poultry
GCA_000684435.1	99.51	0.08	2834417	60	137396	32.8	2676	Poultry
GCA_000684455.1	0.00	0.00	2831926	60	137411	32.8	2680	Poultry
GCA_000684475.1	95.59	0.51	2897110	63	210851	32.9	2757	Poultry
GCA_000684555.1	95.59	0.24	2833656	45	143724	32.9	2641	Poultry
GCA_000684575.1	99.51	0.19	2804667	27	375798	32.8	2601	Poultry
GCA_000684615.1	0.00	0.00	2804567	45	124861	32.9	2609	Poultry
GCA_000684675.1	82.12	0.93	2761528	46	135555	32.8	2553	Poultry
GCA_000684695.1	83.99	0.93	2879188	56	137418	32.8	2706	Poultry
GCA_000684715.1	79.53	0.93	2884669	57	134968	32.9	2733	Poultry
GCA_003336495.1	0.00	0.00	2794799	45	158339	32.7	2623	Poultry
GCA_003336545.1	0.00	0.00	2809930	50	207451	32.7	2647	Poultry
GCA_003336555.1	80.53	0.99	2861584	60	91959	32.7	2689	Poultry
GCA_003336565.1	0.00	0.00	2815966	73	79635	32.7	2652	Poultry
GCA_003336575.1	0.00	0.00	2804495	85	73447	32.7	2653	Poultry
GCA_003336625.1	0.00	0.00	2798548	55	165867	32.7	2623	Poultry
GCA_003336635.1	86.36	1.06	2834199	61	120745	32.7	2663	Poultry
GCA_003343155.1	87.60	0.53	2841298	4	1747067	32.8	2642	Poultry
GCA_003350605.1	0.00	0.00	3092777	5	1736264	32.8	2903	Poultry
GCA_011007015.1	80.31	0.93	2996058	78	109272	32.6	2873	Poultry
GCA_011007195.1	92.40	0.24	2837064	102	58385	32.7	2651	Poultry
GCA_011007235.1	0.00	0.00	2796694	48	120207	32.8	2604	Poultry
GCA_013004085.1	80.73	1.06	2929452	4	2883539	32.9	2742	Poultry
GCA_013867445.1	66.24	2.06	2827070	79	113382	32.7	2678	Poultry
GCA_013867535.1	98.76	0.08	2793999	96	58708	32.7	2658	Poultry
GCA_013868115.1	99.51	0.08	2827128	78	112063	32.7	2674	Poultry
GCA_013868125.1	99.51	0.08	2819005	89	88330	32.7	2680	Poultry
GCA_013868215.1	99.51	0.08	2820206	83	89359	32.7	2678	Poultry
GCA_013868225.1	99.37	0.08	2820532	78	89593	32.7	2671	Poultry
GCA_013868245.1	99.51	0.08	2815833	89	81253	32.7	2681	Poultry
GCA_013868255.1	99.51	0.08	2829110	59	110438	32.7	2669	Poultry
GCA_013868275.1	99.37	0.13	2820740	87	69157	32.7	2666	Poultry
GCA_013868315.1	99.51	0.08	2826888	78	112063	32.7	2674	Poultry
GCA_013868325.1	99.51	0.08	2816829	85	89359	32.7	2677	Poultry
GCA_013868335.1	99.51	0.08	2822445	78	103662	32.7	2672	Poultry
GCA_013868355.1	99.51	0.08	2819741	84	95182	32.7	2677	Poultry

GCA_013868395.1	99.37	0.10	2804903	87	80015	32.7	2661	Poultry
GCA_018680615.1	99.41	0.35	2840042	62	103619	32.6	2661	Poultry
GCA_018680735.1	99.41	0.35	2880219	78	130576	32.6	2741	Poultry
GCA_018680755.1	99.41	0.63	2879090	83	88019	32.6	2751	Poultry
