UNIVERSITY OF SÃO PAULO

DENNIS CARHUARICRA HUAMAN

Genomic Analysis of Bacterial Isolates from Guinea Pigs (*Cavia porcellus*) and Brown-Throated Sloths (*Bradypus variegatus*)

SÃO PAULO

2023

UNIVERSITY OF SÃO PAULO

DENNIS CARHUARICRA HUAMAN

Genomic Analysis of Bacterial Isolates from Guinea Pigs (*Cavia porcellus*) and Brown-Throated Sloths (*Bradypus variegatus*)

Thesis presented to the Inter-Institutional Grad Program on Bioinformatics of the University of São Paulo, to obtain the degree of Master of Science.

Concentration area: Bioinformatics

Advisor: Prof. Dr. João C. Setubal

Co-advisor. Dr. Lenin Maturrano

SÃO PAULO

2023

A mi familia

“Aunque los pasos toquen mil años este sitio,

no borrarán la sangre de los que aqui cayeron”

Pablo Neruda, Canto General

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Prof. Dr. João Setúbal for his constant support, guidance, and patience throughout my research. I deeply appreciate the trust he placed in me to carry out this and other projects that began during the most challenging phase of the pandemic. His timely clarifications, helpful responses, and enthusiasm were important to complete this work.

I also wish to express my deep gratitude to my co-advisor Dr. Lenin Maturrano who gave me the opportunity to learn and grow professionally in the UBIGEM laboratory of the Universidad Nacional Mayor de San Marcos. I greatly appreciate his trust in me during all these years.

I express my gratitude to my colleagues from the UBIGEM laboratory for their formidable work during sampling, isolation and sequencing of *S.* Typhimurium. Especially to Carmen Rodriguez for her valuable contribution and support throughout this process.

I express my gratitude to Iris Lima González, Patricia Locosque Ramos from the Fundação Parque Zoológico de São Paulo and Prof. Dra. Aline Maria Da Silva from the IQ-USP for their work in the isolation and characterization of sloth isolates, as well as for their insightful feedback that contributed to improving this work

Special thanks also to Barny Whitman from the University of Georgia, USA and Nikos Kyrpides, Nicole Shapiro, and Natasha Brown from the Joint Genome Institute, USA, who enabled the genome sequencing of sloth isolates.

Last, I would like to thank to wonderful Latin American public universities USP and UNMSM for providing accessible and high-quality education for all.

RESUMO

CARHUARICRA D. (2023). Análise Genômica de Isolados Bacterianos de Cobaias (*Cavia porcellus*) e Preguiças-de-Garganta-Marrom (*Bradypus variegatus*). Dissertação de Mestrado. Programa de Pós-graduação Interunidades em Bioinformática. Universidade de São Paulo, São Paulo

A genômica revolucionou nossa compreensão do mundo microbiano. Da emergência e epidemiologia de patógenos à compreensão da evolução e adaptação das bactérias aos seus hospedeiros. Apesar dos enormes avanços, persistem lacunas significativas, particularmente no estudo da diversidade e evolução das bactérias residentes nos microbiomas de animais selvagens e no entendimento do surgimento de agentes patogénicos que afetam o gado em países de rendimento baixo e médio. Nesta dissertação, utilizando diferentes abordagens de genômica comparativa, estudo três espécies bacterianas: a patogênica *Salmonella enterica serovar* Typhimurium (S. Typhimurium) isolada de cobaia (*Cavia porcellus*) no Peru e *Kerstersia gyiorum* e *Neisseria* sp. isolado de preguiças-de-garganta-marrom (*Bradypus variegatus*) de vida livre no Brasil.

Salmonella Typhimurium é o principal patógeno que infecta cobaias no Peru. Primeiro, realizo uma caracterização genômica de *S.* Typhimurium de cobaias em Lima-Peru descrevendo a presença de dois diferentes *S*. Typhimurium clusters do *sequence type* ST19, sendo um deles uma variante monofásica carregando o plasmídeo de virulencia pLST e geneticamente relacionado a isolados de humanos. Em segundo lugar, usando um maior conjunto de dados genômicos da América do Sul e filogenias calibradas no tempo, descrevo o surgimento de uma linhagem peruana de *S*. Typhimurium (linhagem B6) associada a porquinhos-da-índia e humanos. O surgimento desta linhagem altamente prevalente em porquinhos-da-índia coincide com a recente intensificação da produção de porquinhos-da-índia e exibe sinais genéticos de adaptação do hospedeiro com múltiplas mutações de perda de função em genes que codificam estruturas envolvidas na interação célula-hospedeiro e na especificidade do

hospedeiro. Estes resultados destacam os riscos da produção pecuária intensiva na emergência de patógenos.

A análise genômica de isolados de preguiça-de-garganta-marrom foi realizada. Primeiro, *K. gyiorum* exibiu variação filogenética e de conteúdo gênico estruturada de acordo com o hospedeiro. Além disso, diferenças no conteúdo de GC, tamanho do genoma, presença diferencial de elementos genéticos móveis (plasmídeos e fagos) e sistemas de defesa sugerem que as populações de *K. gyiorum* seguiram trajetórias evolutivas divergentes que levaram ao estabelecimento de linhagens restritas ao hospedeiro. A presença de genes com diferentes funções pode ser importante para a adaptação de *K. gyiorum* a diferentes nichos. Em segundo lugar, uma nova espécie de *Neisseria*, denominada *Neisseria bradyp*, é caracterizada genomicamente, o que representa a primeira espécie de Neisseria isolada de preguiça-comum. Análises filogenômicas e *Avergage Nucleotide identity* (ANI) do gênero *Neisseria* revelaram que *N. bradyp* é geneticamente distinto das outras 32 espécies de *Neisseria*. Além disso, *N. bradyp* exibiu conteúdo genético exclusivo e apresentou diferenças na sua estrutura capsular, o que pode desempenhar um papel significativo na colonização do hospedeiro.

Os resultados deste trabalho destacam a importância dos dados genômicos e das ferramentas de bioinformática para estudar o surgimento de novas variantes patogênicas e desvendar a evolução e a adaptação das bactérias ao seu hospedeiro.

Keywords: Genômica comparativa, *Salmonella* Typhimurium, cobaias, *Kerstersia gyiorum*, *Neisseria* sp., preguiça-de-garganta-marrom

ABSTRACT

CARHUARICRA D. (2023). Genomic Analysis of Bacterial Isolates from Guinea Pigs (*Cavia porcellus*) and Brown-Throated Sloths (*Bradypus variegatus*). 120 pages. Dissertation of Master Science. Bioinformatics Graduate Program. University of São Paulo, São Paulo

Genomics has revolutionized our understanding of the microbial world. From the surveillance and epidemiology of emerging pathogens to comprehending the evolution and adaptation of bacteria to their hosts. Despite enormous advances, significant gaps persist, particularly in the study of the diversity and evolution of bacteria residing in wildlife microbiomes and the emergence of bacterial pathogens affecting livestock in low- and middle-income countries. In the present work, using different comparative genomics approaches, I study three bacterial species: the pathogenic *Salmonella* Typhimurium isolated from guinea pig (*Cavia porcellus*) in Peru and *Kerstersia gyiorum* and *Neisseria* sp. isolated from free-living brown-throated sloths (*Bradypus variegatus*) in Brazil.

Salmonella Typhimurium is the main pathogen infecting guinea pigs in Peru. First, I perform a genomic characterization of *S.* Typhimurium infecting guinea pigs in Lima-Peru describing the presence of two ST19 clusters, one of them being a monophasic variant carrying the plasmid of virulence (pLST) and genetically related to human isolates. Second, using a larger genomic data collection from South America and time-scaled phylogenetic analysis I describe the emergence of a Peruvian lineage of *S.* Typhimurium (B6 lineage) associated with guinea pigs and humans. The emergence of this highly prevalent lineage in Peruvian guinea pigs coincides with the recent intensification of guinea pig production and exhibits genetic signals of host adaptation with multiple loss-of-function mutations in genes encoding for structures involved in host-cell interaction and host specificity. These results highlight the risks of intensive livestock production in pathogen emergence.

The genomic analysis of brown-throated sloth isolates was performed. First, *K. gyiorum* exhibited phylogenetic and gene content variation structured according to the host. Additionally, differences in GC content, genome size, differential presence of mobile genetic elements (plasmid and phages) and defense systems suggest that *K. gyiorum* populations have followed divergent evolutionary trajectories that led to the establishment of host-restricted lineages. The presence of genes with different functions may be important for the adaptation of *K. gyiorum* to different niches. Second, a new species of *Neisseria*, named *Neisseria bradyp* is characterized genomically, which represents the first known species associated with common sloths. Average Nucleotide Identity (ANI) and phylogenomics analyses within the *Neisseria* genus revealed that *N. bradyp* is genetically distinct from other 32 *Neisseria* species. Additionally, *N. bradyp* exhibited exclusive genetic content that is notably enriched in carbohydrate metabolism and other pathways. Furthermore, it displayed differences in its capsular structure, which may play a significant role in host colonization.

The results of this work highlight the importance of genomics data and bioinformatics tools to study the emergence of novel pathogenic variants and disentangle the evolution and host adaptation of bacteria to their host.

Keywords: Comparative genomics, *Salmonella* Typhimurium, guinea pigs, *Kerstersia gyiorum*, *Neisseria* sp., brown-throated sloths

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

1. General Introduction and Objectives

1.1. Bacteria that colonize mammals

Bacteria show astonishing diversity, and functional capabilities across diverse ecological niches. They have established intimate relationships with eukaryotic hosts that have contributed to the emergence of complex forms of life (McFall-Ngai et al., 2013). Bacteria have evolved various mechanisms to colonize and proliferate in animal tissues and cells, establishing pathogenic or commensal relationships. Pathogenic bacteria have received the most attention from researchers since the inception of microbiology due to their importance in human health. While only in the last decades commensal or mutualistic bacteria associated with animals and humans began to be studied. Bacterial mutualists provide a great variety of benefits to mammal hosts, including nutrients, antibiotic production and shaping the immune system (Gensollen et al., 2016).

Microbiome studies have revealed great microbial diversity in different mammals. However, It is estimated that more than 40 % of all bacterial genomes from well-studied hosts such as humans do not have reference genomes (Yen and Johnson, 2021). Even much less is known about the bacteria that inhabit the microbiomes of other vertebrates, especially wild animals. Several studies have shown that animal microbiomes are a rich source of bioactive compounds that may have biotechnological applications, including antimicrobials, enzymes and probiotics (Wang et al., 2015; Song et al., 2017; Akbar et al., 2019; Stewart et al., 2019). They also can be reservoirs of potential zoonotic diseases that may impact public health (Cantlay et al., 2017; Suminda et al., 2022). Uncovering this “microbial dark matter” is crucial for comprehending the functions of individual bacteria, the diversity they exhibit within and between host populations, and how each bacterium interacts with others and the host (Levin et al., 2021).

1.2. Comparative genomics to understand the biology and evolution of Bacteria

The evolution and diversity of bacterial populations, especially in pathogenic bacteria, has been studied in a variety of ways, using phenotypic (metabolic activity, and drug resistance) and genotypic methods (variation in few locus and sequence analysis of individual genes). However, these methods only capture a small subset of the total variation and therefore have limited resolution. The introduction of genomics, the powerful combination of genome sequencing and bioinformatics analysis, has transformed our understanding of how bacteria function, evolve, and interact with their environment and other bacteria and has provided unexpected insights into microbial diversity (from strains to phyla).

This transformation began in 1995 when Fleischmann et al. (Fleischmann et al., 1995) published the first bacterial genome of the opportunistic pathogen *Haemophilus influenzae*. This first genome of just over 1.9 million base pairs demonstrated how little was known about bacterial physiology. Of a total of 1743 annotated *H. influenzae* genes, almost 40% had unknown function. In the same year, the second bacterial genome became available: *Mycoplasma genitalium,* an intracellular human pathogen (Fraser et al., 1995). The availability of these two genomes gave rise to comparative genomics revealing important differences in genomic composition and organization (Mushegian and Koonin, 1996). Comparative genomics enables many analyses, the most basic of which is the determination of which genes are present or absent in a particular genome with respect to others. Such information may help the understanding of the genetic basis of phenotypic variation

Early comparative studies already showed that large differences in gene content may occur between genomes of the same prokaryotic species. When three *Escherichia coli* genomes became available in 2002, the comparative analysis revealed that only 39.2 % of the genes were shared by these three genomes (Welch et al., 2002). Soon other similar observations were made about other bacterial species, showing the remarkable plasticity of

such genomes, eventually giving rise to the pangenome concept. The pangenome is the set of all non-redundant genes present in a given set of genomes. The differential gene content among genomes from the same species comes about because of extensive horizontal gene transfer (HGT) and gene loss, two of the main forces driving the evolution of prokaryotes (Arnold et al., 2022).

In the last twenty years, the development of cheaper and more accessible sequencing methods caused an exponential increase in the amount of bacterial genomic data (Kim et al., 2020). This in turn stimulated the development of genome informatics, or computational methods for genome analysis, in particular comparative analysis. A few examples are the genomic epidemiology of pathogenic bacteria (Ruan et al., 2020), pathogenesis and niche specialization (Hurtado et al., 2018), discovery of genes associated with virulence and antimicrobial resistance (Mageiros et al., 2021; The CRyPTIC Consortium, 2022) or identification of antigens through reverse vaccinology (Seib et al., 2012)**.**

1.3. Applications of Comparative Genomics

1.3.1. Comparative genomics to study the emergence of pathogenic variants

Over the last century, we have experienced an accelerating prevalence and intensification of emerging and reemerging infectious diseases with substantial human and economic costs. Bacterial zoonotic agents has dominated the emergence of more than 300 human infectious diseases between 1943 and 2004 (Jones et al., 2008). These phenomena are driven mainly by climatic and anthropogenic activities. For example, the farming intensification and rapid dissemination of plants and animals across the world promote close contact between hosts, which provides increased opportunities for transmission leading eventually to the emergence and spread of novel bacterial populations (Jones et al., 2013). The origin of several livestock-associated *Campylobacter jejuni*, *Salmonella enterica* and

Staphylococcus aureus clones coincides with the expansion and intensification of agriculture and domestication (Spoor et al., 2013; Mourkas et al., 2020; Stevens and Kingsley, 2021a).

Several genetic mechanisms involved in the development of pathogen evolution have been documented. For example, recombination and gene point mutations in metabolic or virulence genes and gene acquisition via horizontal gene transfer of antimicrobial resistance and virulence determinants (Viana et al., 2015; Yue et al., 2015; Mourkas et al., 2020; Bakkeren et al., 2022). Genome degradation and pseudogenization is also important mechanism of host adaptation in some species and have preceded the emergence of pathogenic variants (Langridge et al., 2015).

As a result of the dramatic advances in genome sequencing technologies, thousands of pathogen genomes have become available as part of routine surveillance and clinical studies by health systems worldwide and associated with geographic and temporal information. The use of phylogenetic and Bayesian algorithms based on whole-genome sequence helps to predict the spatiotemporal spread of the pathogen in the population (Drummond and Rambaut, 2007; Didelot et al., 2018). It may also provide information about the underlying transmission network within an outbreak. This affords the opportunity to investigate the genomics and timescale of host adaptation of emergent bacterial pathogens.

1.3.2. Pangenome analysis to study gene content variation and host adaptation

Within the bacterial world, there are “generalist” or “specialist” species depending on their ability to inhabit different niches. For example, *Mycoplasma genitalium* is a specialist since it is an obligate pathogen of humans (Casjens, 1998). On the other hand, there are generalist species such as *Escherichia coli* that can inhabit different animal hosts and environments such as soil and water (Foster-Nyarko and Pallen, 2022). At the genetic level, specialists have a smaller genome than generalists. This is because these require a genetic repertoire to colonize and exploit a greater diversity of resources and hosts. Some populations

within a bacterial species can specialize within certain hosts and can accumulate a set of genes and variations that differentiate them from other populations which can eventually lead to an adaptation process (Sheppard et al., 2018). The existence of this variability in gene content within a single species led to the concept of a pangenome which is the complete set of genes that are present in a given species (Tettelin et al., 2008). This set of genes is usually divided into two categories: core genes, which are present across all individuals in a species, and accessory genes, whose presence varies between individuals or strains. Variability in genetic content is explained by two main forces: horizontal gene transfer and gene loss (Arnold et al., 2022).

Pangenome characteristics differ according to the lifestyle and biology of each species. Returning to the previous example, specialists share almost all genes with each other (i.e. have very little strain-to-strain gene content dissimilarity) have a large 'core' and small 'accessory' genome, and are considered to have closed pangenomes (McInerney et al., 2020). In contrast, generalist species typically have open pangenomes in which gene content varies appreciably from one genome to another, these species show small ‘core’ and ‘large’ accessory genomes such as *E. coli* or *Staphylococcus aureus* (Bosi et al., 2016; McInerney et al., 2017). In these species, HGT can have dramatic impacts on the structure of bacterial genomes by altering gene content and providing raw material for natural selection.

In some bacterial populations, the genetic makeup may display patterns of variation that reflect genetic isolation according to the host source as is observed in species such as *Pasteurella multocida* or *Staphylococcus aureus* (Hurtado et al., 2018; Richardson et al., 2018). This host-associated genetic structure can be analyzed by constructing a phylogenetic tree based on whole-genomes from different hosts showing isolates from the same host group . Additionally, gene content analysis may display differential gene pools associated with distinct hosts. This clustering cannot necessarily be interpreted as lineages adapted to the host, however, it may reflect local clonal expansions driven either by selection (adaptive) or genetic drift (neutral diversification) (Sheppard et al., 2018).

Therefore, the pangenomic analysis is a powerful approach of comparative genomics to study the genetic diversity within a specific taxon. The characterization of patterns of sequence variation in the core genome, and differential gene presence/absence in the accessory genome can provide information about the genomic basis of host adaptation.

1.3.3. Taxonomic classification based on genomic sequence analysis

A robust taxonomy is needed to accurately describe microbial diversity. However, to reach this goal, the task is complex and involves the application of numerous laboratory techniques that have evolved over the years to accurately characterize microorganisms. The earliest methods for bacterial species identification were based on their phenotypic characteristics (morphology, source of isolation, pathogenicity); these features, although useful at the beginning, were lately shown to be highly imprecise and subjective (Moore et al., 2010). After the introduction of molecular methods, novel species were recognized using a polyphasic approach which is the combination of phenotypic and genotypic characterization, together with the phylogeny of the strain (Carro et al., 2021). Phylogenetic analysis based on 16S rRNA sequence is now commonly used as the first step in identifying novel organisms. Stackebrandt & Goebel proposed in 1994, that if two strains share less than 97% 16S rRNA gene sequence similarity, they belong to different species (Stackebrandt and Goebel, 1994). This new methodology has greatly increased the rate of discovering novel species in the last decades. However, in many cases, the use of universal genes (such as 16S rRNA) does not provide sufficient resolution to discriminate between different species (Yarza et al., 2014).

Genomics has become a promising methodology for bacterial taxonomy (Chun et al., 2018). Taking into account what was stated by (Wayne et al., 1987): “the complete deoxyribonucleic acid (DNA) sequence would be the reference standard to determine phylogeny and that phylogeny should determine taxonomy”, genome-based taxonomy provides a reproducible, reliable, highly informative means to infer phylogenetic relationships among bacterial strains. The rapid progress of NGS is currently facilitating genome-wide

sequence analyses. The main method used is the Average Nucleotide Identity (ANI) which represents a mean of identity values shared between two genomes (Chun and Rainey, 2014). ANI can be calculated as a mean of identity values of all BLASTN matches between two genomes (ANIb) or using Mummer ultra-rapid aligning tool (ANIm), this last is faster and easily scalable (Richter and Rosselló-Móra, 2009). In consequence, ANI provides a robust resolution to differentiate closely related species. Most studies used ≥95% ANI cutoff as a standard for species demarcation (Jain et al., 2018). Genome phylogenies is another approach that has helped the reclassifications within orders, families, genera, and species, the availability of genomes will give a better view of the position of a strain in the genera. On the other hand, pangenome analysis refers to all the genes contained in all the strains belonging to the same taxonomic group (e.g. species, genera) and also can discriminate species based on differential gene content (Carro et al., 2021).

In recent years, several works have reported the discovery of novel bacterial species based on comparative genomic analysis. For example, *Neisseria* species in African human populations and wild animals expanding diversity of the genera and novel Enterococcus species in a plethora of hosts and environments worldwide (Diallo et al., 2019; Schwartzman et al., 2023). Most frequent analyses performed consist of ANI, genome phylogeny, and gene content variation.

1.4. Motivations and Objectives

This work is divided into two parts:

Part 1: *Salmonella* Typhimurium from guinea pigs in Peru

Salmonella Typhimurium is the most important pathogen that impacts Peruvian guinea pig production. Recurrent outbreaks of *S.* Typhimurum in a continuously growing guinea pig population in Peru may promote the emergence of new lineages carrying specific mutations and changed behavior (e. g. increased virulence or evolve different host preferences). Despite economic relevance and potential impact on public health, genomic data of *S.* Typhimurium from guinea pig farms is not available. Genomic investigation may help understand the underlying mechanism of S. Typhimurium emergence in guinea pig populations and guide the design of targeted interventions and disease eradication.

In this part (Chapters 2 and 3), I analyze genomic sequences of *S.* Typhimurium isolated from guinea pigs in Peru and compare them with publicly available genomes from other sources to gain insights into:

1. The genomic characteristics and diversity of *S.* Typhimurium isolated from guinea pigs in Lima, Peru (Chapter 2)
2. The temporal evolution of *S.* Typhimurium in guinea pig and humans in Peru. (Chapter 3)

Part 2: *Kerstersia gyiorum* and *Neisseria sp.* from brown-throated sloths in Brazil

Bacteria that colonize humans can also inhabit other mammals, including wild animals. Strains from wild hosts may have a different genetic makeup compared to strains of human or livestock origin, potentially exhibiting novel metabolic capacities, increased virulence, or zoonotic potential. Despite the progress and cost reduction in sequencing technologies, whole genome sequences of bacteria inhabiting wild animal microbiomes are scarce. The Bradypus

Research Project (led by the Parque Zoológico de São Paulo) aims to study the bacterial diversity of free-living brown-throated sloths (*Bradypus variegatus*) in the Parque Estadual das Fontes do Ipiranga (PEFI), Brazil. This project has isolated and conducted whole-genome sequencing for two bacterial species: *Kerstersia gyiorum* and *Neisseria sp.*, representing the first-ever genome sequences available for these species in common sloths. This new genomic information allows us to explore the genetic variation (at the phylogenetic and gene content level) compared to isolates from humans and other hosts to determine if they carry genes involved in new metabolic pathways or if it possess differential mobile elements, virulence or antimicrobial resistance genes. Following these premises, in this part (Chapters 4 and 5) our aims were:

3. Analyze the genomic diversity of *K. gyiorum* isolated from brown-throated sloths and compare with human isolates to gain insight into genomic variation and host adaptation (Chapter 4)

4. Characterize the *Neisseria* sp. genomes from brown-throated sloths, compare them with other *Neisseria* species and determine if they belong to a new species (Chapter 5)

Relationship between the two parts

Although the two parts were developed independently of each other, they have much in common in terms of methodology. This fact means that each part to some extent benefited from experience gained in the other part (and vice versa), resulting in a dissertation richer than it would have been if it had only one of these parts.