GCA_018680785.1	99.41	0.35	2878539	88	103574	32.6	2740	Poultry
GCA_018680815.1	99.41	0.35	2880501	79	103576	32.6	2741	Poultry
GCA_018681025.1	99.41	0.35	2831571	52	130576	32.6	2651	Poultry
GCA_018681055.1	99.41	0.35	2881659	58	103574	32.7	2721	Poultry
GCA_018681095.1	99.41	0.35	2834748	52	142608	32.7	2660	Poultry
GCA_018681115.1	99.23	0.44	2983411	363	47606	33.3	2933	Poultry
GCA_018681135.1	92.81	0.67	2834766	52	134925	32.7	2647	Poultry
GCA_018681155.1	85.70	1.36	2832911	56	108216	32.6	2652	Poultry
GCA_018681235.1	21.27	2.08	2929037	63	143469	32.6	2829	Poultry
GCA_018681355.1	99.51	6.16	3170413	410	73275	32.7	3340	Poultry
GCA_018681365.1	96.51	0.81	2858356	79	108224	32.6	2685	Poultry
GCA_018681415.1	84.61	1.36	2859067	85	102621	32.7	2686	Poultry
GCA_018681435.1	90.55	0.67	2854953	74	100127	32.6	2681	Poultry
GCA_018681555.1	49.18	2.31	2958535	58	108302	32.6	2860	Poultry
GCA_018681775.1	8.84	0.99	2904218	63	102142	32.6	2761	Poultry
GCA_018681795.1	99.41	0.63	2912023	83	120545	32.6	2784	Poultry
GCA_018681835.1	99.41	0.35	2860120	94	88211	32.6	2698	Poultry
GCA_018681855.1	99.41	0.35	2859245	55	108244	32.6	2697	Poultry
GCA_018681875.1	99.41	0.35	2854198	60	107910	32.6	2694	Poultry
GCA_018681895.1	99.41	0.63	2859814	59	107910	32.6	2706	Poultry
GCA_018681915.1	99.41	0.35	2860716	64	101953	32.6	2699	Poultry
GCA_018681925.1	99.51	0.35	2904063	66	108199	32.6	2755	Poultry
GCA_018681955.1	99.51	0.35	2905889	54	139742	32.6	2752	Poultry
GCA_018681975.1	99.41	0.35	2866616	66	108199	32.6	2694	Poultry
GCA_018681995.1	99.41	0.35	2858662	54	124196	32.6	2699	Poultry
GCA_018682015.1	99.41	1.96	2940233	182	50871	33.0	2823	Poultry
GCA_018682035.1	99.41	0.35	2899774	65	107909	32.6	2765	Poultry
GCA_018682055.1	99.41	0.35	2897632	113	67553	32.7	2775	Poultry
GCA_018682075.1	98.85	0.35	2825663	100	52010	32.7	2656	Poultry
GCA_018682095.1	99.41	0.62	2900876	191	54106	33.2	2778	Poultry
GCA_018682115.1	99.41	0.42	2860657	178	46578	32.7	2735	Poultry
GCA_018682125.1	99.23	0.42	2941488	373	46656	33.3	2888	Poultry
GCA_020144645.1	99.51	0.08	2806130	67	114861	32.7	2655	Poultry
GCA_020144695.1	99.51	0.08	2806831	59	152353	32.7	2652	Poultry
GCA_020144725.1	99.51	0.08	2800714	73	123682	32.7	2650	Poultry
GCA_020144745.1	99.47	0.23	2800613	113	51660	32.7	2676	Poultry
GCA_020144765.1	88.16	0.53	2802315	65	114838	32.7	2649	Poultry
GCA_020144775.1	72.66	1.77	2800923	82	98376	32.7	2649	Poultry
GCA_020144805.1	51.23	2.16	2800619	77	123682	32.7	2649	Poultry