2. Genomic Characterization of *Salmonella* Typhimurium isolated from guinea pigs with salmonellosis in Lima, Peru

Abstract

Salmonella enterica subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) is one of the most important foodborne pathogens that infect humans globally. The gastrointestinal tracts of animals like pigs, poultry or cattle are the main reservoirs of *Salmonella* serotypes. Guinea pig meat is an important protein source for Andean countries, but this animal is commonly infected by *S.* Typhimurium, producing high mortality rates and generating economic losses. Despite its impact on human health, food security, and economy, there is no genomic information about the *S.* Typhimurium responsible for the guinea pig infections in Peru. Here, we sequence and characterize 11 *S.* Typhimurium genomes isolated from guinea pigs from four farms in Lima-Peru. We were able to identify two genetic clusters (HC100_9460 and HC100_9757) distinguishable at the H100 level of the Hierarchical Clustering of Core Genome Multi-Locus Sequence Typing (HierCC-cgMLST) scheme with an average of 608 SNPs of distance. All sequences belonged to sequence type 19 (ST19) and HC100_9460 isolates were typed *in silico* as monophasic variants (1,4,[5],12:i:-) lacking the *fljA* and *fljB* genes. Phylogenomic analysis showed that human isolates from Peru were located within the same genetic clusters as guinea pig isolates, suggesting that these lineages can infect both hosts. We identified a genetic antimicrobial resistance cassette carrying the *ant(3)-Ia*, *dfrA15*, *qacE*, and *sul1* genes associated with transposons TnAs3 and IS21 within an IncI1 plasmid in one guinea pig isolate, while antimicrobial resistance genes (ARGs) for β-lactam ($bla_{\text{CTX-M-65}}$) and colistin (*mcr-1*) resistance were detected in Peruvian human-derived isolates. The presence of a virulence plasmid highly similar to the pSLT plasmid (LT2 reference strain) containing the *spvRABCD* operon was found in all guinea pig isolates. Finally, seven phage sequences (STGP_Φ1 to STGP_Φ7) were identified in guinea pig isolates, distributed according to the genetic lineage (H50 clusters level) and forming part of the specific gene content of each cluster. This study presents, for the first time, the genomic characteristics of

S. Typhimurium isolated from guinea pigs in South America, showing particular diversity and genetic elements (plasmids and prophages) that require special attention and also broader studies in different periods of time and locations to determine their impact on human health.

Keywords: *Salmonella* Typhimurium; guinea pig; comparative genomics; antimicrobial resistance genes

2.1. INTRODUCTION

Salmonella is one of the most important foodborne pathogens in humans, responsible for 78 million illnesses with a balance of 59,000 deaths each year (Havelaar et al., 2015). *Salmonella enterica* consists of more than 2600 serovars and *S. enterica* serovar Typhimurium is one of the most frequently reported in human infections (European Centre for Disease Prevention and Control, 2017; Centers for Disease Control and Prevention, 2018). Poultry, pigs, and cattle are the main reservoirs of *Salmonella* and constitute vehicles for human infections when there is the consumption of contaminated food (Branchu et al., 2018). In recent decades, the intensification of farm practices has led to the emergence of zoonotic pathogens that threaten animal and human health. The emergence of new variants of *S.* Typhimurium has been reported recently in pigs (U288, ST34) and humans (ST313) and poses a serious threat to health and food safety (Okoro et al., 2015; Bawn et al., 2020; Kirkwood et al., 2021). Furthermore, variants lacking one or both flagellar phases have been widely reported. The monophasic variants of *S.* Typhimurium (4,[5],12:i:-) emerged last century and multiple clones have been reported worldwide and are characterized by not expressing the second-phase flagellar antigen (defined as 1,2 in the antigenic formula) of *S.* Typhimurium (4,[5],12:i:1,2) (Arrieta-Gisasola et al., 2020; Sun et al., 2020). Genomic surveillance provides an opportunity to identify variants, genetic signals of resistance and virulence, or host adaptation of pathogens such as *Salmonella*, facilitating the management of emerging variants that pose a health threat (Gardy and Loman, 2018; Argimón et al., 2021).

The guinea pig (*Cavia porcellus*) is a rodent domesticated in the Central Andes of South America 3000–6000 years ago and used by Andean populations as a food source and for ritual and medicinal purposes (Lord et al., 2020). From that period until today, guinea pig farming has been one of the most important economic and cultural practices for Andean populations (Sánchez-Macías et al., 2018). In recent decades, guinea pig production has undergone an enormous transformation from small-scale family farming with a few dozen animals to large farms with tens of thousands of animals. As was observed in other pathogens, the intensification of farming practices has led to the emergence of host-adapted lineages (Sheppard et al., 2018; Mourkas et al., 2020). In guinea pig farms, *Salmonella* is responsible for high morbidity and mortality affecting the economy and food security in Andean countries such as Peru; although there is no representative information about the regional prevalence of salmonellosis in its commercial production, a recent study performed in 40 animals from 3 farms in La Libertad-Peru found a prevalence of 27.5%, with a higher percentage of positive samples from commercial breeding (Matsuura S. et al., 2010; Moya A. L., 2019). Salmonellosis mainly affects the gastrointestinal tract of guinea pigs, with diarrhea as the most evident clinical manifestation; however, subclinical infection can occur when the guinea pig acts as a latent carrier capable of transmitting *Salmonella* to other animals and humans (Fournier et al., 2015). Recent studies have reported that *S.* Typhimurium is the main etiological agent isolated from cases of salmonellosis in guinea pigs (Marcelo M et al., 2017). Despite its impact on animal health and potential zoonotic source, there is no information on the lineage of *S.* Typhimurium that infects guinea pigs. The specific features of the livestock production system in low and middle-income countries (LMIC), such as animal overcrowding, informal trade, and deficiencies in biosecurity and veterinary service access, promote the emergence of recurrent outbreaks, raising the risk of zoonotic infections and resistance pathogen emergence (Ikhimiukor et al., 2022). A one-health approach is necessary to study the impact of guinea pig salmonellosis on the health of the Peruvian population using modern genomics tools. The objective of this work was to perform a genomic analysis of *S.*

Typhimurium isolates obtained from guinea pigs in four farms in Lima-Peru and compare them with available genomic sequences of S. Typhimurium from human origin from Peru.

2.2. METHODS

Bacterial Isolates

In this study we analyzed 11 *S.* Typhimurium isolates collected from guinea pigs reared in four different farms located in Lima, Peru, between 2015–2016 (Table A2.1). The isolates were recovered from liver samples obtained from salmonellosis clinical cases. *Salmonella* isolation was performed according to the International Organization for Standardization (ISO) 6579:2002. Liver tissue fragments were aseptically transferred to buffered peptone water and incubated at 37 °C for 18 h. Enriched cultures were transferred to Rappaport Vassiliadis Soya (RSV) broth and then incubated at 42 °C for 24 h. An aliquot was streaked onto XLD (Xylose Lysine Deoxycholate) agar and incubated at 37 °C for 24 h. Presumptive *Salmonella* colonies were confirmed as *S.* Typhimurium by PCR reaction using primers according to (Jamshidi et al., 2010). All isolates were stored at −80 °C in 25% glycerol diluted in BHI broth.

Whole Genome Sequencing and Assembly

Confirmed *S.* Typhimurium isolates were selected for sequencing. Genomic DNA was extracted and purified using the PureLink™ Genomic DNA Mini Kit (Invitrogen, Waltham, MA, USA) and concentration was measured using Qubit dsDNA HS assay (Invitrogen). For library preparation, 1 ng of DNA was required for the Nextera XT protocol, and subsequent sequencing was carried out using 2 × 250 bp reads on the Illumina Miseq platform (Illumina, San Diego, CA, USA) at the Laboratory of Molecular Genetics and Biology, Faculty of Veterinary Medicine, UNMSM. The fastq files were retrieved and evaluated using FastQC v0.11.9 (Andrews S., 2010), the trimming of low-quality reads was performed with Trimmomatic v0.39 (Bolger et al., 2014) using the options LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:50, and de novo assembly was performed with SPAdes v3.14.1 (Bankevich et al., 2012). The genome completeness and contamination parameters

to evaluate the assembly quality were calculated with the checkM software (Parks et al., 2015). Finally, the Prokaryotic Genome Annotation Pipeline (PGAP) version 14 April 2022.build6021 was used for annotation, using default parameters.

Phylogenetic and Genetic Diversity Analysis of *S.* Typhimurium from Guinea Pigs

In order to phylogenetically characterize *S.* Typhimurium genomes isolated from guinea pigs, we used 21 reference genomes representing different lineages of *S.* Typhimurium isolated from different hosts and countries and 5 genomes from Peruvian human isolates. These genomes were downloaded from the GenBank database, and their information is detailed in Table A2.2. We used Snippy v 4.6.0 (https://github.com/tseemann/snippy) (accessed on 1 June 2022). to generate a core genome alignment. The complete genome sequence of S. Typhimurium LT2 was used as a reference strain (Accession number: NC_003197.1). Gubbins v2.4.1 software (Croucher et al., 2015) was run for five iterations to remove recombinant regions before the phylogenetic reconstruction. From the core-genome alignment, we reconstruct a maximum-likelihood tree using IQ-TREE v1.6.12 (Nguyen et al., 2015) based on a GTR + Gamma nucleotide substitution model with 1000 bootstrap replicates. Tree visualization and annotation were created using ggtree v3.0.4 package (Yu et al., 2017). We computed pairwise SNP distances between genomes from the core-genome alignment using SNP-dists v0.6 (https://github.com/tseemann/snp-dists) (accessed on 3 June 2022).

MLST, Serotype, Prophage, Plasmid, and Antibiotic Resistance Genes [ARGs] Profiles Prediction and Comparative Genomics

We used *mlst* v2.19.0 tool (https://github.com/tseemann/mlst) (accessed on 16 May 2022) to determine multilocus sequence type [MLST] using the *Salmonella enterica* scheme from PubMLST. Serovar prediction was performed using SeqSero2 in the default k-mer-based mode (Zhang et al., 2019c; Banerji et al., 2020). ARGs, virulence genes, and plasmid replicon types were annotated using the Resfinder, Virulence factor database [VFDB], and PlasmidFinder databases with the ABRICATE v. 1.0.1 tool

(https://github.com/tseemann/abricate) (accessed on 10 June 2022). ABRICATE was run using a minimum DNA identity of 80% and a minimum coverage of 80%. We searched all genomes for phage elements using the PHASTER database (Arndt et al., 2016). We took the "intact" phage elements as defined by a phage score of > 90 and "questionable" phage with a score of > 80. These sequences were checked manually and compared between genomes using BLASTN to evaluate sequence conservation. The genomic comparison was performed using BLAST Ring Image Generator (BRIG) (Alikhan et al., 2011). The genetic contexts of ARGs detected in genomes were performed in R using the ggplot2 and gggenes R packages (Wickham, H., 2016).

2.3. RESULTS

Genomic Characteristics and Diversity of S. Typhimurium Isolates Obtained from Guinea Pigs

We sequenced the whole genome of 11 *S.* Typhimurium isolates from guinea pigs that died of salmonellosis in four farms in Lima-Peru, sampled between 2015 and 2016 (Table A2.1). The genome size of these eleven sequences ranged between 4.85–5.10 Mb and %GC content was between 52.08–52.22%. The assembly quality assessed with the completeness and contamination parameters from checkM software was >99% and <1% for all genomes, respectively (Table 2.1). *In-silico* MLST typing using *S. enterica* schema from pubMLST showed that all genomes from guinea pigs belong to the ST19 type, the most prevalent genotype in serovar Typhimurium. To obtain a good resolution on *S.* Typhimurium diversity on guinea pig farms, the H100 level of the Hierarchical Clustering of Core Genome Multi-Locus Sequence Typing (HierCC-cgMLST) scheme from Enterobase was used to categorize our isolates. The 11 isolates were classified into two distinct HierCC-HC100 clusters: 9 isolates belonging to the HC100_9757 cluster and 2 to the HC100_9460 cluster.

Table 2.1. Genomic statistics of 11 *S.* Typhimurium isolates from guinea pigs in Lima, Peru.

Isolate	Biosample ID	Average coverage	Number of contigs	Number of bases	Number of CDS	%GC	MLST	completeness *	contamination *	HierCC HC100
SMVET11	SAMN28944802	61.012	44	4,851,410	4,514	52.23	19	100	0.39	9460
SMVET12	SAMN28944803	76.632	45	4,902,895	4,571	52.14	19	100	0.08	9757
SMVET19	SAMN28944804	93.268	61	5,088,465	4,780	52.08	19	100	0.39	9757
SMVET20	SAMN28944805	106.307	39	4,849,526	4,514	52.23	19	100	0.39	9460
SMVET21	SAMN28944807	83.118	44	5,020,376	4,699	52.10	19	100	0.08	9757
SMVET22	SAMN28944808	115.677	54	5,095,938	4,803	52.07	19	100	0.08	9757
STc10	SAMN28944806	30.112	175	5,046,798	4,725	52.22	19	100	0.08	9757
STc12	SAMN28944811	36.224	65	5,017,046	4,694	52.13	19	100	0.08	9757
STc8	SAMN28944810	31.676	80	4,915,140	4,563	52.09	19	99.69	0.15	9757
STc9	SAMN28944809	61.838	39	4,891,950	4,548	52.15	19	100	0.08	9757
VET1	SAMN28944812	316.099	86	4,910,420	4,557	52.12	19	100	0.9	9757

*Values calculated by CheckM software

A core-genome alignment of the 11 isolates from guinea pigs and the *S.* Typhimurium LT2 reference strain was used to construct a matrix of pairwise SNP distances to explore genetic diversity between the S. Typhimurium isolates (Figure A2.1). The SNP distance between the two clusters (HC100_9757 and HC100_9460) ranged from 590 to 623 SNPs with an average of 608 SNPs, compared to the distance between HC100_9757 and the LT2 reference strain (576 to 606 SNPs) and between H100_9460 and LT2 (478 and 471 SNPs), suggesting a large divergence between these two clusters of guinea pig isolates (Figure A4.1). We also evaluated the intra-H100-cluster diversity; whereas two HC100_9460 isolates were closely related (3 SNPs difference), the HC100_9757 cluster contains important diversity

between isolates (2-200SNP). We observed two sub-clusters in HC100_9757 that corresponded to the HC50-67422 and HC50-9757 classification; the distance between these sub-clusters was 188-225 SNPs whereas intra-cluster distance was less than 100 SNPs (Figure A2.1).

Molecular serotyping using SeqSero2 software revealed that two HC100_9460 (SMVET11 and SMVET20) isolates lacked part of the fljAB operon sequence and were typed as monophasic variant (1,4,[5],12:i:-). Local inspection of the *fljAB* operon in HC100_9460 isolates shows that the fljA and fljB genes were replaced by a phage sequence that we named STGP_Φ2 (~90 kb). Instead, the other lineage identified in guinea pigs, HC100_9757, contained both intact genes (*fljA* and *fljB*) but an inversion in the hin gene was observed (See Figure A2.2).

Phylogenomic Analysis of *S.* Typhimurium Isolated in Peru

We performed a phylogenomic analysis of 11 *S.* Typhimurium isolated from guinea pigs within a context of 22 additional *S.* Typhimurium complete genomes from diverse hosts and locations. We also included 5 genomes from human origins isolated in Peru that were downloaded from the GenBank database. The core-genome alignment length of the 37 genomes was 4,448,045 bp (90% of 4,951,383 bp LT2 reference strain). The phylogenetic tree constructed with 6015 SNPs in the core genome revealed two phylogroups [α and β] as described previously by (Bawn et al., 2020). All genomes from guinea pig origin were clustered into the α phylogroup, which includes isolates from humans and domestic animals (Figure 2.1). The two H100 clusters identified for guinea pig isolates (HC100_9757 and HC100_9460) are highlighted in the phylogenetic tree. These two lineages associated to guinea pigs were more distant to β phylogroups strains (>850 SNPs) and less distant from the LT2 reference strain (<600 SNPs) and phage type U288 (<670 SNPs); the SNP distance matrix of all genomes is depicted in Figure A2.3. Interestingly, two isolates from humans in Peru (sampled in 2012) also grouped within guinea pig lineages, suggesting that these lineages can infect

both hosts (Figure 2.1). We were unable to obtain clinical information from Peruvian human isolates to confirm the source of contamination.

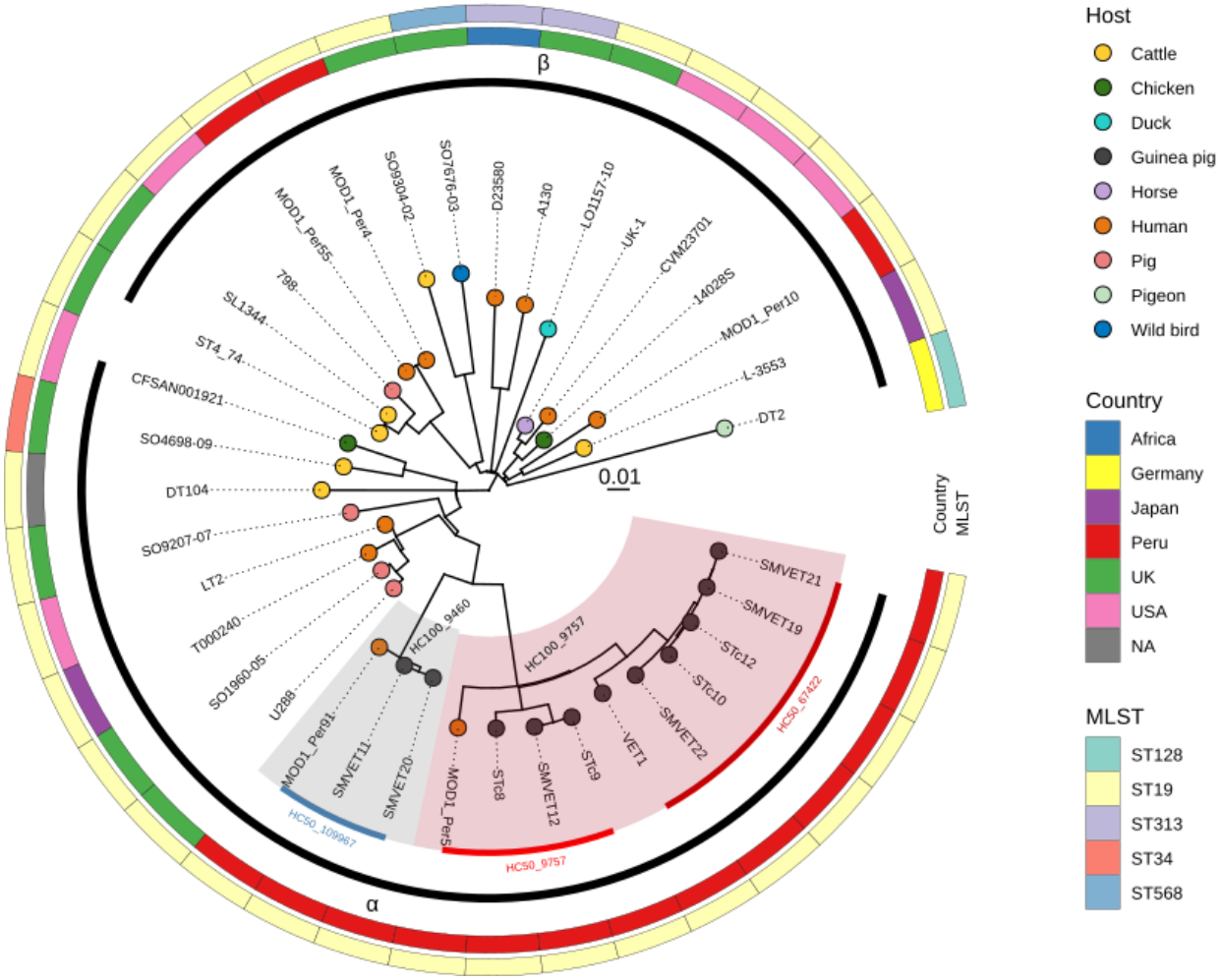

Figure 2.1. Phylogenetic tree of *S.* Typhimurium genome sequences isolated from guinea pigs in a context of 26 *S.* Typhimurium isolates from different hosts and locations including five sequences from Peruvian human sources. Two genetic clusters containing guinea pig isolates are highlighted in gray [HC100_9460] and red [HC100_9757]. The latter contains, in turn, two subclusters of H50: HC50-67422 and HC50-9757, that are represented by dark red and red bars, respectively.

ARGs, Plasmid, and Virulence Factor Repertoire of *S.* Typhimurium from Guinea Pigs and Humans in Peru

Since *S.* Typhimurium is an important human pathogen, we investigated the presence of ARGs and virulence genes in Peruvian isolates. We detected the chromosomal *aac(6')-Iaa* gene for aminoglycoside resistance in all isolates from guinea pigs and humans. Surprisingly, the rest of the *S.* Typhimurium genomes from guinea pigs do not present other ARGs except the SMVET22 isolate that contains a resistance gene cassette with four ARGs: *ant(3')-Ia, dfrA15, qacE,* and *sul1*, that confer resistance to aminoglycoside, trimethoprim, quaternary ammonium, and sulfamide, respectively (Figure 2.2A). This antimicrobial resistance cassette was 9678 nucleotides in length and inserted into an IncI1 plasmid of ~90 kb. We detected transposases flanking these resistance genes: IS*1326* and TnAs3, belonging to the IS*21* and Tn3 families, respectively (Figure 2.2B).

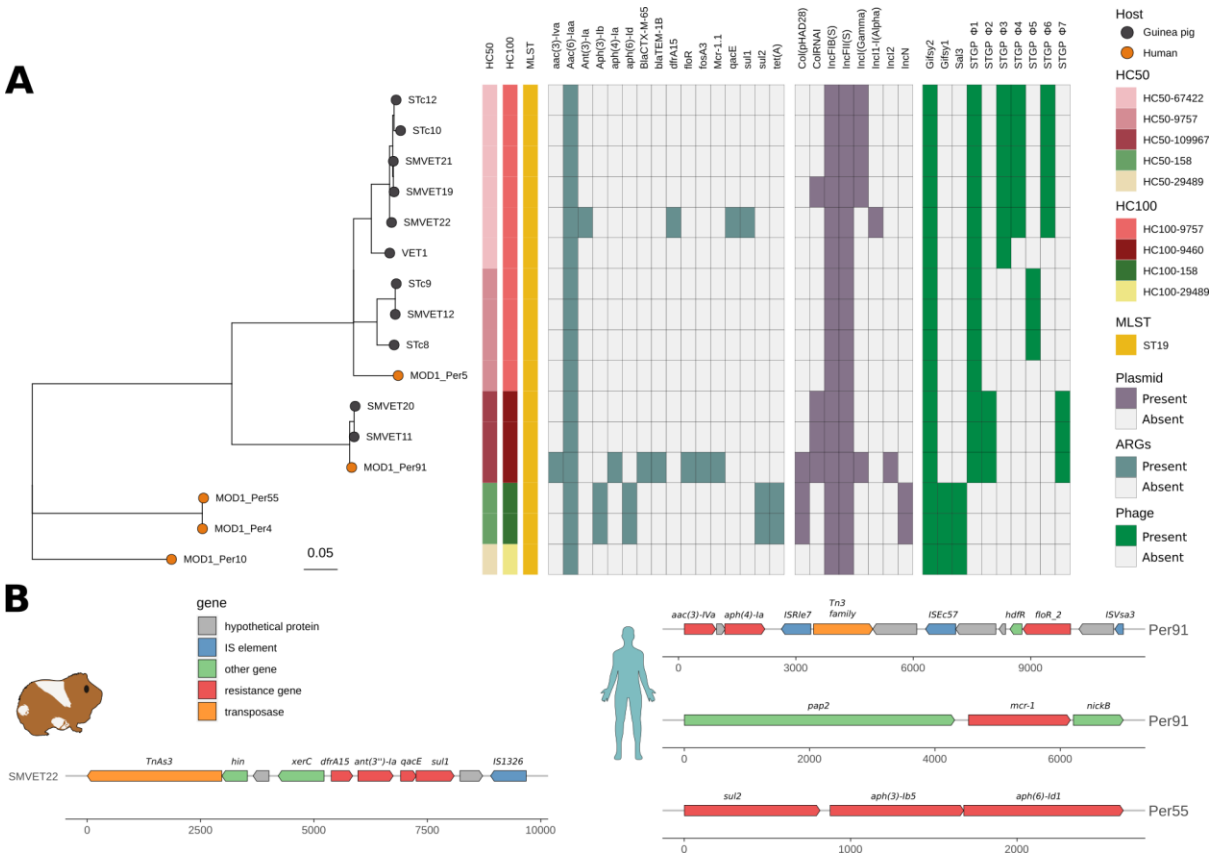

Figure 2.2. ARGs, plasmid, and phage profile in Peruvian *S.* Typhimurium isolates. (A) Phylogenomic tree based on SNP alignment of *S.* Typhimurium isolates including 11 from guinea pigs and 5 from human origin. Heatmap showing the presence/absence of ARGs (blue), plasmid replicon

(purple), and prophages (green). (B) Genetic context of ARGs detected in guinea pig and human isolates; the annotation of each gene in different colors is shown as per the key.

Three Peruvian human isolates contained at least four ARGs, including the β-lactam resistant genes *bla*CTX-M-65 and *bla*TEM-1B and the *mcr-1* colistin-resistant gene in the MOD1_Per91 isolate. We detected four mobile genetic elements (MGEs) in the context of the *aac(3)-IVa*, *aph(4)-Ia,* and floR resistance genes; ISV*sa3* and ISEc57 were flanking the *floR* gene (IS91 and IS21 family elements, respectively), while the ISRle7 (IS6 family element) and Tn3 family transposases were in the context of *aac(3)-IVa* and *aph(4)-Ia*. On the other hand, the *mcr-1* gene mobilized by an IncI2 plasmid was flanked by the pap2-nikB genes; we did not identify copies of ISA*pl1* in the neighborhood of the *mcr-1* gene (Figure 2.2B).

Eight types of plasmid replicons detected in Peruvian isolates belonged to the F type. All the analyzed genomes shared IncFII and IncFIB replicons; these correspond to a plasmid highly similar to the virulent plasmid pSLT from the LT2 reference strain, carrying the virulence operon (*spvRABCD*), but no ARGs. An Incly (Gamma) plasmid of ~93 Kb of size was detected in four guinea pig isolates of the cg-ST272 type; the annotations of this plasmid do not reveal the presence of any virulence or AMR genes. On the other hand, Col (pHAD28), Incl2, and IncN plasmids were found in the human origin isolates, of which IncI2 detected in the MOD1_Per91 human isolate harbored the *mcr-1* resistance gene (Table A2.3).

The *S.* Typhimurium genomes studied presented 144 different virulence genes associated with flagella, capsules, plasmids, adhesion systems, and type 3 secretion systems (T3SS) encoded in the pathogenicity islands SPI-1 and SP-2. All the genomes studied presented the main fimbrial operons necessary during tissue colonization: *fim* (encodes type I fimbriae), *lpf* (Long polar fimbriae), *bcf*, and the *csg* operon (Thin aggregative fimbriae—Tafi), except for the *pef* operon (Plasmid-encoded fimbriae), which was present in 81.1% of them. Additionally, the virulence genes located in the pSLT plasmid, involved in phenotypic virulence in rodents, were present in at least 29 of the genomes studied: *mig-14* (100%), *rck* (83.8%), and *spvBC* (78.4%). T3SS-related genes and pathogenicity islands (SPI-1 to SPI-5) were

present in all studied genomes; these include the *misL*, *sptP*, *avrA*, *pipB* genes; and the operons *inv*, *mgt*, *sic*, *sip*, *spa*, *ssa*, *ssc*, *sse*, *prg*, and *sic* (Figure A2.4).

Differential Gene Content in *S.* Typhimurium Isolates from Guinea Pigs Is Driven by Phages

We used PHASTER to examine phage content in all Peruvian *S.* Typhimurium genomes. We found the Gifsy2 (NC_010393) phage present in all Peruvian genomes containing the virulence factor *sodC1as*, expected for *S.* Typhimurium. Whereas the Gifsy-1 (NC_010392) and Sal3 phages were found only in human isolates, we detected five intact (we named these as STGP_Φ1–5) and two questionable (STGP_Φ6 and STGP_Φ7) phage elements in guinea pig isolates (Figure 2.2A). A STGP_Φ1 phage of ~26 kb was detected in all guinea pig isolates. This phage possesses partial sequence coverage with *Edwardsiella* spp. phage GF-2 (defined by PHASTER), commonly identified in the *S.* Typhimurium DT104 type. Interestingly, pairwise genome alignment using the BRIG tool showed gene content variation in guinea pig genomes driven by the differential presence of prophage sequences (Figure 2.3). HC50_67462 carried three phages: STGP_Φ3 (~38 kb), STGP_Φ4 (~35 kb), and STGP_Φ6 (~39 kb), absent in other guinea pig genomes. No virulence factor was identified in these phages (Figure 2.3A); instead, three genomes belonging to the HC50_9757 cluster carried exclusively STGP_Φ5, an intact phage of 42 kb. Finally, one intact (STGP_Φ2) and one partial (STGP_Φ7) phage were present in genomes from the HC50-109967 cluster (Figure 2.3B). A BLAST search for STGP_Φ7 in the NCBI virus database showed a match with a Myoviridae phage of *Escherichia* spp., ESSI2_ev239 (NC_049392.1) with >85% of identity and 76% of coverage.

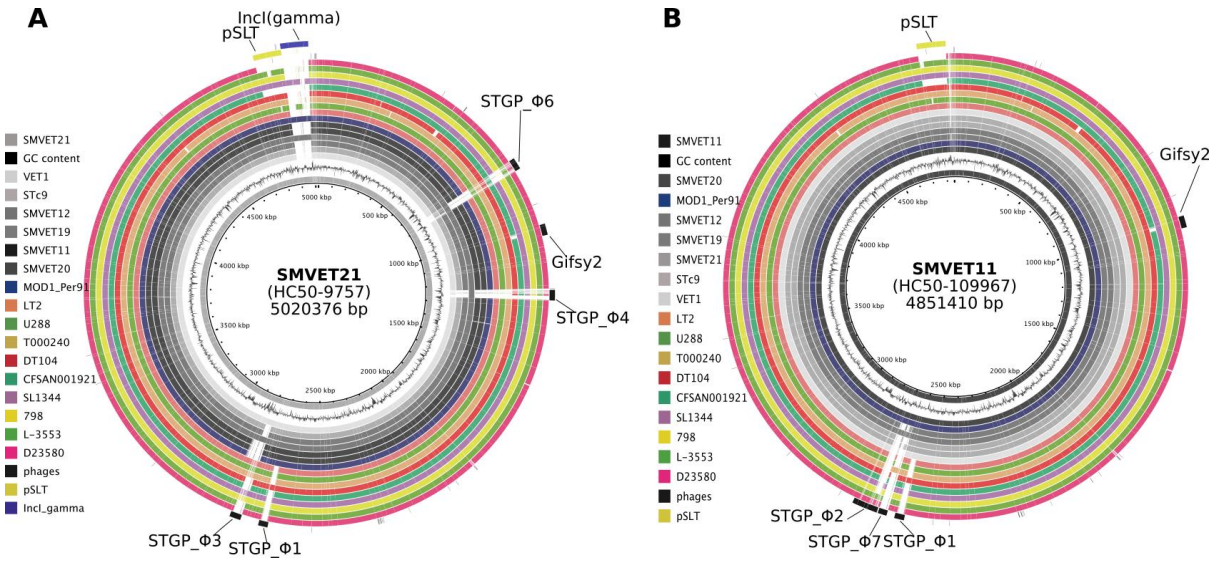

Figure 2.3. Comparative genomic showing differential prophage profiles in guinea pig isolates. Each ring represents the genome of a single *S.* Typhimurium isolated mapped to the (A) SMVET2 1 strain (H50-9757) or (B) SMVET11 strain (H50-109967); from the inner to the outermost ring, the first (innermost) ring shows the genome size in kbp, followed by SMVET21 (A) or SMVET11 (B), GC content (black), 6 guinea pig genomes (gray scale), and complete reference genomes from different host and location (colors). The predicted phage regions (black), pSLT (yellow), and Inclγ (gamma) (blue) plasmids are in the outermost ring.

2.4. DISCUSSION

In this study, we present a characterization based on the genomic sequences of *Salmonella enterica* subsp. *enterica* serovar Typhimurium isolated from guinea pigs in farms located in Lima-Peru during the period 2015–2016. In Peru, various studies showed the high susceptibility of guinea pigs to infection by *S.* Typhimurium, causing in them a disease characterized by high morbidity and mortality. It is considered the most critical pathogen that affects guinea pig production (Chauca L., 1997; Matsuura S. et al., 2010; Marcelo M et al., 2017). Salmonellosis has been recognized in guinea pigs for several decades, but recently its characteristics have been analyzed at the molecular level (Salvatierra R et al., 2018; Duran Gonzalez et al., 2021). A data set of 11 *S.* Typhimurium genomes from guinea pigs were analyzed based on their phylogenetic relationship with other *S.* Typhimurium isolates from

humans and various animal species. Additionally, we analyzed the presence of antibiotic resistance genes [ARGs] and mobile genetic elements, like plasmids and prophages.

Our analyzes reveal the presence of two very divergent genetic clusters of *S.* Typhimurium (HC100_9460 and HC100_9757 with ~600 SNPs of difference) circulating in guinea pigs reared in farms located in Lima, Peru. *In silico* serotype prediction revealed that the HC100_9460 lineage is a monophasic variant of *S*. Typhimurium. Many lineages of monophasic *S*. Typhimurium with different deletion types of the second-phase flagellar genomic region (*fljA*, *fljB* and *hin* genes) have been reported worldwide (Ido et al., 2014; Sun et al., 2020). The guinea pig HC100_9460 lineage lacks both the *fljA* and *fljB* genes due to the insertion of a phage sequence (STGP_Φ2) but maintains the *hin* gene. In Colombia and Spain, the most prevalent monophasic variants isolated from humans show the full deletion of the *fljAB* operon including the *hin* gene (Arrieta-Gisasola et al., 2020; Cuenca-Arias et al., 2020).

The HC100_9460 and HC100_9757 clusters are found within the phylogroup α, a clade composed of strains from well-characterized epidemics in domestic animals, while clade β harbors lineages that infect wild birds (Branchu et al., 2018; Bawn et al., 2020). In addition to differential host preferences, clade α isolates contain several lineages with multiple ARGs, while most clade β isolates contain few or no ARGs (Bawn et al., 2020).

Previous studies have used DNA-fingerprinting-based methods to evaluate the genetic variability of *S.* Typhimurium from guinea pigs. The BOX-PCR technique was used in 20 isolates from two farms in Lima in 2015 determining the presence of a single DNA band pattern that suggested a clonal population infecting guinea pigs (Salvatierra R et al., 2018). In contrast, the ERIC-PCR technique determined at least 7 different patterns in *S*. Typhimurium isolates from guinea pigs of the Peruvian coast collected between 2016 and 2018, showing greater diversity in samples taken in 2018 (Huamán et al., 2020). Although DNA fingerprinting techniques are difficult to reproduce and show a limited ability to differentiate lineages, our

results from genomic data confirm that there are at least two lineages circulating in guinea pig farms in Lima.

On the other hand, the phylogenetic analysis revealed that two (out of five) available genomes from human isolates from Peru were grouped within guinea pig clusters, suggesting the possibility of transmission of *S.* Typhimurium from guinea pigs to humans and vice versa. Although there is no evidence of transmission of *S*. Typhimurium to humans through contact with guinea pigs on farms or at the community level in Peru, a case of salmonellosis was reported in a US family after consumption of guinea pig meat in 2006. It is the first probable association between the consumption of guinea pig meat and human infection for non-typhoidal salmonellosis (Fournier et al., 2015). In 2017, an outbreak of *Salmonella* Enteritidis was reported in children through exposure to pet guinea pigs in several US states (Robertson et al., 2018). Although the probability of human salmonellosis infection by contact with pet guinea pigs is rare and unlikely (Dróżdż et al., 2021), it may be more common in farm environments where a large number of animals and salmonellosis outbreaks are more recurrent, as was described in poultry, pig, and cattle farms (Hendriksen et al., 2004; Stevens et al., 2009; Lanzas et al., 2010). Although the number of samples in this study limits the ability to predict dominant or more prevalent lineages, the fact that the same lineages have been found in both hosts suggests that these lineages may be epidemiologically important. However, it is necessary to expand the number of guinea pig and human genomes in different geographic regions and over longer periods of time to have a better view of the diversity and evolution of *S.* Typhimurium.

Despite the intensification of guinea pig farming in recent decades, the administration of antibiotics is generally used to treat infections and not as a growth promoter as is commonly used in poultry or pig farms. Therefore, resistant strains are not expected to be found in guinea pigs in the same magnitude as in poultry or pig farms. All *S*. Typhimurium genomes from guinea pigs and humans from the HC100_9757 and HC100_9460 clusters carried the gene for resistance to aminoglycosides *aac(6')-Iaa*. However, this gene is cryptic, meaning

that it does not produce phenotypic resistance (Salipante and Hall, 2003; Leon et al., 2018). In the isolate SMVET 22 from guinea pigs, a resistance genetic cassette was detected containing ARGs associated with resistance to aminoglycoside (*ant(6)-Iaa*), trimethoprim (*dfrA12*), sulfamethoxazole (*sul1*), and quaternary ammonium (*qacE*).

Aminoglycoside resistance genes are displayed in high frequency in *S*. Typhimurium strains isolated from human feces and animal-based food. Other ARGs frequently described were against tetracyclines, trimethoprim, beta-lactams, fluoroquinolones, and macrolides (Almeida et al., 2018; Rodrigues et al., 2020). Resistance to aminoglycosides, trimethoprim, and sulfonamide may be favored by the use of streptomycin, gentamicin, chloramphenicol, nitrofurans, and sulfa-trimethoprim for the treatment of infectious diseases such as salmonellosis in guinea pig colonies in Peru (Matsuura S. et al., 2010; Salvatierra R et al., 2018). Elevated phenotypic resistance to antimicrobials has been reported in previous studies. Streptomycin [aminoglycoside] resistance was identified in 30% of *S*. Typhimurium isolated from healthy and infected guinea pigs from farms in Lima (Peru) in 2015 (Salvatierra R et al., 2018), whereas sulfamethoxazole-resistant strains were found in 68% of isolates from farms of the same city during 2016 and 2018 (Huamán et al., 2020). Although the use of penicillin is not prescribed in guinea pigs because it is toxic to them (Madge, 1969), strains with a resistant phenotype to ampicillin have been found (Salvatierra R et al., 2018; Huamán et al., 2020).

Previous works in *S*. Typhimurium isolates from guinea pigs using PCR detected several ARGs including qnrB,D (quinolones), tetA,B,C (tetracyclines), and sul1,2 (sulfamethoxazole), with 23%, 71%, and 57% prevalence, respectively (Huamán et al., 2020), which raises concern about the possible increase in antimicrobial resistance in guinea pig farms. It is interesting that human isolate MOD1-Per91, sampled in 2012, phylogenetically related to the monophasic HC100_9460 lineage from guinea pigs, carried various resistance genes with clinical significance (*floR*, *fosA3*, $bla_{\text{CTX-M-65}}$, $bla_{\text{TEM-1B}}$, and *mcr-1* against florfenicol, fosfomycin, extended-spectrum β-lactamases, and colistin). Studies suggest that the transmission of ARGs by horizontal transfer in zoonotic bacteria may be of high risk in the

emergence and dissemination of bacteria carrying ARGs between humans and farm animals (Carroll et al., 2017).

Plasmids are common mobile elements in *S.* Typhimurium and the tracking is important because they can harbor genes encoding virulence factors, ARGs, and other important genes associated with environment adaptation. In this study, all strains isolated from guinea pigs and humans carried the IncFiB (s) and IncFII (s) plasmids (Figure 2.2). The FIB (IncFIB) and FII plasmids are commonly found in *S*. Typhimurium (Aljahdali et al., 2020) and may encode virulence factors and ARGs in Enterobacteriaceae species (Deng et al., 2011; Han et al., 2012). The Inclγ and ColRNAI plasmids were also identified in guinea pigs and human isolates. The Inclγ plasmid is mainly detected in Escherichia coli and is associated with horizontal transfer and dissemination of ARGs, such as ESBL (spectrum-extended β-lactamases) and AMPc β-lactamases (Carattoli et al., 2021). ColRNAI is also recognized for harboring ARGs against ampicillin, streptomycin, sulfamethoxazole, and tetracycline, but is present at a low frequency in *S. enterica* (Miller et al., 2020). Interestingly, strain SMVET22 carries a plasmid named IncI1 harboring several ARGs (*ant(3')-Iaa, dfra15, qarE* and *sul1*) in a cassette with a transposable element TnAs3. This plasmid has been described strictly in *Enterobacteriaceae*, commonly associated to *Salmonella enterica*, implicated with the spread of ESBL (Leverstein-van Hall et al., 2011), and related to the presence of transposable elements (Johnson et al., 2011). TnAs3, an insertion element, is abundantly present in plasmids and strongly associated with ARGs and their spread of them in external environments (Razavi et al., 2020).

The identification of phage elements in the genomes of *S.* Typhimurium of Peruvian origin revealed an important diversity of phages. Seven intact [STGP_Φ1–5] and questionable [STGP_Φ6 and STGP_Φ7] phages were identified among the 11 *S.* Typhimurium isolates from guinea pigs, in addition to Gifsy-2, which is found in most isolates of the serovar Typhimurium (Bawn et al., 2020). The presence of phage sequences in Peruvian isolates shows a lineage-dependent profile suggesting that variation in genetic content is due to

dynamic loss/acquisition of phages (Mottawea et al., 2018). In all the guinea pig genomes, the STGP_Φ2 phage was detected, which has a partial match with the *Edwardsiella* spp. bacteriophage GF_2 and is also commonly found in isolates of type DT104 (Parker et al., 2021), although no virulence genes or ARGs were detected.

The pathogenic ability of *S.* Typhimurium is mediated by genes encoding virulence factors involved in adherence, invasion, intracellular survival, and dissemination. All genomes analyzed in this study were very similar in terms of presence/absence of virulence genes, especially in genomes from clusters HC100_9757 and HC100_9460. The content of genes encoding virulence factors in isolates from humans and guinea pigs suggests that the mechanisms capable of causing infection and disease in both hosts are similar; however, in humans, the infection causes a mild and self-limiting disease (Gal-Mor et al., 2014), unlike the infection in guinea pigs, which causes a disease with high mortality rates and serious injuries to tissues such as the liver. Further studies must be carried out in order to clarify the pathogenesis of salmonellosis in guinea pigs.

In conclusion, the genomic analysis demonstrated the existence of two clades of *S.* Typhimurium circulating among guinea pigs in Lima, including a monophasic *S.* Typhimurium lineage. The close genetic relationship between the guinea pig isolates and some human isolates from Peru found in this study makes it necessary to conduct a more exhaustive investigation to establish possible epidemiological connections between guinea pig and clinical human isolates. The sustained increase in the production of guinea pig meat represents a growing risk to public health due to recurrent outbreaks of salmonellosis in these animals. Future works should consider a greater amount of genomic data from different sources including humans and other animals to determine the impact of *S.* Typhimurium from livestock production systems on human health as well as to monitor the genetic determinants of antimicrobial resistance.