GCA_020144825.1	44.24	2.23	2801588	75	117338	32.7	2646	Poultry
GCA_020144835.1	21.38	2.08	2801048	81	90711	32.7	2649	Poultry
GCA_024384555.1	99.23	0.08	2769040	65	110508	32.7	2610	Poultry
GCA_025153865.1	99.41	0.35	2837952	63	88347	32.6	2664	Poultry
GCA_025153935.1	99.51	0.69	2914227	70	190620	32.7	2739	Poultry

GCA_025868925.1	99.41	0.35	2904928	125	67553	32.7	2787	Poultry
GCA_000189435.2	0.00	0.00	2770352	87	92179	32.7	2614	Sheep
GCA_000189455.2	88.71	1.60	2811260	96	76542	32.8	2656	Sheep
GCA_000588835.1	85.88	1.13	2604446	100	40173	32.7	2689	Sheep
GCA_000684515.1	0.00	0.00	2775578	36	204662	32.9	2573	Sheep
GCA_001197935.1	0.00	0.00	2771187	22	645298	32.8	2560	Sheep
GCA_001208645.1	90.21	1.17	2762610	13	645402	32.7	2528	Sheep
GCA_001209085.1	0.00	0.00	2750111	12	440561	32.7	2516	Sheep
GCA_001680915.1	0.00	0.00	2777893	65	122121	32.8	2607	Sheep
GCA_001681105.1	0.00	0.00	2778845	63	105410	32.8	2610	Sheep
GCA_002902425.1	0.00	0.00	2575746	401	14564	32.8	2740	Sheep
GCA_003353135.1	91.31	1.05	2835200	29	193883	32.9	2671	Sheep
GCA_003353145.1	91.99	1.01	2834945	28	227498	32.9	2669	Sheep
GCA_003353195.1	99.40	0.08	2834922	25	285450	32.9	2666	Sheep
GCA_003353205.1	99.51	0.64	2837160	16	447541	32.7	2645	Sheep
GCA_003353235.1	90.33	1.07	2836056	28	228010	32.9	2669	Sheep
GCA_003353265.1	0.00	0.00	2835359	27	380234	32.9	2670	Sheep
GCA_003353315.1	0.00	0.00	2817083	35	374034	32.8	2670	Sheep
GCA_003490125.1	99.49	0.08	2772340	1	2772340	32.9	2549	Sheep
GCA_003627835.1	0.00	0.00	2778005	1	2778005	32.8	2557	Sheep
GCA_003955945.1	99.49	0.08	2864632	1	2864632	32.9	2698	Sheep
GCA_004771335.1	0.00	0.00	2764228	1	2764228	32.9	2509	Sheep
GCA_004771515.1	0.00	0.00	2768115	1	2768115	32.9	2551	Sheep
GCA_004771975.1	0.00	0.00	2778443	1	2778443	32.8	2564	Sheep
GCA_004772055.1	99.51	0.22	2695389	1	2695389	32.9	2422	Sheep
GCA_004772175.1	98.95	0.22	2794042	1	2794042	32.9	2587	Sheep
GCA_004772215.1	99.49	0.08	2841948	1	2841948	32.9	2656	Sheep
GCA_004797195.1	0.00	0.00	2761328	1	2761328	32.9	2543	Sheep
GCA_014876755.1	0.00	0.00	2755024	1	2755024	32.7	2888	Sheep
GCA_020074555.1	99.49	0.08	2749822	53	266642	32.8	2576	Sheep
GCA_023383215.1	99.49	0.08	2816287	43	242893	32.8	2659	Sheep
GCA_900458725.1	99.45	0.22	2871371	3	2726628	32.8	2659	Sheep
GCA_900635315.1	99.49	0.08	2843795	1	2843795	32.9	2650	Sheep
GCA_900636695.1	99.49	0.64	2836670	1	2836670	32.9	2626	Sheep
GCA_918422865.1	99.51	0.24	2814282	35	442045	32.7	2617	Sheep
GCF_023654805.1	99.49	0.22	2759214	41	120969	32.8	2571	Yak
GCF_023654815.1	0.00	0.00	2699401	29	311319	32.7	2490	Yak
GCF_023654845.1	99.49	0.22	2778672	44	121841	32.8	2590	Yak
GCF_023654865.1	99.49	0.22	2758993	41	120969	32.8	2570	Yak
GCF_023654885.1	99.49	0.22	2757870	42	119065	32.8	2570	Yak
GCF_023654895.1	0.00	0.00	2712938	20	240611	32.7	2490	Yak
GCF_023654925.1	0.00	0.00	2713333	21	240611	32.7	2492	Yak
GCF_023654945.1	90.32	0.83	2694577	31	167361	32.7	2485	Yak
GCF_023654965.1	0.00	0.00	2712991	21	240815	32.7	2493	Yak
GCF_023654985.1	0.00	0.00	2664607	17	310112	32.7	2428	Yak
GCF_023655005.1	91.15	0.83	2713287	22	213377	32.7	2497	Yak
GCF_023655015.1	91.93	0.83	2664549	20	238701	32.7	2428	Yak
GCF_023655045.1	93.64	0.83	2664779	18	291497	32.7	2428	Yak

GCF_023655055.1	0.00	0.00	2665503	18	291497	32.7	2429	Yak
GCF_023655085.1	0.00	0.00	2665472	20	238834	32.7	2430	Yak
GCF_023655105.1	91.11	0.83	2717271	30	311427	32.8	2521	Yak
GCF_023655125.1	0.00	0.00	2717350	27	318173	32.8	2523	Yak
GCF_023655135.1	91.45	0.86	2729473	17	307533	32.7	2527	Yak
GCF_023655165.1	0.00	0.00	2741178	21	250136	32.6	2535	Yak
GCF_023655185.1	92.01	0.86	2764565	36	169549	32.7	2559	Yak
GCF_023655205.1	0.00	0.00	2666264	20	238701	32.7	2428	Yak
GCF_023655225.1	99.49	0.22	2760097	43	121841	32.8	2567	Yak
GCF_023655245.1	0.00	0.00	2664750	22	238701	32.7	2427	Yak
GCF_023655255.1	98.35	0.18	2730615	16	375705	32.7	2524	Yak
GCF_023655285.1	99.49	0.22	2760413	44	120969	32.8	2568	Yak
GCF_023655295.1	98.95	0.10	2738660	22	275068	32.7	2527	Yak
GCF_023655315.1	92.47	0.86	2738421	19	275068	32.7	2526	Yak
GCF_023655345.1	0.00	0.00	2737913	19	275068	32.7	2526	Yak
GCF_023655365.1	99.49	0.22	2780666	45	120969	32.8	2588	Yak
GCF_023655385.1	91.37	0.86	2738360	22	275067	32.7	2528	Yak
GCF_023655395.1	0.00	0.00	2779717	37	210578	32.7	2575	Yak
GCF_023655405.1	81.09	1.83	2779477	36	204419	32.7	2575	Yak
GCF_023655445.1	77.21	1.88	2779633	37	204419	32.7	2575	Yak
GCF_023655465.1	0.00	0.00	2779988	39	204419	32.7	2575	Yak
GCF_023655485.1	0.00	0.00	2715586	43	110022	32.7	2488	Yak
GCF_023655495.1	0.00	0.00	2776787	26	183916	32.7	2606	Yak
GCF_023655525.1	76.05	1.88	2779484	36	210578	32.7	2576	Yak
GCF_023655545.1	0.00	0.00	2813929	32	164213	32.7	2630	Yak
GCF_023655565.1	77.06	1.88	2778930	35	210578	32.7	2575	Yak
GCF_023655585.1	85.24	1.26	2760223	25	183916	32.7	2584	Yak
GCF_023655605.1	0.00	0.00	2761472	26	183916	32.7	2586	Yak
GCF_023655625.1	76.66	1.88	2779845	38	204419	32.7	2575	Yak