2.5. CONTRIBUTIONS

A version of this chapter was originally published by Carhuaricra D et al. in Microorganism:

Carhuaricra Huaman, D. E., Luna Espinoza, L. R., Rodríguez Cueva, C. L., Duran Gonzales, C. G., Rosadio Alcántara, R. H., Setubal, J. C., et al. (2022). Genomic Characterization of Salmonella Typhimurium Isolated from Guinea Pigs with Salmonellosis in Lima, Peru. Microorganisms 10, 1726. doi: 10.3390/microorganisms10091726.

This work was carried out in collaboration with the Laboratory of Genetics and Molecular Biology (UBIGEM) of the Faculty of Veterinary Medicine of the Universidad Nacional Mayor de San Marcos, led by Dr. Lenin Maturrano. UBIGEM members performed the sampling, isolation, microbiological characterization and whole-genome sequencing of S. Typhimurium strains.

My primary contributions to this paper include (i) bioinformatics data analyses, (ii) results interpretation and (iii) writing the manuscript.

This work was carried out in collaboration with the Laboratory of Genetics and Molecular Biology (UBIGEM) of the Faculty of Veterinary Medicine of the Universidad Nacional Mayor de San Marcos, led by Dr. Lenin Maturrano. UBIGEM members performed the sampling, isolation, microbiological characterization and whole-genome sequencing of S. Typhimurium strains.

In this section my primary contributions include: (i) whole genome sequencing, (ii) bioinformatics data analyses, (iii) results interpretation and (iv) draft the manuscript.

3. Population genomics reveals the emergence of a *Salmonella* Typhimurium lineage in Peru linked to the intensification of guinea pig production

ABSTRACT

Salmonella enterica serovar Typhimurium is one of the most important Non-Typhoidal *Salmonella* (NTS) worldwide. In Peru, *S.* Typhimurium is the third most prevalent serovar in humans but it is the main pathogen of guinea pigs producing recurrent outbreaks and generating economic losses. The Peruvian guinea pig industry has experienced a dramatic increase in the last decades, posing a challenge for the emergence of new pathogens. In this study, we perform large-scale comparative genomics of *S.* Typhimurium from Peru in a regional context identifying an exclusive Peruvian lineage (named B6 lineage) prevalent in guinea pigs and humans. Phylogenetic dating supports that the emergence of the B6 lineage coincides with the rise in guinea pig numbers in Peru in the second half of the 20th century. B6 lineage evolution was dominated by loss-of-function mutations in several virulence genes including *oafA*, *fimC* and *bcsE* encoding for structures involved in host-cell interaction and host specificity (O5-antigen, fimbria and biofilm) potentially impacting virulence and pathogenicity. For example, the 7bp deletion in the *oafA* gene generates the loss of O5-antigen and is associated with immune escape. In the last decade, quinolone-resistant genes (*qnr*) and mutations in *gyrA* have become more prevalent in B6 isolates and Class-1 integrons carrying antimicrobial-resistant cassettes were identified in recent years. Taken together, our in-depth population genomic analysis highlights the evolution of a Peruvian *S.* Typhimurium lineage with crucial loss-of-function genetic events. We speculate that guinea pig farm intensification played a role in the emergence of this lineage which poses signals of host adaptation while maintaining zoonotic capacity.

3.1. INTRODUCTION

Non-typhoidal *Salmonella* (NTS) is one of the most important etiological agents of foodborne infections worldwide (Majowicz et al., 2010). Livestock are major reservoirs of *Salmonella* and animal-sourced foods play a significant role in human NTS (Li et al., 2019a). Among the more than 2600 distinct serovars, Enteritidis and Typhimurium are at the top of the list of prevalence in many hosts. Globally, *Salmonella* Typhimurium was the most commonly isolated serovar from people for many years, but in recent times, *S.* Enteritidis has surpassed *S*. Typhimurium in prevalence across various regions, particularly in Europe (Lan et al., 2009). In Peru, serovar Typhimurium is the third most common serovar isolated from humans after Infantis and Enteritidis (Garcia et al., 2019).

S. Typhimurium is considered a broad-host-range serotype because it is frequently isolated from many host species, including domestic animals and wildlife (Branchu et al., 2018). The epidemiology of *S.* Typhimurium suggests that several lineages (pathovariants) are highly host-adapted, for example, U288 in pigs (Kirkwood et al., 2021), definitive type (DT) phage 8 in ducks (Rabsch et al., 2002) and sequence type (ST) 313 associated with invasive NTS (iNTS) in humans (Kingsley et al., 2009). Anthropogenic selection and livestock intensification contributed to the emergence of new variants of *S*. Typhimurium in the last decades (Bawn et al., 2020; Stevens and Kingsley, 2021b).

Guinea pig farming in Peru has been facing recurrent *S*. Typhimurium outbreaks affecting the economy and health of farmers (Chauca L., 1997). The genomic analysis of *S*. Typhimurium isolated from guinea pigs revealed the presence of two ST19 lineages (> 600 SNP of distance) circulating in Lima-Peru, one of these being a monophasic variant (Carhuaricra Huaman et al., 2022). At the same time, human *S*. Typhimurium genomes from Peru became available, all of them of the ST19 genotype (Hurtado et al., 2022). A comparative analysis is needed to evaluate the genetic relatedness between isolates from both hosts.

Guinea pigs have represented an important meat source for the Andean population since pre-Hispanic times. Currently, Peru stands as the main global producer of guinea pig meat and their breeding represents a significant source of economic sustenance for many Andean and coastal families (Chauca Francia, 2013). The dramatic increase in the guinea pig population in the last decades represents a new niche favorable for the emergence of novel variants of *S*. Typhimurium.

In this study, we perform a populational genomics analysis of *S*. Typhimurium from Peru in the South American context. Our analysis demonstrates the emergence of a Peruvian lineage which coincides with the intensification process of guinea pig production in the last decades. The B6 Peruvian lineage shows specific signatures of pseudogenization in genes involved in the interaction with the cell host.

3.2. METHODS

Sampling and Isolation

We collected fecal samples of guinea pigs from six farms across the country (Lima, Pasco, Cusco, Cajamarca and San Martin regions) between 2019 and 2022 (Table A3.1). Samples were taken using sterile spoons and stored in containers at 4°C until bacteria isolation. Fecal pellets were transferred for enrichment to Rappaport-Vassiliadis broth and incubated at 42°C for 24h. After incubation, one loopful of each broth was streaked onto xylose lysine deoxycholate agar (XLD agar) and incubated at 37°C overnight. The plates were examined for morphologically typical *Salmonella* colonies. Among the positive samples, presumptive Salmonella isolate was picked and confirmed by PCR using primers according to (Jamshidi et al., 2010).

Whole Genome Sequencing and Assembly

Thirteen *Salmonella* isolates identified as *S*. Typhimurium by PCR were subject to genomic DNA extraction using Purelink Genomic DNA Mini Kit (Invitrogen, Waltham, MA,

USA). A DNA library was constructed with the Nextera XT DNA sample preparation kit and the associated Nextera XT Index kit (Illumina, San Diego, CA, USA). Whole genome sequencing of pooled samples was carried out using a MiSeq Illumina platform generating 2x250bp paired-end reads. Low-quality bases were trimmed with Trimmomatic v. 0.33 and the final quality of Fastq files was checked using FastQC. Genomes were assembled *de novo* using SPades v 3.1.4 with the default parameters. Genome coverage was determined using BBMap version 35.49. Assembly stats are summarized in Table A3.1. All genomes are available in the NCBI under the Bioproject accession number (Submission in process).

Genomic Dataset Collection

We retrieved publicly available genomes of *S*. Typhimurium from Peru and other South American countries. Recently, 90 genomes mainly from Peruvian human patients between 2004 to 2012 were presented by (Hurtado et al., 2022) and 11 genomes isolated from guinea pigs between 2015 to 2016 were presented by our group (Carhuaricra Huaman et al., 2022). Additionally, we downloaded 659 *S*. Typhimurium genomes from all countries in South America available in the Enterobase database (accessed in May 2023). Reference genomes from different lineages were also obtained from Genbank. Our full dataset consists of 700 genomes of *S.* Typhimurium including the newly sequenced genomes (see Table 3.1) and details of each isolate is available in (Table A3.2).

Table 3.1: List of genomes of *S.* Typhimurium by country used in this study

Country	Number of genomes	Source
Argentina	47	Enterobase
Brazil	226	Enterobase
Chile	199	Enterobase
Colombia	9	Enterobase
Ecuador	41	Enterobase
Guyana	4	Enterobase
Paraguay	5	Enterobase

Peru	146	Enterobase (n=32), Genbank (n=101), This study (n=13)
Reference	21	Enterobase
Suriname	1	Enterobase
Venezuela	1	Enterobase

Phylogenomics and population structure

We aligned 699 genomes to the reference *S*. Typhimurium genome strain SL1344 (RefSeq NC_003197.1) using Snippy (v4.3.7) (https://github.com/tseemann/snippy). Recombinant sites were identified with Gubbins using 10 iterations and removed from the alignment (Croucher et al., 2015). The core genome alignment was extracted with SNP-sites. A maximum likelihood (ML) phylogenetic tree was inferred using IQ-TREE2 v2 (Minh et al., 2020) modeled using a general time-reversible (GTR) substitution model +Γ and 1000 ultra-fast bootstrap replicates. Tree visualization was carried out with the ggtree package (Yu et al., 2017) in R environment 4.3. Population structure was investigated with rhierbaps (Tonkin-Hill et al., 2018) using the core genome alignment as input.

Temporal phylogeny construction

To investigate the temporal signal of the Peruvian B6 isolates a regression analysis was performed of the root-to-tip branch distances within the maximum likelihood phylogeny as a function of the year of collection using TempEst v1.5 (Rambaut et al., 2016). The resulting data was visualized in R using ggplot2 (Wickham, H., 2016). We use BEAST v1.10.4 (Drummond and Rambaut, 2007) to date evolutionary events of the B6 lineage. To calibrate the molecular clock, we incorporated isolation dates for each genome (indicated by year of collection). We employed a GTR+Γ substitution model and tested different combinations of molecular clock models (strict or uncorrelated relaxed with an underlying lognormal distribution) and tree priors (constant coalescent or exponential growth coalescent). Each

model was fitted using a Markov chain Monte Carlo with 300 million iterations, sampling every 20,000 iterations. We evaluated the adequacy of sampling from the stationary distribution by ensuring that the effective sample size (ESS) of key parameters exceeded 200. Since our alignments consisted of single nucleotide polymorphisms (SNPs), we explicitly specified the number of constant nucleotides in the model according (https://groups.google.com/forum/#!topic/beast-users/QfBHMOqImFE).

Pseudogene identification

Pseudogenes and fragmented genes were identified using the Pseudofinder tool (Syberg-Olsen et al., 2022). Pseudofinder's 'Sleuth' module was used to identify pseudogenes by querying (using BLASTN, e-value cutoff of 1E-5) reference gene sequences of *S.* Typhimurium LT2 reference strain against each isolate genome. We filtered SNPS/indels/insertions that cause nonsense or frameshift mutations and the disruption was manually confirmed through gene nucleotide sequence alignment using MAFFT (Katoh and Standley, 2013).

Inferring genetic relatedness using SNPs between isolates from different hosts

To study genetic relatedness between B6 isolates and identify possible clusters of transmission intra or inter-host species we use a threshold-based approach based on SNP distances (Thorpe et al., 2022). First, the core genome was generated with Snippy (https://github.com/tseemann/snippy) using the FD0185333 strain as a reference, after removing phage and plasmid sequences. Then, we computed pairwise SNP distances between genomes from the core-genome alignment using SNP-dists v0.6 (https://github.com/tseemann/snp-dists). Further, we constructed a network using *igraph* package (Csárdi and Nepusz, 2006) in which every isolate was initially linked to every other isolate, and then subsequently removed links if the SNP distance between isolates exceeded the selected threshold. We carried out this analysis using 4 different SNP thresholds: 5, 10, 20, and 40 SNPs.

Antimicrobial-resistant genes and plasmid identification

To identify ARGs, we utilized the AMRFinderPlus database (Feldgarden et al., 2021) with the abriTAMR v 1.0.14 pipeline (Sherry et al., 2023) using 90% identity and 90% coverage cutoff. Class 1 Integrons were predicted by IntegronFinder2.0 (Néron et al., 2022) using genome assemblies as input. Additionally, PlasmidFinder database (Carattoli and Hasman, 2020), was utilized to identify plasmid replicons. Graphical representations of Class 1 integrons and ARGs cassettes were produced with Easyfig (Sullivan et al., 2011).

3.3. RESULTS

Phylogenomics of South American isolates reveal the emergence of a *S.* Typhimurium lineage in Peru

A total of 146 Peruvian *S.* Typhimurium genomes were analyzed, of which 13 were newly sequenced genomes from guinea pig (Table A3.1), 102 genomes were retrieved from recently published data by Hurtado et al. (2022) and Carhuaricra et al., 2022, and the remaining were publicly available genomes recovered from Enterobase. To contextualize Peruvian isolates, we download additional 534 publicly available genomes from South American countries available in Enterobase (Figure 3.1A). Overall, our full dataset consists of 700 *S*. Typhimurium genomes (Table A3.2). All isolates belonged to the most common sequence type 19 (ST19)

A maximum likelihood phylogenetic tree was constructed based on the core genome SNPs of 700 *S.* Typhimurium and the outgroup *S.* Enteritidis P1029. The *S.* Typhimurium population structure consisted of eight BAPS (B1 to B8) with clusters B2, B3, B4, B6 and B8 being monophyletic (Figure 3.1B). Whereas most clades contain isolates from multiple geographic origins, the B6 cluster was composed of 90 genomes (89 from Peru, and 1 from Ecuador) mainly from human and guinea pig origin. The B6 lineage contained 61 % (89/146) of total isolates from Peru (Figure 3.1C). Further, this clade is characterized by a relatively

long branch which is indicative of a larger number of single base pair mutations has occurred prior to the emergence of this lineage (Figure 3.1B).

Out of the 38 guinea pig genomes from Peru, 36 belonged to the B6 lineage, while the remaining two genomes from Peru were grouped within the polyphyletic B5 cluster, along with eight guinea pig genomes from Ecuador. On the other hand, from a total of 92 human isolates from Peru, 45 belonged to the B6 lineage and the rest were distributed in the B3 (n=10), B4 (n=6), B5 (n=4) and B7 (n=27) lineages.

In our previous work (Carhuaricra Huaman et al., 2022), using only 11 isolates we report the presence of two ST19 (HC100_9460 and HC100_9757) clusters in guinea pigs from Lima-Peru, based on the H100 level of the Hierarchical Clustering of Core Genome Multi-Locus Sequence Typing (HierCC-cgMLST) scheme. The isolates classified as HC100_9757 were located into the B6 lineage which also contain isolates from other regions of Peru (Table A3.1).

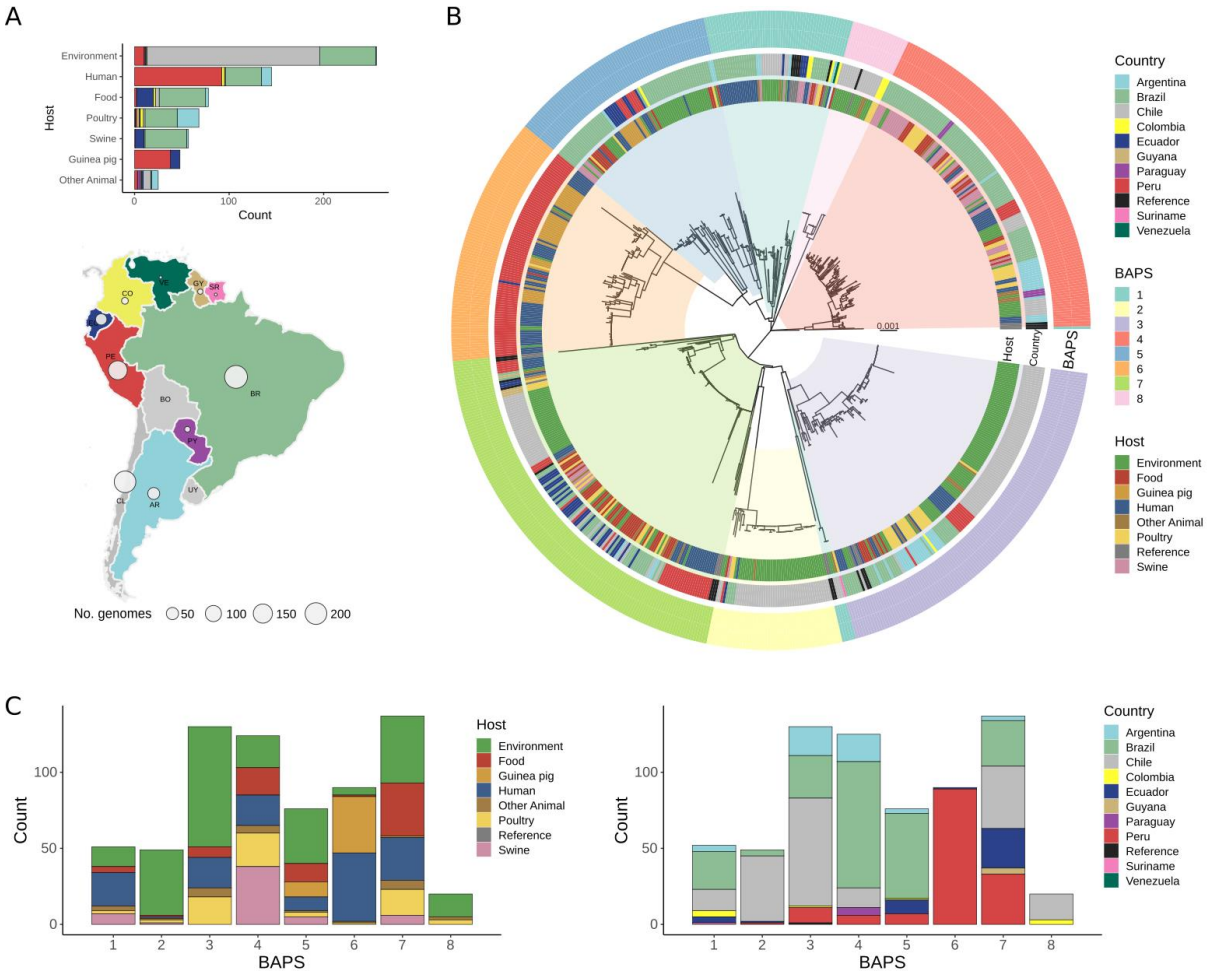

Figure 3.1. Genomic diversity of *S.* Typhimurium in South America. A. South American map showing the number of isolates per country and a barplot with the number of isolates by host source. B. ML tree based on core genome alignment of 700 isolates with clades highlighted by their respective BAPS cluster, as determined by Rhierbaps. Host origin, country of isolation and BAPS are represented by respective rings surrounding the phylogenetic tree. C. Barplots showing the number of isolates by each BAPS cluster and colored according to host and country source.

Peruvian *S.* Typhimurium B6 lineage emerged in the last century and coincides with the increase in guinea pig production

To estimate the age of emergence of the B6 Peruvian lineage we used the core-genome alignment of the 89 Peruvian isolates with known isolation dates ranging from 2004 to 2022. We removed prophage sequences and recombinant regions to recover the population clonalframe and improve the timed-measured approximation of the phylogeny. The temporal

signal was calculated with Tempest with R^2 of 0.3 and a correlation of 0.5 (Figure 3.2A), supporting the temporal analysis. Bayesian phylodynamic analyses using the highest supported model of a relaxed clock and Bayesian coalescent skyline tree prior revealed that the emergence of the B6 lineage occurred in 1968 (95% Highest Posterior Density: 1946–1989). Much sublineage diversification was more recent (1975-1990). The emergence and spread of B6 lineage in Peru coincides with the increase in guinea pig production in the last fifty years, the guinea pig population in Peru passed from nearly 374,000 in 1961 to more than 6.8 million in 1994 (Figure 3.2B).

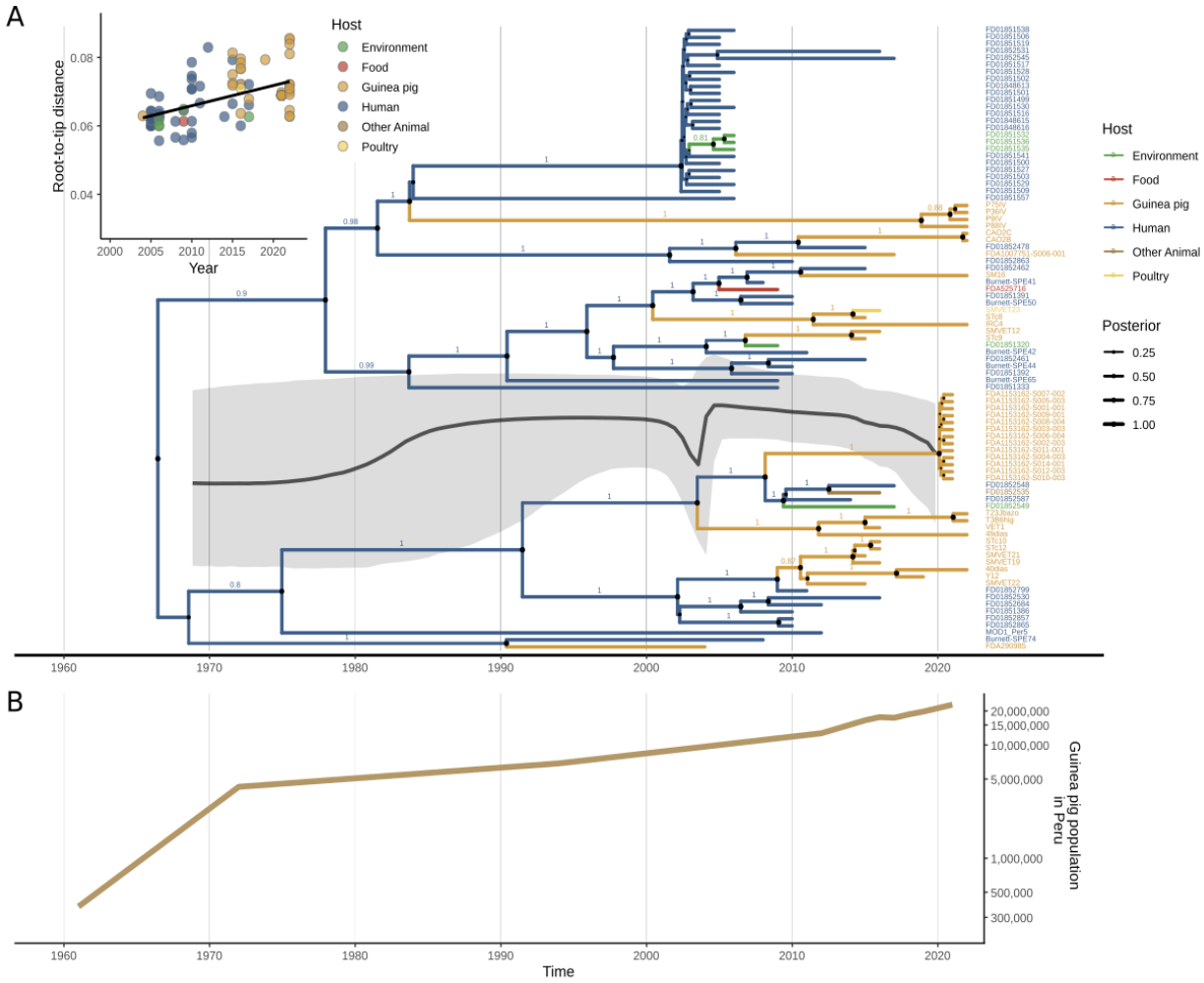

Figure 3.2. Time-scaled phylogenetic analysis of the emergence of the B6 *S.* Typhimurium lineage. A. Linear regression of root-to-tip genetic distance against the year of sampling calculated using Tempest v1.2 and time-scaled phylogenetic tree of 89 Peruvian isolates Colours represent different host sources according to the legend on the left of the tree. Skyline plot (black and gray lines)

is superimposed. The number on the branches represents the posterior probability. Branches are weighted by the support for the host changes; thicker branches have higher support. B. Graph of the estimated number of domestic guinea pigs in Peru from 1961 to 2021.