GCF_023655645.1	93.36	1.27	2779688	37	204419	32.7	2575	Yak
GCF_023655665.1	87.60	1.27	2779923	38	204419	32.7	2575	Yak
GCF_023655675.1	79.53	1.83	2779484	36	210578	32.7	2575	Yak
GCF_023655705.1	83.86	1.27	2779385	35	210578	32.7	2575	Yak
GCF_023655725.1	81.04	1.83	2779922	38	204419	32.7	2575	Yak
GCF_023655745.1	82.16	1.27	2779632	37	204419	32.7	2575	Yak
GCF_023655765.1	99.23	0.24	2814152	31	164213	32.7	2634	Yak
GCF_023655785.1	99.51	0.29	2748826	36	154249	32.7	2569	Yak
GCF_023655805.1	99.23	0.24	2814077	33	164213	32.7	2632	Yak
GCF_023655825.1	99.23	0.24	2813929	32	164213	32.7	2630	Yak
GCF_023655845.1	99.23	0.24	2779485	36	210578	32.7	2575	Yak
GCF_023655865.1	99.23	0.24	2778606	38	204419	32.7	2576	Yak
GCF_023655885.1	99.23	0.24	2813977	32	164213	32.7	2631	Yak
GCF_023655905.1	99.23	0.24	2777904	36	210578	32.7	2575	Yak
GCF_023655925.1	99.23	0.24	2777823	35	225183	32.7	2575	Yak
GCF_023655945.1	99.23	0.24	2777883	35	210578	32.7	2575	Yak
GCF_023655955.1	99.23	0.24	2778971	35	225549	32.7	2576	Yak
GCF_023655985.1	99.23	0.24	2778924	35	225549	32.7	2575	Yak
GCF_023656005.1	99.23	0.24	2813929	32	164213	32.7	2631	Yak

GCF_023656025.1	99.23	0.24	2813929	32	164213	32.7	2632	Yak
GCF_023656045.1	99.23	0.24	2813829	31	164213	32.7	2633	Yak
GCF_023656065.1	99.23	0.24	2812518	29	164213	32.7	2632	Yak
GCF_023656085.1	99.23	0.24	2811899	30	164213	32.7	2631	Yak
GCF_023656095.1	99.23	0.24	2778924	35	225549	32.7	2576	Yak
GCF_023656105.1	91.25	1.27	2812362	30	164213	32.7	2632	Yak
GCF_023656145.1	83.13	1.27	2779955	37	204419	32.7	2575	Yak
GCF_023656985.1	76.79	1.88	2814115	32	164213	32.7	2631	Yak
GCF_023657025.1	70.10	1.66	2701041	61	85094	32.8	2492	Yak
GCF_023657035.1	42.10	1.76	2813929	32	164213	32.7	2631	Yak
GCF_023657065.1	42.92	1.19	2779641	37	210578	32.7	2575	Yak
GCF_023657085.1	99.51	0.22	2761560	28	183916	32.7	2585	Yak
GCF_023657095.1	99.23	0.24	2778769	34	225549	32.7	2576	Yak
GCF_023657125.1	99.51	0.22	2745556	27	190154	32.7	2559	Yak
GCF_023657145.1	99.51	0.22	2746134	34	190154	32.7	2564	Yak
GCF_023657165.1	99.51	0.22	2747904	32	150839	32.7	2562	Yak
GCF_023657185.1	99.23	0.24	2779689	37	204418	32.7	2574	Yak
GCF_023657195.1	99.23	0.24	2779085	36	210578	32.7	2575	Yak
GCF_023657215.1	99.51	0.50	2700812	64	84621	32.8	2494	Yak
GCF_023657245.1	99.23	0.24	2778477	37	204419	32.7	2574	Yak
GCF_023657265.1	99.23	0.24	2779330	35	210578	32.7	2575	Yak
GCF_023657275.1	99.23	0.24	2779632	37	204419	32.7	2575	Yak