Pseudogene accumulation occurred previous to the divergence of B6 *S.* Typhimurium lineage

We identified gene sequences with nonsense/frameshift mutations that were present in almost all B6 isolates and practically absent in the rest of South American isolates: *fimC* (10bp insertion) encodes for a type 1 pilus chaperone FimC; *bcsE* (W17*) encodes a cyclic di-GMP binding protein; STM0458 (deletion of 678 bp position) encodes a cysteine synthase, *agp* (deletion at the position 276) encodes a glucose-1-phosphatase, STM2789 (S32*) encodes a glutarate-dioxygenase, and *oafA* (7bp deletion) encodes an acetyltransferase. The last one was also detected in isolates of different lineages (Figure 3.3A and Table 3.2).

The disruption of the *oafA* gene has been associated with the loss of the O-antigen in *S.* Typhimurium O:5-negative. Whereas *fimC* plays a role in the fimbrial organization, the *bscE* gene is part of the *bcsEFG* operon involved in cellulose production. Overall, *oafA*, *fimC* and *bcsE* genes are involved in functions related to the formation of different bacterial surface structures (O-antigen, fimbria and biofilm), in consequence, disruption in these genes may affect the interaction of the bacteria with their host.

STM2789, STM0458 and *agp* genes have metabolic functions, the first two are associated with amino acid metabolism (Lysine degradation and cysteine metabolism) and the last is involved in carbohydrate metabolism (glycolysis/gluconeogenesis).

Furthermore, two hypothetical pseudogenes were present in B6 isolates and isolates grouped by a deeper branch (Figure 3.3A). Frameshift mutations in *rtcB* (deletion at the position 845) and STM0057 (deletion at the position 39). The *rtcB* gene produces a tRNA-splicing ligase and is associated with tRNA biogenesis function. STM0057 gene encodes a citrate-sodium symporter involved in organic anion transmembrane transporter activity.

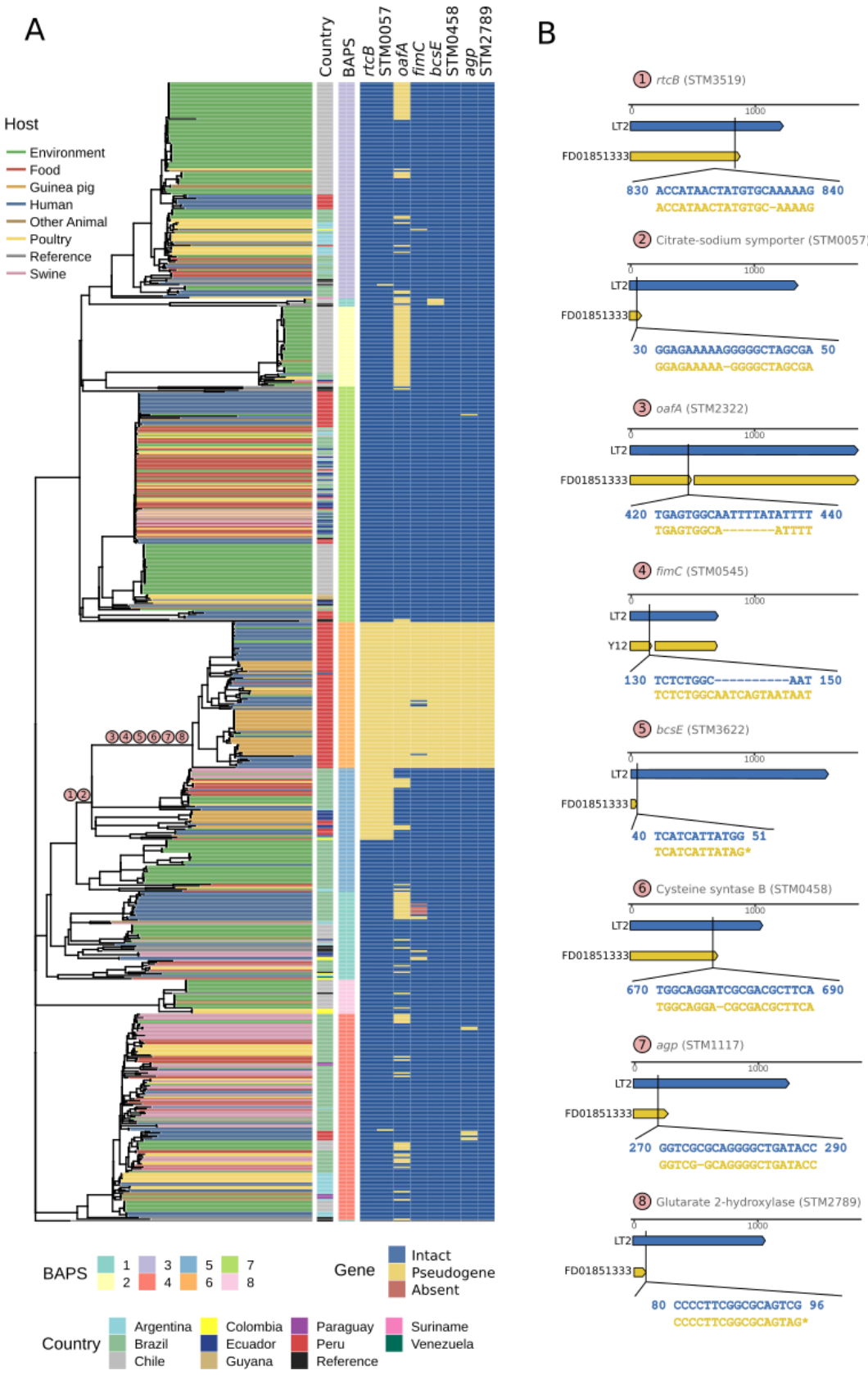

Figure 3.3. Pseudogene profile in B6 lineage isolates. A. Maximum-likelihood phylogenetic tree based on variation (SNPs) in the core genome of 700 *S*. Typhimurium. Each isolate is color-coded by

the host. The country of isolation and the BAPS clusters are represented in colored strips. The gene variants profile (intact or pseudogene) are indicated by bars color coded as indicated in the key. B. Graphic representation of mutational events of eight conserved pseudogenes in B6 lineage. Intact and pseudogene versions are represented by blue and yellow arrows and the exact position of mutation is shown for each gene in sequence alignment.

Table 3.2. List of pseudogenes and non-synonym mutations identified in B6 isolates.

Deletions are indicated in red, whereas insertions in blue.

	Gene name	Short name	Mutation type	mutation	AA posicion	Reference	Reference gene
1	RNA-splicing ligase	*rtcB*	frameshift/ stop codon	AAA	283	LT2	STM3519
2	Citrate-sodium symporter		frameshift	GCT	15	LT2	STM0057
3	acetylation of the O-antigen (LPS)	*oafA*	frameshift/ stop codon	ATTTTATATC	146	LT2	STM2232
4	fimbrial chaperone	*fimC*	frameshift/ stop codon	AATCAGTAATAGC	50	LT2	STM0545
5	Cyclic di-GMP binding protein	*bcsE*	stop codon	TGG (W) -> TAG (*)	13	LT2	STM3622
6	Cysteine synthase B		frameshift/ stop codon	GAT	226	LT2	STM0458
7	glucose-1-phosphatase	*agp*	frameshift/ stop codon	GCG	92	LT2	STM1117
8	Glutarate 2-hydroxylase		stop codon	TCG (S) ->TAG (*)	29	LT2	STM2789
9	Glycerol dehydrogenase		Non synonym mutation	GCC (A) -> GTC (V)	324	LT2	STM3529
10	Bifunctional aspartokinase/homoserine dehydrogenase 1		Non synonym mutation	GAA (E) -> AAA (K)	524	LT2	STM0002
11	putative hydroxymethyltransferas		Non synonym	GCC(A) -> ACC(T)	218	LT2	STM0019

	eb (chitinase)		mutation				
12	Phosphotransferase	*rcsD*	Non synonym mutation	CGT (R) -> AGT (S)	567	LT2	STM2269
13	NADPH-dependent curcumin reductase	*yncB*	Non synonym mutation	GTC (V) -> GCC (A)	155	LT2	STM1589
14	NADH-quinone oxidoreductase subunit L	*nuoL*	Non synonym mutation	GTG(V) -> GGG(G)	478	LT2	STM2318
15	3-isopropylmalate dehydratase small subunit 1	*leuD*	Non synonym mutation	GCA (A) -> ACA (T)	24	LT2	STM0110

Core SNP distance revealed the genetic closeness of human and guinea pig isolates within B6 lineage

The maximum pairwise distance between B6 isolates was 204 SNPs and the average pairwise distance was 119 SNP (Figure 3.4A). We use a network approach to cluster isolates with SNP cut-offs of 5, 10, 20, and 40 (Figure 3.4C). With 5 SNP cut-off, we identified nine clusters, the larger cluster was composed of 21 human and 3 environment isolates from six different regions during 2005-2009, and the second larger cluster was composed of 13 guinea pig isolates taken in 2021; two clusters with 4 and 3 isolates from guinea pig were identified both belonging to two farms in Lima sampled between 2016-2022 and one cluster with 3 isolated from Cusco in 2022. Interestingly, two isolates from guinea pig and chicken from Lima were closely related as 3 SNPs. With 10 SNP cut-off, one cluster consists of seven guinea pig isolates and one human isolate, all isolated in Lima during 2012-2016.

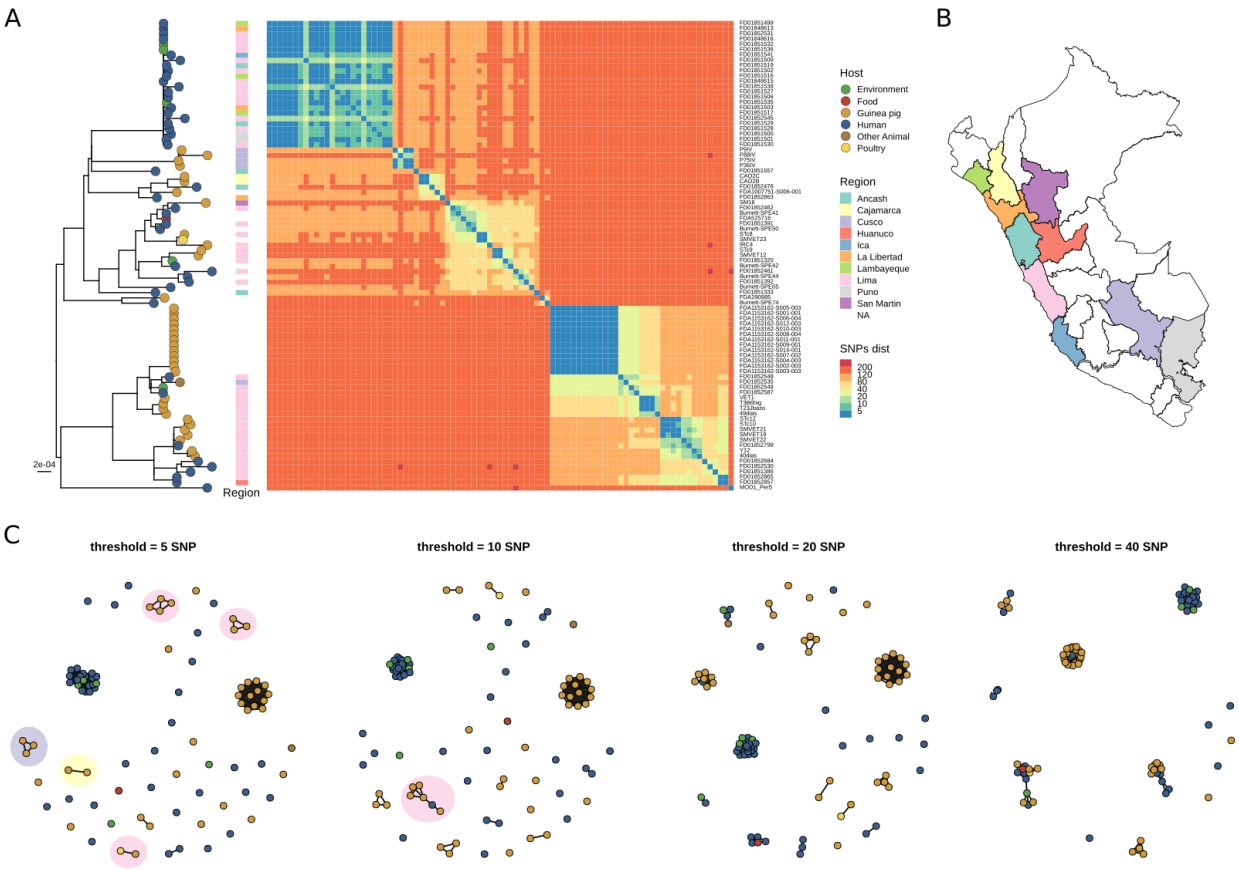

Figure 3.4. SNP distance and putative transmission clusters between B6 isolates. A. Maximum-likelihood phylogenetic tree of 89 Peruvian isolates of B6 lineage. Tip points represent the host source. The region of isolation is represented by a column. The Heatmap of the SNP distance matrix is represented in a gradient with different cut-offs. B. Map of Peru indicating the regions from where samples were taken. C. Networks of putative transmission clusters where each node represents an isolate colored according to the host, each link connects two nodes if the distance between both isolates is less than the chosen SNP distance threshold.

Genetic determinants of AMR are more prevalent in most recent isolates

Genes and mutations associated with antimicrobial resistance were predicted in 89 B6-lineage isolates using AMRfinderPlus. Fifteen different mobilizable ARGs were predicted conferring resistance to six antibiotic classes including aminoglycoside, chloramphenicol, quinolone, sulfonamide, tetracycline and trimethoprim. Additionally, we found four different non-synonymous mutations in the *gyrA* gene associated with quinolone resistance (Figure

3.5A). Most isolates do not contain any genetic element associated with AMR (n=64/89), especially those that were isolated before 2013, which are primarily of human origins (n=39/64). Interestingly, the most common AMR genetic signature is associated with quinolone resistance (n=18/89) in isolates since 2014 (Figure 3.5B). Only five isolates contain genetic elements that confer resistance to at least 3 antibiotic classes (Figure 3.5C), of them three were from guinea pigs isolated between 2015 and 2022, including the 49dias isolate which contained eight ARGs to six different antibiotic classes. Two isolates (SM16 and AMVET22) harbor Class-1 integrons carrying multiple ARGs inserted in IncI1 plasmids (Figure 3.5).

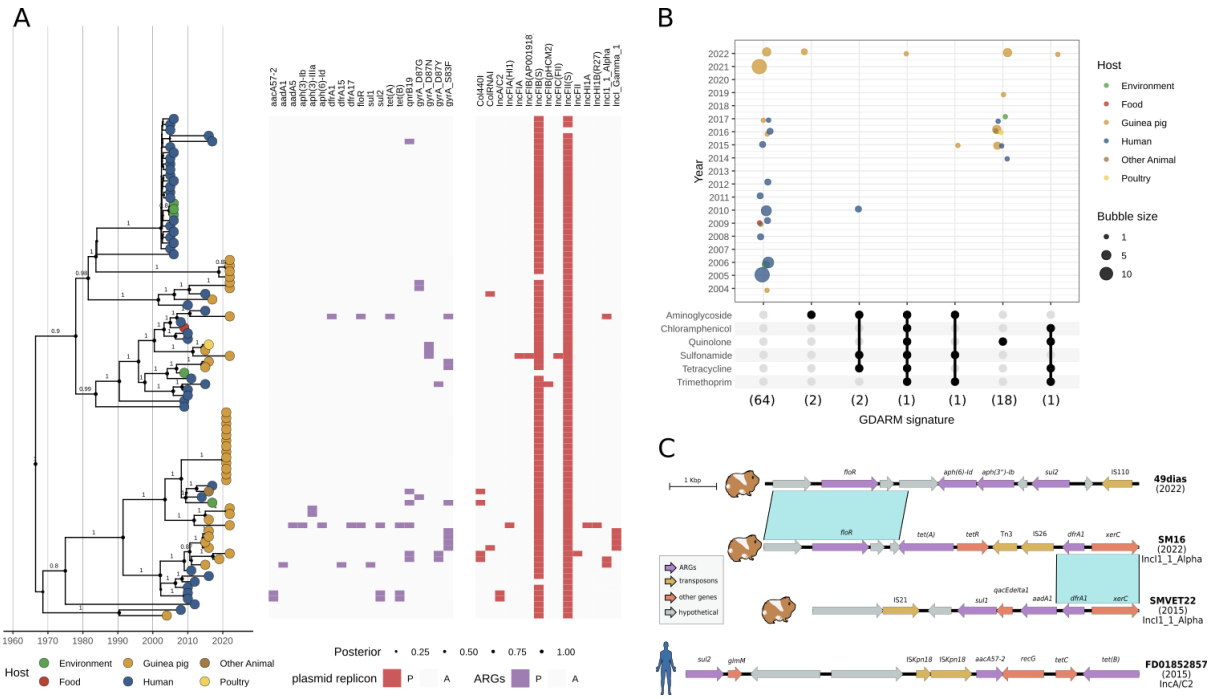

Figure 3.5 Profile of genetic determinants of antimicrobial resistance in B6 lineage. A. ML. phylogenetic tree with tip points color representing host source. The presence (P)/absence (A) matrix of AMR genetic elements and plasmid replicons are represented by purple and red bars. B. Temporal AMR trends in *S*. Typhimurium B6 lineage (2004–2022). The combination matrix illustrates the predicted AMR patterns of *S*. Typhimurium. In this matrix, genome-predicted resistance is denoted by black circles. The vertical arrangement of these black circles represents the resistance profile. The count of isolates for each resistance profile is specified in parentheses. The bubble plot presents the proportional count of isolates (bubble size) corresponding to each resistance profile (combination

matrix) over different years (y-axis). The bubbles' colors indicate the host sources, aiding in the identification of host-specific AMR trends. C. Representation of antimicrobial resistance gene clusters in four genomes with more than 3 ARGs. Arrows represent coding sequences with ARGs showed in purple.

3.4. DISCUSSION

Salmonella Typhimurium is the third-most common serotype associated with human salmonellosis in Peru after serotypes Infantis and Enteritidis (Garcia et al., 2019; Quino et al., 2020), but, it is the main serotype isolated in cases of salmonellosis in guinea pigs (Marcelo M et al., 2017). After the analysis of 146 *S*. Typhimurium genomes from Peru contextualized in 554 genomes from other South American countries, here we describe the emergence of a *S*. Typhimurium lineage (designated as B6 lineage) composed uniquely of isolates from Peru. The B6 lineage harbors 61% (n=89) of all *S*. Typhimurium Peruvian isolates, including 50% (45/92) of human isolates and almost all guinea pig genomes from Peru (34/36). Whereas human isolates from Peru were distributed in different clusters across the phylogeny, guinea pig isolates were almost exclusive of the B6 lineage. The broad diversity of human isolates in Peru may be due to the introduction of lineages by travelers or contaminated imported food (van der Bij and Pitout, 2012; Ashton et al., 2017). The international trade of animals for production can lead to the intercontinental dispersion of pathogens as recently described in the poultry supply chain and their influence in the global spread of *S*. Enteritidis (Li et al., 2021). Peru is the lead producer of guinea pigs worldwide and does not import guinea pig breeds; consequently, a geographic-specific clade is expected (Fenske et al., 2019). Based on the analysis of 11 *S.* Typhimurium genomes from guinea pigs, we previously reported two different genetic groups circulating in Peru (Carhuaricra et al., 2022). In the present analysis, the 48 guinea pig genomes from Peru (n=36) and Ecuador (n=12) confirm that these are distributed in two genetic clusters: B6 lineage, where most Peruvian guinea pigs belong and a clade into B5 cluster composed of 10 Ecuadorian and two Peruvian isolates. The last was

previously described as a monophasic variant (1,4,[5],12:i:-) lacking the *fljA* and *fljB* genes (Carhuaricra Huaman et al., 2022).

B6 lineage has been predicted to have emerged in the second half of the 20th century (1968, IC95% 1947-1989). Dating the emergence of B6-lineage provides clues about the influence of guinea pig intensification on their evolution. Guinea pigs were domesticated in 2000-6000 B.C. (Lord et al., 2020) and for millennia have been raised and consumed throughout the Central Andean (Morales, 1994; Chauca L., 1997). In 1961, the estimated population of guinea pigs was 374,000 individuals, raised basically for home consumption under a family system production. Since 1966, the National Institute of Agricultural Innovation (INIA) has carried out a project for the modernization of guinea pig production with the development of improved breeds and technifying the production systems turning guinea pig production from a subsistence activity into a profitable business in the highlands and the coastal region of Peru (Chauca L., 1997; Chauca Francia, 2013). As a result, the current guinea pig farming produced a 60-fold increase of the guinea pig population in last 60 years (the last estimate was 22,714,440 in 2021). The dramatic increase in animal numbers kept in high-density units and low genetic diversity favors the emergence and spread of zoonotic pathogens (Jones et al., 2013). Since the introduction of domesticated animals and the increased interaction between humans and animals during the neolithization process thousands of years ago, specific *Salmonella* serovars adapted to hosts have arisen (Key et al., 2020). The emergence of epidemic variants *S*. Typhimurium DT104, U288 and ST34 in Europe associated with livestock in the last decades illustrates that *Salmonella* evolve in farmed animals over short timescales (Bawn et al., 2020; Kirkwood et al., 2021; Stevens and Kingsley, 2021b). For instance, the use of antibiotics as a growth promoter has been associated with the emergence of the multidrug-resistant *S*. Typhimurium ST34 lineage in pigs in 2005 which has become dominant in the UK (Petrovska et al., 2016; Tassinari et al., 2020). The significant increase in the guinea pig population over the last few decades represents a new niche which can be exploited by newly adapted lineages. Currently, salmonellosis is the

most important disease that affects guinea pig production causing high mortality and morbidity rates. The presence of *S*. Typhimurium has been reported even in healthy animals highlighting the high prevalence of this pathogen in guinea pigs.

The B6 lineage arises from the B5 polyphyletic lineage with a long branch, which indicates a rapid differentiation of the ancestral populations through the accumulation of mutations. Loss of function mutations or genomic degradation has been associated with a process of host adaptation in *Salmonella* (Holt et al., 2009; Langridge et al., 2015; Key et al., 2020). Several lineages of *S*. Typhimurium have evolved by accumulating pseudogenes such as the invasive ST313 in humans or lineages associated with wild birds such as DT4 in passerines (Bawn et al., 2020; Pulford et al., 2021). Here we show that nonsense and frameshift mutations have generated putative pseudogenes during the emergence of the B6 lineage.

The most relevant disrupted genes in B6 isolates were *oafA*, *fimC* and *bcsE* which are involved in the biosynthesis of surface and extracellular structures (i.e., O-antigen, fimbria and biofilm). Disruption of these structures may alter the interaction with the host cells potentially impacting the virulence and pathogenicity. The 7-bp deletion in the *oafA* gene disrupts O:5 LPS production, as previously documented in invasive *S*. Typhimurium (Van Puyvelde et al., 2022). This particular version was not limited to B6 isolates; instead, it appears highly prevalent throughout the phylogeny within our dataset (see Figure 3.4) suggesting that it has emerged multiple times and may potentially confer a selective advantage. A recent study detected IgA-escape *S*. Typhimurium mutants in the caeca of vaccinated mice after a challenge with a wild-type *S*. Typhimurium (Diard et al., 2021). Sequencing of these variants revealed the same 7 bp mutation in the *oafA* gene, indicating that one strategy employed by *S*. Typhimurium to evade adaptive immunity is the loss of the O-antigen. Several NTS vaccine candidates in development are O-antigen-based, including glycoconjugate vaccine (Micoli et al., 2018). Moreover, the loss of the O5-antigen can impact immunogenicity and vaccine development.

The *fimC* gene is required for the biosynthesis of type 1 fimbriae. This structure is involved in bacterial adhesion to the host cells and has been observed that variation in type 1 fimbriae affects *Salmonella* host specificity (Yue et al., 2015; Kolenda et al., 2019). Recently, genomic surveillance in the US identified a passerine *S*. Typhimurium lineage with unique pseudogenes including *fimC* suggesting that pseudogenization in virulence genes may have driven the emergence of the passerine-adapted *S*. Typhimurium (Fu et al., 2022). On the other hand, several studies have suggested that inactivation of *bscE* decreases significantly (Solano et al., 2002) or moderately (Fang et al., 2014) cellulose production in *S*. Typhimurium. The disruption in cellulose production may affect virulence as was described in a Shiga toxin-producing *E. coli* outbreak in 2011. The Stx-producing O104:H4 strains presented a C-terminal truncation of *bcsE* gene and were cellulose negative, this probably contributed to their higher virulence because “naked” curli fibers enhance inflammatory response (Richter et al., 2014). Recent work highlights that the intact *bcsE* gene plays an active role in biofilm production, motility and epithelial cell invasion, and when *bcsE* is disrupted, *S*. Typhimurium loses the invasion ability (Özdemir et al., 2021). While the studies mentioned above provide insights into the individual effects of disrupting the *oafA, fimC,* and *bcsE* genes on host interactions, predicting the combined impact of these mutant genes on the *Salmonella* phenotype remains uncertain. To comprehensively assess the consequences of these mutations on virulence and pathogenicity, it is imperative to conduct both *in vivo* and *in vitro* studies.

While *S*. Typhimurium infections in humans produce mainly gastrointestinal symptoms, *S.* Typhimurium strains circulating in guinea pigs can cause acute and fulminant infections, leading to death within a few hours after exposure. Experimental infections by *S*. Typhimurium VET1 strain (which belongs to the B6 lineage) have revealed the colonization of various organs, including the liver, spleen, and nervous system (Mejía et al., 2019). Enrofloxacin (fluoroquinolone) is frequently used to treat salmonellosis in guinea pigs (Ortiz G. et al., 2022). The overuse of this antibiotic in farms may explain why the most common genetic mechanism of resistance in guinea pig isolates is related to quinolones (presence of *qnrB19* gene and

mutations in *gyrA* gene). Additionally, we observe that most isolates (mainly from humans) before 2014 do not contain ARGs or mutations that confer resistance. However, the isolates are not equally distributed in time, whereas most human isolates in B6 lineage ranged between 2004 to 2009, the majority of guinea pig isolates were sampled after 2014.

Taken together, our data contribute to a better understanding of the evolutionary history associated with the emergence of a new *Salmonella* Typhimurium variant in Peru. Guinea pig intensification in the last decades may be the niche for the emergence and expansion of this variant that shows signals of host adaptation through loss-of-function events in virulent genes. Although it maintains zoonotic potential.

3.5. CONTRIBUTIONS

This work was carried out in collaboration with the Laboratory of Genetics and Molecular Biology (UBIGEM) of the Faculty of Veterinary Medicine of the Universidad Nacional Mayor de San Marcos, led by Dr. Lenin Maturrano. UBIGEM members performed the sampling, isolation, microbiological characterization and whole-genome sequencing of S. Typhimurium strains.

In this section my primary contributions include: (i) whole genome sequencing, (ii) bioinformatics data analyses, (iii) results interpretation and (iv) draft the manuscript.

Our intention is to prepare a manuscript based on the results of this chapter and submit it for publication in the months to follow.

4. Comparative genomic analysis of *Kerstersia gyiorum* isolates from free-living brown-throated sloth (*Bradypus variegatus*) at the Parque Estadual das Fontes do Ipiranga, Brazil

Abstract

Kerstersia gyiorum is a gram-negative bacterium isolated from various animals, including humans, where it has been associated with infections. Knowledge of the basic biology of *K. gyiorum* is essential to understand the evolutionary strategies of niche adaptation and how the organism contributes to infectious disease; however genomic data is very limited, especially from non-human hosts. In this work, we sequenced 12 *K. gyiorum* genomes isolated from healthy free-living brown-throated sloths (*Bradypus variegatus*) in the Parque Estadual das Fontes do Ipiranga (São Paulo, Brazil), and compared them with genomes from human origin in order to gain insights into genomic diversity, evolutionary history, and potential ecological differentiation and host specialization of the species. Phylogenetic analysis revealed a *K. gyiorum* population structured according to the host, with sloth isolates clustering together. Additionally, differences in genome size and genetic content suggest potential host-associated lineages. Despite sloth isolates coming from a single place, the intra-sloth population diversity of *K. gyiorum* was divided into three populations with differences of > 1000 SNPs between them, suggesting the circulation of various *K. gyiorum* clusters in sloths. Genes involved in mobilome and defense mechanisms against mobile genetic elements (MGEs) were enriched in the accessory genome as the main source of variation. An IncN2 plasmid exclusive of sloth isolates was identified and a total of five phage sequences in all *K. gyiorum* genomes. A search for defense mechanisms identified 14 different systems including CRISPR-cas and restriction-modification (RM) with strain-level variation; these elements may prevent further element flow and explain the low amount of MGE in *K. gyiorum* genomes. Gene content variation may be important for the adaptation to different niches. Collectively, this study furthers our understanding of population structure, diversity, and host adaptation of *K. gyiorum* strains.

Keywords: *Kerstersia gyiorum*, brown-throated sloth, comparative genomics, phylogenomics, population structure

4.1. INTRODUCTION

Kerstersia gyiorum is a Gram-negative coccobacilli bacteria classified into the *Alcaligenaceae* family (Coenye et al., 2003), and has been isolated from various patients with wounded limbs, respiratory, and urinary tract infections and other infectious diseases (Almuzara et al., 2012; Pence et al., 2013; Ogawa et al., 2016). Despite being isolated from various types of clinical samples, its pathogenic capacity and contribution to the development of infection have been little explored. Previous genomic analysis of *K. gyiorum* isolated from a Chinese patient allowed the identification of putative virulence factors and antimicrobial resistance genes (Li et al., 2019b).*K. gyiorum* is ubiquitously distributed in a variety of habitats, including other mammals, insects and soil (Gupta et al., 2012; Wan et al., 2016; Dalmutt et al., 2020; Shen et al., 2022). However, until now, only few *K. gyiorum* genomes from human origin are publicly available.

The brown-throated sloth (*Bradypus variegatus*) is an arboreal mammal from the Xenarthra superorder distributed throughout Central and South America inhabiting the Amazon, Caatinga and Atlantic Forest biomes (Hayssen, 2010). Sloths are considered important reservoirs of various zoonotic pathogens including *Leishmania, Trypanosoma cruzi, Anaplasma* spp*., Ehrlichia* spp., and arboviruses such as dengue, mayaro, oropouche, among others (Catenacci et al., 2018; Muñoz-García et al., 2019; Calchi et al., 2020; Sant'Ana et al., 2020). Anthropogenic activities produce rapid changes in the habitat available to wildlife intensifying the interaction between wild animals, humans and domestic animals, increasing the risk of pathogen transmission (Kruse et al., 2004). The increased interaction between wild animals and the human microbiome and exposure to antibiotics has been shown to decrease bacterial diversity in the wild host, at the same time as increasing the chance for the development of pathogenic strains and increasing the prevalence of antimicrobial resistance

genes (ARGs) (Fernandes et al., 2022; Jia et al., 2022). Therefore, it is important to know the genomic diversity of potential pathogens in wild animals that can be transmitted to humans and investigate their virulence and ARGs.

The Parque Estadual das Fontes do Ipiranga (PEFI) comprises an area of 527 hectares of an important remaining fragment of Atlantic Forest surrounded by an urban region in the city of São Paulo, Brazil (Barbosa et al., 2002). It is a critical biodiversity area where few studies aiming the conservation of fauna and flora are carried out, for example, the Bradypus Research Program. Within this research program, a study concerning bacterial diversity in the oral and rectal cavity of *Bradypus variegatus* was previously conducted and species such as *K. gyiorum* were first reported.

In this work, we sequenced and characterized *K. gyiorum* genomes isolated from free-living brown-throated sloth (*Bradypus variegatus*) from The PEFI in Brazil under the Bradypus Research Program and performed the first comparative genomic analysis for the species.

4.2. MATERIALS AND METHODS

Ethic statements

Animal capture and sample collection were conducted by a previous study with permission of Instituto Chico Mendes da Conservação da Biodiversidade (ICMBio) and Sistema de Autorização e Informação em Biodiverdidade (SISBIO) license number 49627 and it was approved by the scientific board of São Paulo Zoological Park Foundation.

Sample Collection and Bacterial Isolation

In a previous study, rectal swabs were obtained from wild brown-throated sloth (*Bradypus variegatus*) for culturable bacterial diversity analysis. Among the results, 12 strains of *Kerstersia gyiorum* were identified from 10 individuals.

Disk diffusion susceptibility test

The Kirby-Bauer disk diffusion test was performed to evaluate the susceptibility of *K. gyiorum* isolates to antimicrobials (Bauer et al., 1966). The inhibition zones of each isolate were measured and compared to the standard defined by Clinical and Laboratory Standards Institute (CLSI, 2021).

Evaluation of partial 16S ribosomal RNA gene (16S rRNA)

Bacterial fresh colonies growth in Tryptic Soy Broth (TSB) were prepared for DNA extraction using Wizard® Genomic DNA Purification Kit (Promega, cat#1125), according to the manufacturer's protocol. A PCR to amplify the V1 - V9 region of 16S rRNA was carried out using the set of primers: 27F (5'AGAGTTTGATCMTGGCTCAG 3') and 1401R (5'CGGTGTGTACAAGACCC 3' and the reaction were performed in 50µL final volume reactions comprising 1X Taq buffer (Tris-HCl 20mM and KCl 50 mM), 1.5 mM $MgCl_2$, 200 µM of dNTPs, 30-50ng of genomic DNA, 0.3 µM of each primer and 2U Taq polymerase (Invitrogen). The gene amplification was performed with a Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems®) and consisted of initial denaturation at 95°C for 2 minutes followed by 40 cycles of denaturation at 95° C for 2 minutes, annealing at 50° C for 1 minute and extension at 72° C for 3 minutes. An additional extension step at 72ºC for 10 minutes was performed. The amplified PCR products were purified using GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, cat# 27-9602-01) following the manufacturer's protocol. Sequencing reactions were performed on an automated sequencer, Model ABI 3730 DNA Analyzer (Applied Biosystems). To the purified PCR products (5.0 µL) were added 4.0 µL of BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and 1.0 µL primer for sequencing (0.5 µmol each). The primers used in the sequencing reaction were 27F (5'AGAGTTTGATCMTGGCTCAG 3'), 1401R (5'CGGTGTACAAGACCC 3') and 782R (5'ACAGGGTATCTAATCCTGT 3'). The sequencing program consisted of 25 cycles at 95° C for 20s, 50° C for 15s and 60° C for 60s.

Whole-genome sequencing and assemblies

DNA was extracted from pure culture using the Wizard® Genomic DNA Purification Kit (Promega, cat#1125). DNA was sent to Joint Genome Institute (JGI) for sequencing using Illumina technology. The 12 genomic assemblies of *K. gyorum* from sloths were annotated with the Prokaryotic Genome Annotation Pipeline (PGAP). Additionally, we retrieved 5 genomic sequences of *K. gyiorum* from Genbank public database as 08 November of 2022. The basic metrics of all sequences used for this study are shown in Table 4.1.

Average Nucleotide Identity (ANI) Assessment and Phylogenomic Analysis

The Average Nucleotide identity (ANI) was calculated for all pairs of genome sequences with the aim of determining if all isolates belong to the same species. We used the Python pipeline PyANI v0.2.7 (https://github.com/widdowquinn/pyani) to generate ANIm values using MUMmer for alignment. A heatmap was generated from the pairwise ANI values matrix using the *pheatmap* package in R v4.2.

For phylogenomic analysis, we used Snippy v4.6.0 (https://github.com/tseemann/snippy) to generate a core genome alignment. The complete genome sequence of *Kerstersia gyiorum* str. SWMUKG01 was used as a reference strain (Accession number: CP033936). We used Gubbins v2.4.1 (Croucher et al., 2015) to remove recombinant regions prior to the phylogenetic reconstruction. From the core-genome alignment, we reconstruct a maximum-likelihood tree using IQ-TREE v1.6.12 based on a GTR nucleotide substitution model with 1000 bootstrap replicates (Nguyen et al., 2015). ggtree v3.0.4 package was used for visualization and annotation of the phylogenetic tree (Yu et al., 2017).

Pangenome Reconstruction and Functional Annotation

The annotated genome assemblies were used as input for panaroo (Tonkin-Hill et al., 2020). Orthologous genes were grouped using sequence identity of > 95% and query coverage of > 70 % which determined the core and accessory genes of 17 *K. gyiorum* genomes (including public sequences). The pangenome was represented on a heatmap using

the presence/absence matrix of orthologous genes. Pangenome and core curves were generated from the presence/absence matrix of orthologous groups retrieved from panaroo using the R package vegan.

We used the Clusters of Orthologous Group (COG) database recently updated (Galperin et al., 2021) for functional annotation performing a conserved-domain search with RPS-BLAST against COG domains. COG categories were assigned to each orthologous gene of the core and accessory genome. If a gene was assigned to more than one COG category for functional annotation analysis, it was defined as 'ambiguous' category.

Characterization of MGE, defense mechanism systems, virulence factors and ARGs

We selected two command-line tools to identify phage sequences in the *K. gyiorum* genomes. PhiSpy (Akhter et al., 2012) and VirSorter2 (Guo et al., 2021). Both tools performed well when evaluated for their ability to predict prophages in bacterial genomes (Roach et al., 2021). After individual prediction for each genome, we use *blastn* to identify sequences predicted by both tools. For plasmid prediction, we used PlasmidFinder (Carattoli and Hasman, 2020), whereas Integronfinder2 (Néron et al., 2022) and MGEfinder (Durrant et al., 2020) were used to identify integrons and other MGE. Defense systems were detected in the genomes of *K. gyiorum* with DefenseFinder (Tesson et al., 2022) using default settings. ABRICATE (https://github.com/tseemann/abricate) was used to screen the assembly sequences against the CARD and RESFINDER databases for ARGs detection and against the Virulence Factor Database (VFDB) for virulence factor identification. We set the percentage of identity and coverage at 80% and 70%, respectively. The ggtree package was used for plotting a heatmap of presence-absence genes.

Analysis of accessory gene content

We used the presence/absence matrix of accessory genes recovered from get_homologues to perform a PCA analysis to assess clustering between isolates from different hosts using the ggfortify package. Then we used BLAST Ring Image Generator

(BRIG) v 0.95 (Alikhan et al., 2011), to compare genomes from different clusters. Variable regions identified by BRIG comparison were extracted and manually checked for accurate annotation using EasyFig version 2.2.273 (Sullivan et al., 2011), in order to determine and visualize the gene content of these regions.

4.3. RESULTS

General features of *K. gyiorum* genomes

We isolated and sequenced 12 strains of *K. gyiorum* from free-living brown-throated sloth rectal swabs which were sampled during a clinical examination at Zoológico of São Paulo between 2014 and 2016. These 12 draft genomes plus five publicly available sequences from the Genbank Database were analyzed (Table 4.1). Overall, the genome size of these 17 strains ranged from 3.77 to 3.99 Mb, with an average length of 3.84 Mb. The genomic size of strains of human origin (3.95-3.99 Mb) was longer than those strains isolated from sloths (3.77 to 3.81 Mb) at least in 0.24 Mb (Figure A4.1A). The GC content ranged from 62.35 to 62.69 %; a small difference is observed between isolates from different hosts (Figure A4.1B).

Table 4.1. Statistics summary of 17 *K. gyiorum* genomes

Strain	Country	Host	%CG	no. of contigs	No. of bases	N50	Completeness	Accession Number
1483E	Brazil	Sloth	62.67	24	3,790,355	304,202	99.76	JALJYH000000000
155D	Brazil	Sloth	62.67	24	3,794,361	286,375	99.76	JALJXS000000000
186F	Brazil	Sloth	62.69	19	3,781,143	421,928	99.53	JALJYN000000000
186J	Brazil	Sloth	62.69	15	3,781,438	426,601	99.53	JAOQMY000000000
262I	Brazil	Sloth	62.67	23	3,807,173	299,337	99.53	JALJYO000000000
2780G	Brazil	Sloth	62.67	22	3,807,569	386,286	99.53	NZ_JALJXY000000000
3324E	Brazil	Sloth	62.69	19	3,781,341	421,905	99.53	JALJYL000000000
3415D	Brazil	Sloth	62.66	23	3,809,345	386,286	99.53	JALJYP000000000
3415G	Brazil	Sloth	62.67	21	3,809,165	809,377	99.53	NZ_JAOQOW000000000.
381J	Brazil	Sloth	62.67	25	3,790,540	304,202	99.76	NZ_JALJXX000000000
4201G	Brazil	Sloth	62.69	20	3,781,254	316,545	99.53	JALJXQ000000000

652G	Brazil	Sloth	62.68	22	3,770,845	466,140	99.53	JALJYI000000000
CCUG_47000	USA	Human	62.35	22	3,986,919	385,940	99.53	NZ_VZPC00000000
CG1	USA	Human	62.43	32	3,943,087	215,440	99.53	NZ_LBNE00000000
KG0001	Ghana	Human	62.45	17	3,983,113	485,120	99.53	NZ_JANKLF000000000
DSM_16618	USA	Human	62.36	9	3,982,678	801,770	99.53	NZ_SGWZ00000000
SWMUKG01	China	Human	62.43	1	3,945,801	3,945,801	99.29	NZ_CP033936.1

Phylogenetic Analysis reveals a structured population of *K. gyiorum*

A maximum likelihood phylogeny was computed on the core-genome alignment of 17 genomes of *K. gyiorum* as well as other members of the *Alcaligenaceae* family: *Alcaligenes faecalis* strain DMS-30030, *Alcaligenes aquatilis* strain QD168 and *Bordetella bronchiseptica* strain NCTC10543 which served as outgroups (Figure 4.1A). The ANI values for all pairs of genomes ranged from 99 to 99.998 % which means that all 17 genomes belong to *K. gyiorum* species and are closely related despite differences in genomic size and GC content (Figure 4.1A). The phylogenetic analysis showed host-associated structure in *K. gyiorum* population with all isolates from sloths grouping in a single clade distant from human isolates (Figure 4.1B).

The 12 *K. gyiorum* isolates sequenced in this study were isolated from 10 brown-throated sloth individuals. SNP distance and population structure between isolates reveal an important genetic diversity of *K. gyiorum* in the sloth population at PEFI. Three clusters with at least 1000 SNPs of difference between these populations were observed (Figure 4.1C). Although the number is not significant, the two individuals (B and J) that had two *K. gyiorum* isolates showed 0 or 2 SNPs difference, which suggests that the intra-host diversity of *K. gyiorum* may be the result of a single clone, unlike other commensal bacteria, such as *Escherichia coli,* which maintains multiple lineages within an individual's gastrointestinal microbiota.

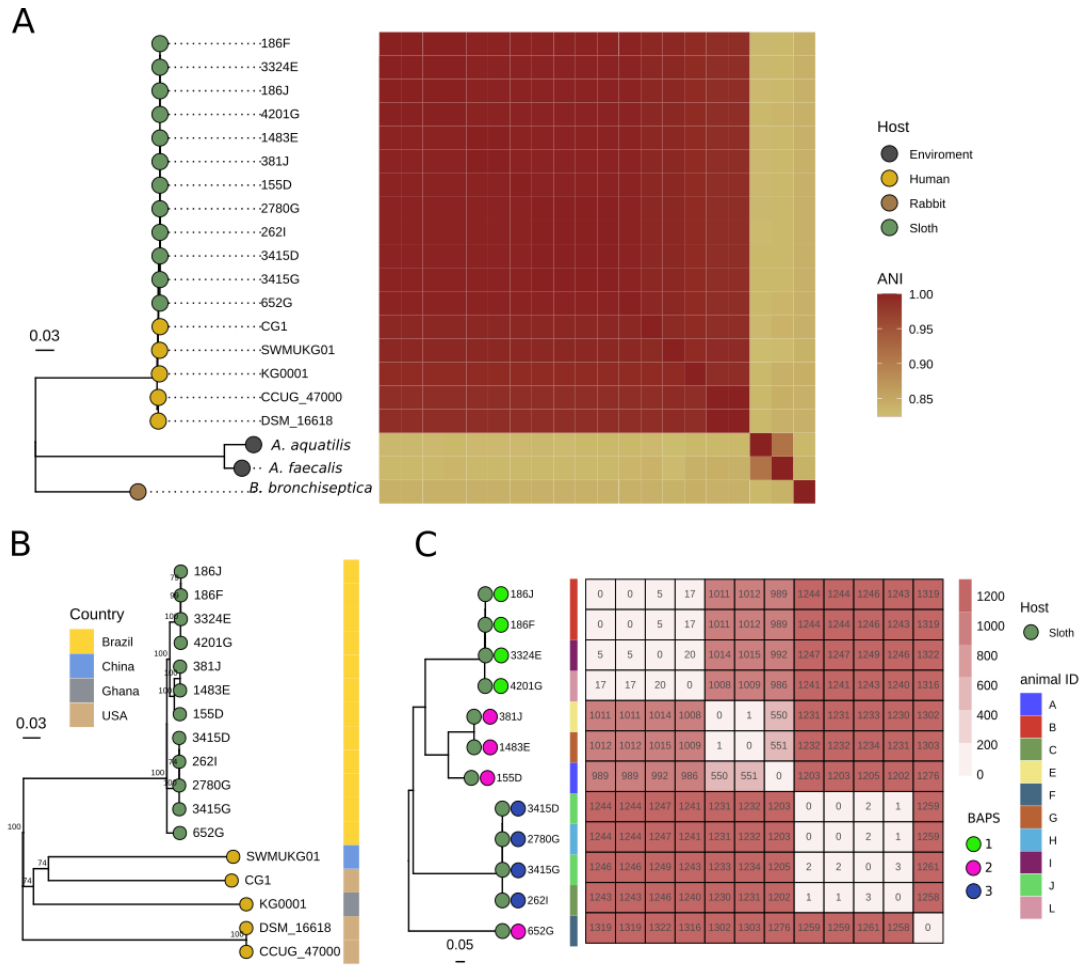

Figure 4.1. Genetic relatedness and phylogenomics of *K. gyiorum* strains. A) Phylogenetic tree based on SNP alignment of core-genome of *K. gyiorum* isolates, *Alcaligenes fecali* DMS-30030 and *Alcaligenes aquatilis* QD168 were used as outgroups coupled to a heatmap of average nucleotide identity (ANI) values. *K. gyiorum* strains share > 95 % ANIm belonging to the same species. B) Phylogenetic tree of 17 *K. gyiorum* with host and country information. C) SNP distance matrix for 12 *K. gyiorum* isolates from brown-throated sloths in PEFI-Brazil.

Pangenome and core genome analysis

All CDSs identified in each isolate of *K. gyiorum* were employed for the pangenome reconstruction of this species. We identified a total of 4,199 different orthologous genes in 17 genomes of *K. gyiorum* representing the pangenome (Figure 4.2A), where 2974 (70%) genes were shared by all strains (core genome), and 1225 (30%) orthologous genes were present in the accessory fraction (cloud and shell genome). The pangenome rarefaction curve plotted

versus the number of genomes shows a trend where more genomes are added to the analysis, the pangenome will increase in size, in other words, an open pangenome (Figure 4.2B). Heap s' law allows us to calculate mathematically if the pangenome is open or closed on the value of α in the equation $n=\kappa N^{-\alpha}$. The $\alpha=0.75$ confirms the openness of *K. gyiorum* pangenome. However, the gene frequency distribution across strains shows a closed pangenome pattern with most genes present in all strains (Domingo-Sananes and McInerney, 2021); this can be explained by the low strain diversity that we used in this analysis (Figure 4.2C).

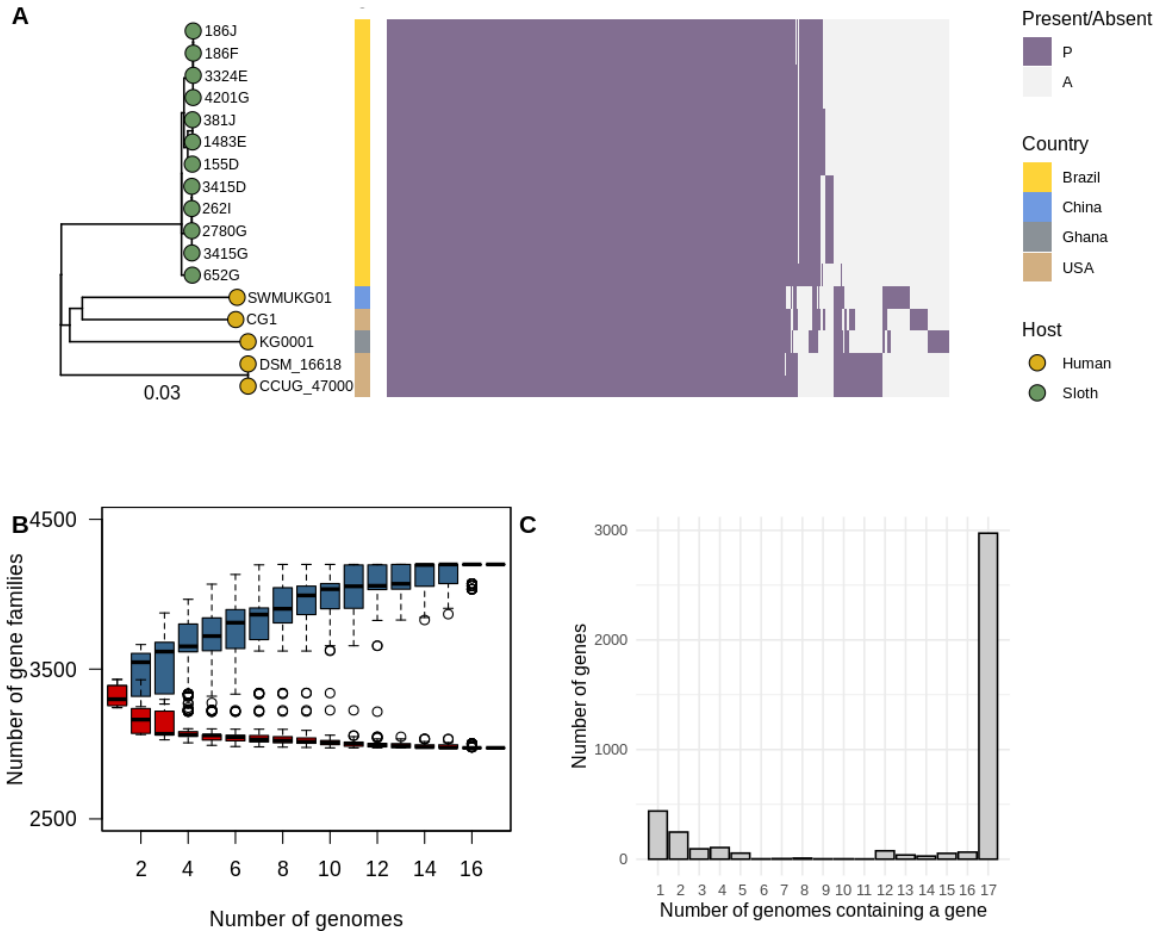

Figure 4.2. Pangenome of *Kerstersia gyiorum*. A) Phylogenetic tree of 17 *K. gyiorum* genomes alongside a matrix of presence/absence of all 4375 genes from the pangenome. B) curves for pan- (blue) and core- (red) genomes of *K. gyiorum* as was calculated using *vegan* package in R, the power-law regression model confirmed the pangenomes was open (α =0.7) C) Gene frequency distribution across strains.

Functional assignation of *Kerstersia gyiorum* pangenome fractions

The attribution of the Cluster of Orthologous Genes (COG) categories to each gene of the pangenome and fractions, core and accessory genomes, are depicted in Figure 4.3 (Detailed description in Table A4.1 and A4.2). First, 3124 (74.4 %) of all 4199 genes from the pangenome were assigned to a single COG category, while 451 (10.7 %) had an ambiguous COG category annotation (more than one COG category), and 622 (14.8%) lacked a COG annotation (unclassified). The most frequent categories in the pangenome were K (Transcription), E (Amino acid transport and metabolism), P (Inorganic ion transport and metabolism), J (Translation, ribosomal structure and biogenesis) and M (Cell wall/membrane/envelope biogenesis), these five categories accounted for 7.66, 7.31, 5.24, 5.07 and 4.81 % of the *K. gyiorum* pangenome, respectively. When we observed the functional enrichment of core and accessory genome (Figure 4.3B), we found that E, K, J, P and I categories represent more than 35.9% of all genes in the core genome, instead, the accessory genes are mainly enriched by categories K, X (Mobilome: prophages, transposons), V (Defense mechanisms) and L (Replication, recombination and repair) representing together 25.7 % of the accessory genome. The relative number of genes present in the accessory was several fold higher than in the core genome for the following informative COG assignments: Mobilome: prophages, transposons (X, 20.8-fold), Intracellular trafficking, secretion, and vesicular transport (U, 2.44-fold), and Defense mechanisms (V, 2.04-fold).

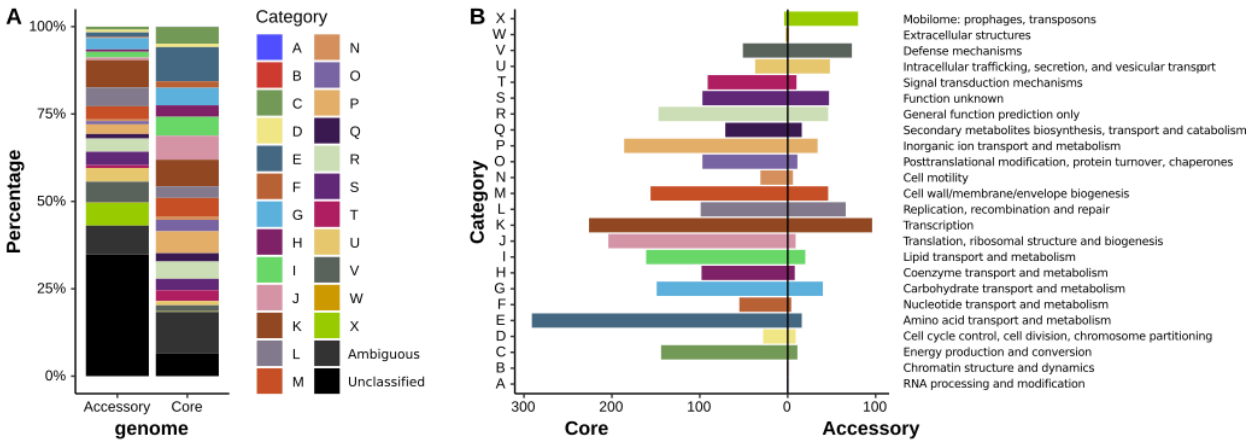

Figure 4.3 Functional annotations of *K. gyiorum* core and accessory genome A) COG categories genes within the core and accessory genomes of *K. gyiorum*. Each category is graphed as a percentage of the total number of genes in the core or accessory genomes. B) Back-to-back barplot comparing the functional annotation of core and accessory genes showing that Mobilome: prophages, transposons (X), Intracellular trafficking, secretion, and vesicular transport (U), and Defense mechanisms (V) categories are more frequent in the accessory compared to the core genome.

The specific cluster of genes identified in sloth and human isolates

The comparison of the accessory portion from the pangenome allows us to recognize the variability of bacterial genome composition. Using principal component analysis (PCA) we observe that gene content variation discriminates *K. gyiorum* genomes according to host origin (sloths and humans) based on the presence/absence of 1225 accessory genes (Figure 4.4A). Two clusters were observed in human isolates. *Scoary* was used to determine which sequences are differentially present in each host. We identified and extracted 76 genes that were exclusively present in sloth isolates and 54 genes that were present only in isolates from human origin. The functional annotation of these genes revealed different functional assignments (Figure 4.4B). While metabolic pathways categories: I (Lipid transport and metabolism), E (Amino acid transport and metabolism), G (Carbohydrate transport and metabolism) and Q (Secondary metabolites biosynthesis, transport and catabolism) are more common in human isolates, the categories related with mobilome: prophages, transposons (X), Intracellular trafficking, secretion, and vesicular transport (U), and Defense mechanisms (V) categories are more frequent in sloth isolates. The full list of genes and functions assigned for each one is shown in Tables A4.3 and A4.4.

The set of genes that are exclusively present in 12 *K. gyiorum* genomes from sloths was arranged in 4 regions (RS01-RS04) with sizes between 8.4 kb and 21 kb (Figure 4.4C). RS01 contains genes poorly characterized, RS02 contains genes involved in metabolic function: “energy production and conversion” (C) and “Carbohydrate transport and metabolism” (G) categories), while the RS03 region contains CRISPR-associated proteins

(cas) that belong to the Defense mechanism (V) category (Figure 4.4D). The RS04 was predicted to be an IncN2 plasmid sequence of 21 Kb (Figure 4.5B).

On the other hand, we found two regions present in all human isolates but absent in sloth isolates: RH01 and RH02 with 18 and 14 genes, respectively (Figure A4.2). For both regions, genes were enriched for categories: “energy production and conversion” (C), “Amino Acid transport and metabolism” (E), “Carbohydrate transport and metabolism” (G) and “lipid transport and metabolism” (I).

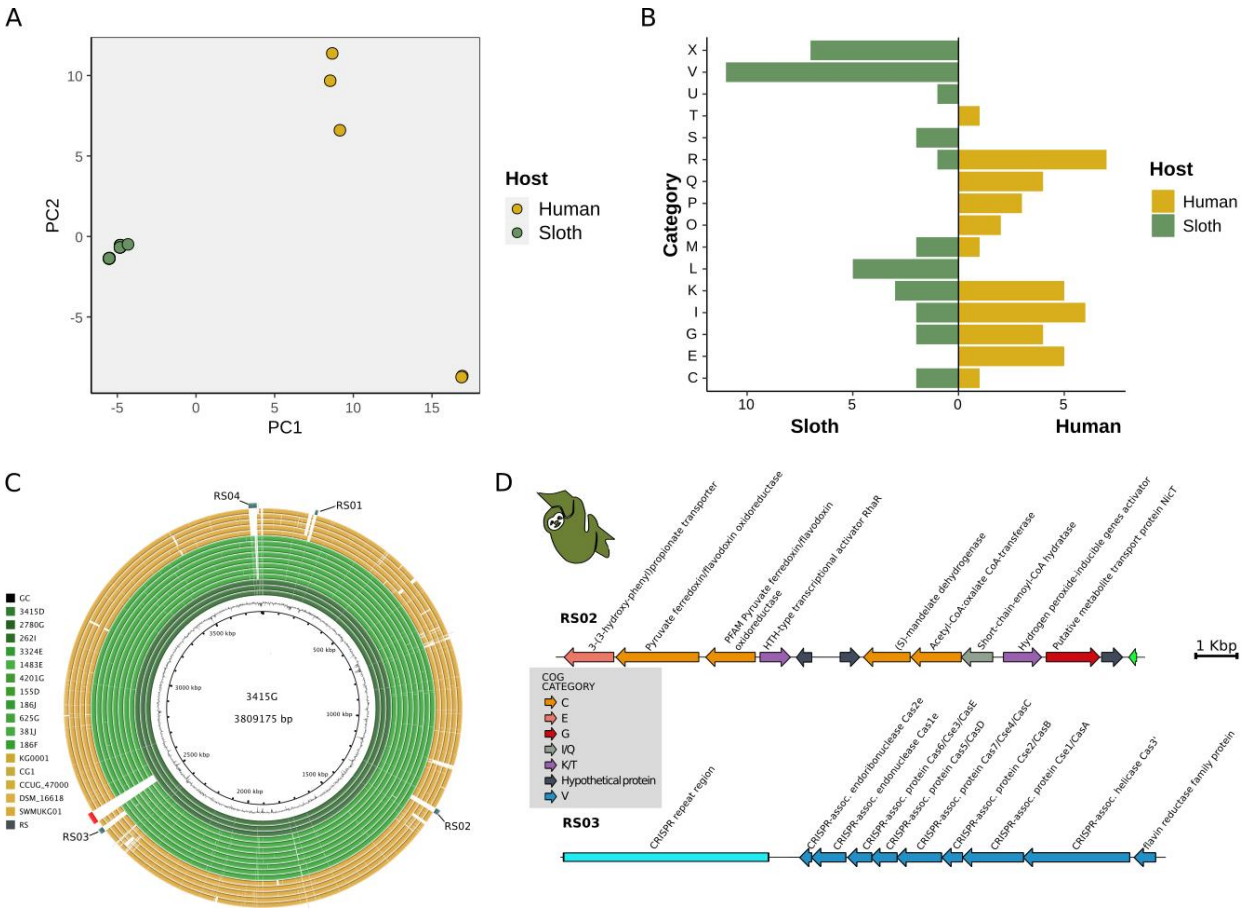

Figure 4.4. Differential gene presence between isolates from different host origins. A) PCA based on accessory gene presence/absence matrix. B) Functional annotation using COG categories to classify specific gene pools reveal different functions in a group of genes of sloths and human isolates C) Blast alignment of all 17 genomes of *K. gyiorum* reveals 4 genetic regions (RS01-RS04) present in all 12 sloth isolates but absent in human isolates D) Gene context of RS02 and RS03 regions of *K. gyiorum* genomes from sloths colored according to COG category

Mobile genetic element (MGE) and Defense Systems detection and characterization in *K. gyiorum* genomes

We used plasmidfinder for plasmid replicon identification. Inc1Q, Inc2Q and IncN2 replicons were detected with at least 70/70% of identity/coverage sequence limit; IncN2 and Inc2Q were present in sloth isolates but absent in human isolates (Figure 4.5A). The IncQ1 replicon was detected only in SWMUKG01 strain isolated from humans, but no plasmids were reported in a previous study (Li et al., 2019b). The contig containing the IncN2 replicon was ~21Kb long with 29 coding regions (Figure 4.5B). Interestingly, a BLAST search of this sequence against the "nr" database returned a match with 70% coverage and 95% identity with the pXap41 plasmid found in the bacterium *Xanthomonas arboricola* pv. pruni (Pothier et al., 2011). This plasmid carries genes encoding type III effectors and helper genes that are absent in the probable plasmid of *K. gyiorum.* In this plasmid, we identified genes involved in pili formation, plasmid conjugation (*traF, G, H, L* and J), and segregation (*par A, B*).

Phispy and virsorter2 were used to predict phage sequences in all genomes. In total, 26 phage sequences were predicted by both tools (Table A4.5). From these, 10 were predicted by virsorter2 and 16 by Phispy. After performing a BLASTn search between predictions by each software, we consider only sequence regions that were predicted by both tools. Finally, five different phage sequences were obtained (Table A4.6). We gave these provisional names: *Kerstersia gyiorum* phage Kgϕ1 through Kgϕ5 (Figure 4.5 A, C). Kgϕ1 (27.7 Kb) was the only detected in sloth isolates (just in four isolates). Each of the five human isolates has a different phage sequence (Kgϕ2-Kgϕ5), except for Kgϕ2 (15.8 Kb) which was present in two human isolates (DSM_16680 and CCUG_47000 strains). All phages were predicted to be dsDNA phages. To predict other MGE, we used Integronfinder and MGEfinder to seek integrons and transposable elements. We did not detect any integrase, transposase or insertion sequence (IS) in *K. gyiorum* genomes.

Microbes can defend themselves against phages and other MGEs using a variety of systems. We used DefenseFinder tool to identify all anti-viral systems in *K. gyiorum* genomes. We detected 14 different defense systems including the most common systems in prokaryotes: *cas* and Restriction-Modification (RM) (Figure 4.5A and Table A4.7). In all sloth isolates two CRISP-cas type I operons were predicted, one of them including CRISPR spacers (Figure 4.4C). Additionally, other systems were present exclusively in sloth isolates: the Defense Retro-Transcriptase (DRT) system (*drt3ab* and *drt4*) and Dynamins (LeoA and LeoBC) both involved in anti-viral defense functions.

More diverse defense mechanisms were found in human isolates including RM type I, II and IV; BREX, CBASS, Rst, and Gabiija.

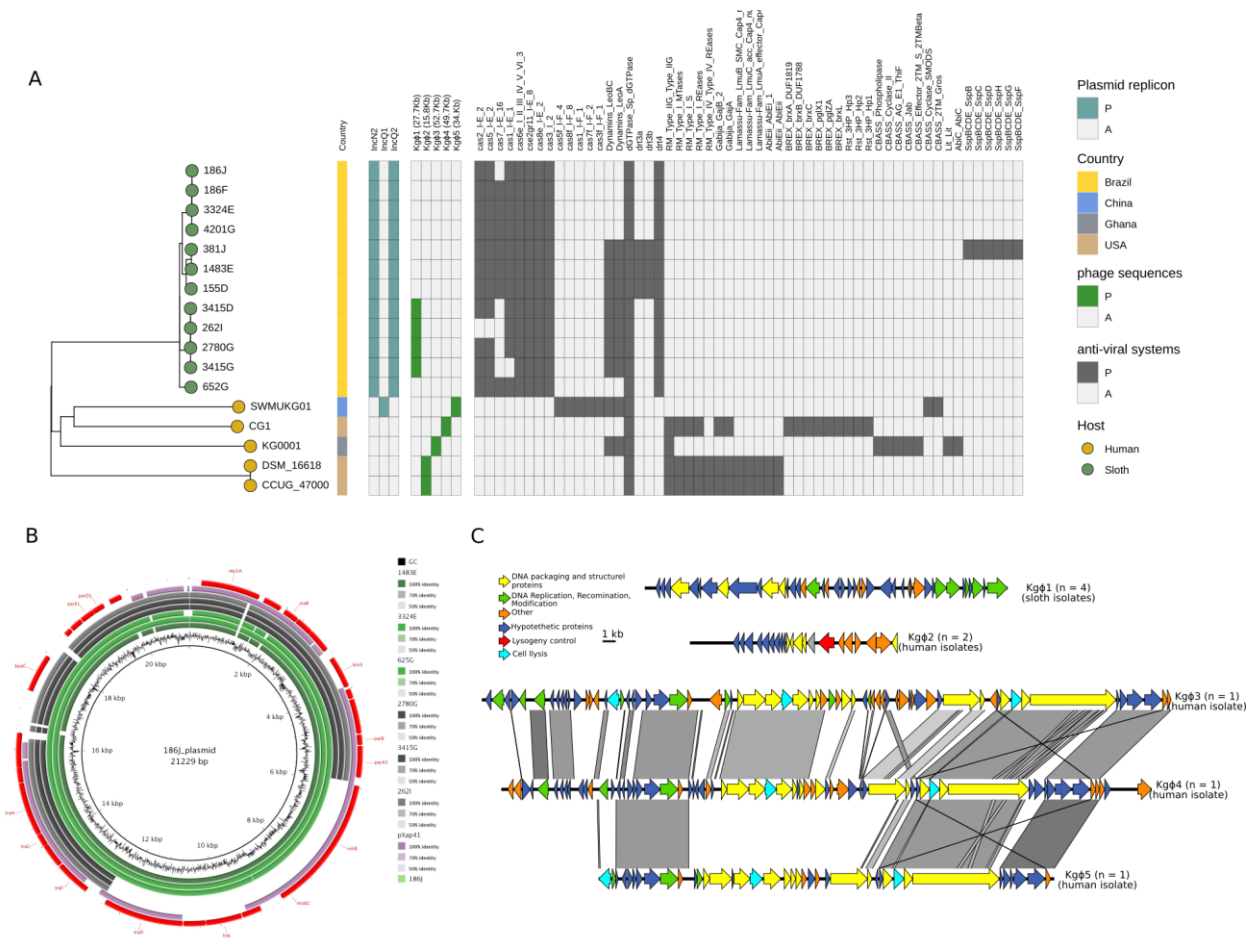

Figure 4.5 Mobile genetic element (MGE) and Defense Systems of *K. gyiorum*. A) presense absense matrix of plasmid replicons, phage sequences, and defense systems predicted for 17 *K. gyiorum* genomes B) Comparison of the partial IncN2 plasmid (~ 21 kb) sequence of *K. gyiorum* from sloths with pXap41 plasmid from plant pathogen *Xanthomonas arboricola* pv. pruni, shows partial alignment (70% of coverage) between these sequences. C) Map of five predicted phages in *K. gyiorum* genomes.

Virulence Factor and Antimicrobial Resistance Genes Prediction

Our analysis shows the detection of 51 virulence factors in *K. gyiorum* genomes. These genes are associated with adhesion, biofilm formation, capsule, flagella, LPS, iron uptake and secretory system (Figure 4.6). Most of these sequences are conserved in isolates from both hosts. The flagellar regulon of *K. gyiorum* was described by Li et al., 2019 in the SWMUKG01 genome, here we identified operons, *cheARWY*, *flgCDEFGHIK*, *flhABC*, *fliFGIMPORS*, and *motAB* conserved in all isolates with exception of *fliR* that was lacked by four isolates from sloths. Cell surface components like capsule and lipopolysaccharides (LPSs) are essential virulence factors in some Gram-negative bacteria. We identified a highly conserved operon *tviBC* that is involved in the biosynthesis of LPS and capsule in pathogenic bacteria like *Acinetobacter baumannii* or *Escherichia coli*. Other genes involved in capsule formation *kpsU*, *rfbA* and *wcaJ*, the last only present in human isolates. On the other hand, five genes associated with iron uptake were identified in *K. gyiorum*, the *sitABC* operon is involved in Fe uptake in *Staphylococcus*, and was conserved in *K. gyiorum*. The *bcr* gene was identified only in three isolates of human origin and codifies a permease that maintains appropriate levels of cellular alcaligin levels in *Bordetella* species (which is also part of the *Alcaligenaceae* family).

Antimicrobial resistance in clinical isolates of *K. gyiorum* was reported previously for ciprofloxacin and cephalosporins (Pence et al., 2013). Here we detected 20 chromosomal genes mainly encoding multidrug efflux pump systems against fluoroquinolones, tetracycline, phenicols, cephalosporins, as well as other antimicrobials. CeoAB-OpcM efflux pump with resistance to aminoglycoside and fluoroquinolone; MexGHI-OpmD with resistance to

fluoroquinolone and tetracycline, MuxABC-OpmB efflux pump system with resistance to aminocoumarin, macrolide, monobactam and tetracycline; and the complex *smeABC* of resistance to aminoglycoside and cephalosporin.

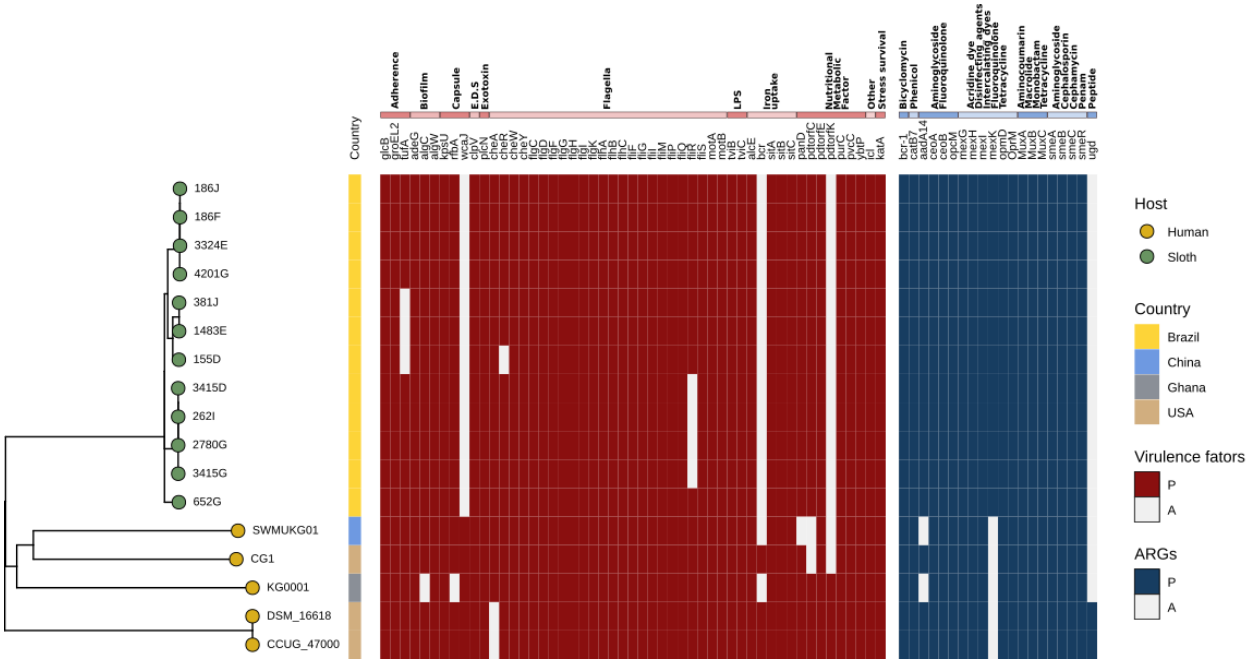

Figure 4.6 Resistome and virulome of *K. gyiorum*. Phylogenetic tree of 17 *K. gyiorum* genomes alongside a heatmap of presence/absence of virulence factor (red) and ARGs (blue) predicted by abricate tool using VFDB and CARD databases, respectively.

4.4. DISCUSSION

Kerstersia gyiorum has been isolated from multiple environments including wastewater, aquatic mammals, and insects and is associated with infections in humans. However, until now only *K. gyiorum* genomes of clinical origin have been described. Here we add new genomic data of *K. gyiorum* isolated from healthy free-living brown-throated sloths (*Bradypus variegatus*) and perform a comprehensive comparative analysis to gain insights into the genetic diversity, population structure, and evolution of this bacterium species.

Our analyses show substantial differences between the *K. gyiorum* genomes isolated from sloths and humans at various levels. The phylogenomic analysis shows that isolates from sloths cluster together, whereas human isolates segregate distant from sloth clade suggesting

that *K. gyiorum* populations are structured according to host. Additionally, we observed differences in genome size (in bp) and GC content between isolates from both hosts. This pattern may be evidence of host-adapted lineages of *K. gyiorum* but other explanations are also valid, for example, physical barriers in different host niches followed by genetic drift (Sheppard et al., 2018). Sloths have a peculiar ecology. Due to their lifestyle, they have generated physiological adaptations that make their digestive system a unique environment. Some works have described that the gut microbial population in sloths shows a distinctive diversity compared to other herbivores (Delsuc et al., 2014; Dill-McFarland et al., 2016). This niche condition significantly different from the human microbiome would favoured the fixation of specific adaptive signatures that lead to a different ecotype (ecotype model). Other isolates from different locations and hosts are necessary to assess these hypotheses on *K. gyiorum* population.

On the other hand, the *K. gyiorum* genetic diversity in the sloth population of PEFI was revealed by SNP distance and genetic content. The 12 *K. gyiorum* isolates from 10 sloth individuals were distributed in 3 clades, which differ between them by more than 1000 SNPs corresponding to approximately 99.98% ANI. In other bacteria species, this genetic distance is indicative of separated lineages or strains, for example, in *Salmonella enterica* ser. Typhimurium approximately 400-600 SNPs (99.99% ANI) is used to define different strains (Branchu et al., 2018). Genetic content differences between isolates intra-sloths were mainly enriched by genes related to phage and plasmid sequences (Figure A4.3), this suggests that horizontal gene transfer (HGT) is important to generate genetic diversity in *Kerstersia gyiorum.*

HGT plays an important role in strain diversification in bacteria that are rapidly facing new ecological challenges and opportunities. (Rocha and Bikard, 2022). We identified a partial IncI2 plasmid sequence in all sloth isolates, with partial identity/coverage to a virulent plasmid of *Xanthomonas arboricola* pv. pruni (Pothier et al., 2011). We do not identify virulence or antimicrobial-resistant genes in this plasmid; however, due to short-read sequencing, we can not recover the full plasmid sequence. Additionally, a unique phage sequence of 27.7 kb of

size was predicted in four isolates from sloths, and another four phage sequences in human isolates, one in each genome. No transposable elements or plasmids were predicted in human isolates. In general, our analysis revealed few MGE in *K. gyiorum* genomes. Instead, the presence of MGE in the *alcaligeneaceae* family is abundant as was described recently (Ellabaan et al., 2021). The most representative genus of *alcaligeneaceae* family, *Achromobacter* spp. and *Bordetella bronchiseptica* contain large quantities of MGE in their genome in contrast with *Bordetella pertussis* or *K. gyiorum* as we now describe. The small number of MGEs identified might be related to the presence of multiple defense mechanisms in *K. gyiorum* genomes. We detected different anti-MGE defense systems in genomes from distinct hosts. CRISPR-Cas and RM systems are the most widespread systems in prokaryotes (Tesson et al., 2022), in sloth isolates the most prevalent system was CRISPR-cas (Type I E), whereas RM (Type I, II and IV) in human isolates.

We identify 51 putative VF genes associated with adhesion, biofilm formation, capsule, flagella, LPS, iron uptake and secretory system, almost all of them are present in isolated from both hosts, but *wcaJ* (capsule), *bcr* (iron uptake) and *pdtorfk* (metabolism) were only detected in human isolates from clinical samples. These factors would play a role in pathogenesis, *wcaJ* encodes a glucosyltransferase and experimental deletion of this gene was associated with less virulence and phage susceptibility in *Klebsiella pneumoniae* (Cai et al., 2019), while *bcr* is involved in iron uptake regulation in pathogenic *Bordetella* species (Brickman and Armstrong, 2005). *K. gyiorum* was reported to have a motility variable (Coenye et al., 2003; Almuzara et al., 2012), however, the flagella regulon was conserved in all *K. gyorum* genomes, whereas *K. similis,* other species of the genus, lacks flagella (Vandamme et al., 2012). In *B. bronchiseptica* motility and/or flagella play an important role during infection and recently the same was demonstrated for *B. pertussis* (Hoffman et al., 2019).

Here, we identified a group of genes encoding efflux pump systems that potentially may be involved in aminoglycoside, tetracycline and phenicol resistance. All of these genes were chromosomally located, not associated with MGE and conserved in all *K. gyiorum*

genomes. However, disk diffusion susceptibility testing revealed susceptibility to almost all antibiotics in sloth isolates, resistance to norfloxacin (quinolone) was observed in two isolates (Table A4.7), suggesting that genes predicted may not be involved in phenotypic resistance. In human *K. gyiorum* isolates, resistance to ciprofloxacin (quinolone) was reported in different countries including Brazil (Pence et al., 2013; Li et al., 2019b; Pires et al., 2020). The main mechanism of fluoroquinolone resistance in Gram-negative bacteria is the accumulation of mutations in DNA gyrase and DNA topoisomerase IV (Bush et al., 2020). There is no information on *gyrAB* mutations that may confer resistance against fluoroquinolones in *K. gyiorum* and we don't find polymorphism in the alignment of these genes in our dataset. The low levels of resistance to fluoroquinolones reported in *K. gyiorum* may be due to efflux proteins as was observed in Enterobacteriaceae (Poole, 2000). A standardized CLSI criterion for antimicrobial susceptibility testing and interpretation of results for *K. gyiorum* is needed for a correct understanding and determination of resistance in this bacteria.

The present study also constitutes a comprehensive comparative genomic characterization of *K. gyiorum*. We constructed the pangenome for the species and according to Heaps' law calculation (α = 0.77) it was defined as an 'open' pangenome. The open or closed nature of a pangenome is bound to the lifestyle of the bacterial species (Tettelin et al., 2008). An open pan-genome indicates that species have a high capacity to exchange genetic material and also indicates a free-living lifestyle with metabolic versatility (Rouli et al., 2015; McInerney et al., 2017). This definition is appropriate to the *K. gyiorum* lifestyle, which is found in multiple animal hosts including the environment. However, our comparative analysis limited by the availability of genomes from only two mammal hosts (human and sloth) is not representative of the full diversity of this species. The number and diversity of genomes used in the analysis impact on the accurate measurements of pangenome openness (McInerney et al., 2020). Therefore, it is desirable to increase the number of *K. gyiorum* sequences from diverse hosts and geographic settings to gain a better understanding of their genomic diversity and pangenome.

K. gyiorum has an average genome size of 3.84 Mb, the core genome (2976 OG) represents 90% of an average strain genome (3280 CDS) and around 70% of the *K. gyiorum* pangenome (4199 OG). The large core genome size in *K. gyiorum*, despite the open pangenome, may be due to the small number of genomes in the analysis and will decrease when more genomes are added to the analysis. Differences in functional annotation were noted for pangenome portions, whereas the core genome was enriched for metabolic categories (E, K, J, P and G) the accessory genome was enriched for information storage and processing (K and L), mobile genetic elements and prophages (X) and defense mechanism (V) categories. Typically, core genomes are enriched for housekeeping functions, whereas the accessory genome is often associated with selfish elements (plasmids, phage, and transposons), genes involved in protein trafficking and defense, and niche-specific functions (McInerney et al., 2017; Innamorati et al., 2020). Variable gene content in isolates from different hosts was characterized. Whereas specific genes in 12 sloth isolates were represented by plasmid, prophage sequences (X) and anti-viral system (V), metabolic categories (I, E, G, Q) were differentially present in 5 human isolates.

In conclusion, here we present 12 new genomic sequences of *K. gyiorum* isolated from free-living brown-throated sloths in Brazil and perform a comprehensive comparative genomic analysis for the species. Our study reveals a structured population of *K. gyiorum* according to hosts suggesting that these lineages are host-specialized. An open pangenome was determined for the species in concordance with *K. gyiorum* versatility despite few MGEs in genomes analyzed because of the high diversity of defense mechanism systems. Differences in gene content in isolates with different functional assignations would suggest differences in metabolic capacities in lineages from different hosts. This study represents the first comparative analysis of *K. gyiorum* and provides new insight into mechanisms that drive the diversification of the species in different hosts.

4.5. CONTRIBUTIONS

This work was carried out in collaboration with the Laboratory of Microbiology and Molecular Biology of the Fundação Parque Zoológico de São Paulo. Led by the researchers: Iris Lima Gonzalez and Patricia Locosque Ramos who performed the sampling, isolation and microbiological characterization of Kerstersia gyiorum strains.

This work also involved the collaboration and supervision of Prof. Aline Maria Da Silva from the IQ-USP.

Genome sequencing was done by the Joint Genome Institute, thanks to a recommendation by Professor Barny Whitman, from the University of Georgia, USA.

In this chapter, my primary contributions include: (i) bioinformatics data analyses, (ii) results interpretation and (iii) draft the manuscript.

Our intention is to prepare a manuscript based on the results of this chapter and submit it for publication in the months to follow.

5. Genomic characterization of a novel *Neisseria* species (*Neisseria bradyp*) isolated from the oral cavity of brown-throated sloths (*Bradypus variegatus)* in Brazil

ABSTRACT

Neisseria is a highly diverse genus that commonly inhabits animal mucouse. Two *Neisseria* strains were isolated from the oral cavity of brown-throated sloths (*Bradypus variegatus*) in Brazil and characterized using comparative genomics tools. After phylogenomics and ANI analysis of 420 genomes representing 33 different species, the two isolates from sloths were identified as belonging to a novel species that we named *Neisseria bradyp.* The GC content of these isolates was around 52.5 %, and the genome size was around 2.4 Mbp. Pangenomic analysis of the genus revealed the presence of 480 exclusive genes in *N. bradyp* mainly enriched in functions related to carbohydrate metabolism and cell wall, membrane and envelope biogenesis. *N. bradyp* represents the first *Neisseria* species isolated from common sloths.

Keywords. *Neisseria bradyp*, *Neisseria*, brown-throated sloth, comparative genomics

5.1. INTRODUCTION

The genus *Neisseria* comprises a diverse group of species that colonize the mucosal surface of animals and humans (Liu et al., 2015). This genus includes the globally significant pathogens *N. meningitidis* and *N. gonorrhoeae* which cause invasive diseases in humans (Bennett et al., 2014). Non-Pathogenic *Neisseria* (NPN) are harmless commensal members of the microbiota; although some have been identified as opportunistic pathogens including some *Neisseria* species that have the ability to cross the species barrier and cause infections in humans, highlighting the potential zoonotic significance of these bacteria (Heydecke et al., 2013; Merlino et al., 2021). While *N. meningitidis* and *N. gonorrhoeae* have been extensively studied, far less attention has been paid to other *Neisseria* species (Liu et al., 2015).

Neisseria has been isolated from the oral cavity, respiratory and digestive tracts, and stool of many mammals, reptiles and avian hosts and even was found as a free-living bacteria in rivers and soils (Liu et al., 2015). Variation at phenotypic, genetic and morphological levels with some members showing rod and coccoid morphology confirms the remarkable diversity of this bacterium (Bennett et al., 2014). Population genomics and phylogenetics analysis have helped to study the genetic diversity of *Neisseria* and helped define species reclassification (Bennett et al., 2012; Maiden and Harrison, 2016). A genomic inspection into capsule structure in pathogenic and non-pathogenic *Neisseria* has revealed various capsule structures in different species with a probable exchange of capsule locus and other virulence factors between species by horizontal gene transfer (Marri et al., 2010; Clemence et al., 2018).

The Bradypus Research Program led by the São Paulo Zoological Park Foundation is concerned about the bacterial diversity in the oral and rectal cavity of the brown-throated sloth (*Bradypus variegatus)* in The Parque Estadual das Fontes do Ipiranga (PEFI), an important remaining fragment of Atlantic Forest surrounded by an urban region in the city of São Paulo, Brazil. The brown-throated sloth is an arboreal mammal from the Xenarthra superorder distributed throughout Central and South America inhabiting the Amazon, Caatinga and Atlantic Forest biomes (Hayssen, 2010). *B. variegatus* is classified as “least concern” by the International Union for Conservation of Nature (IUCN). As part of this program, we isolate two *Neisseria* spp. from the oral cavity of two individual free-living sloths. No previous records on the isolation of *Neisseria* in Xenartra order were reported until now; however, a study using 16S RNA sequencing of the gut microbiome of *B. variegatus* revealed that *Neisseria* sp. is the most abundant bacterial phyla in the foregut. The 16S RNA sequence was not similar to any previously reported *Neisseria* species suggesting that a novel species would habit these hosts (Dill-McFarland et al., 2016).

In this study, we report the isolation and genomic sequencing of a novel *Neisseria* species, which was isolated from the oral cavity of brown-throated sloths and characterized

following a comparative genomic approach. Based on the findings, we conclude that these isolates represent a novel species of the genus *Neisseria*.

5.2. MATERIAL AND METHODS

Ethics statement

Animal capture and sample collection were conducted by a previous study with permission of Instituto Chico Mendes da Conservação da Biodiversidade (ICMBio) and Sistema de Autorização e Informação em Biodiverdidade (SISBIO) license number 49627 and it was approved by the scientific board of São Paulo Zoological Park Foundation.

Whole genome sequencing

DNA was extracted from pure culture using a commercial kit. DNA was sent to the Joint Genome Institute (JGI) for sequencing using Illumina technology. The two *Neisseria bradyp* genomes from sloths were annotated with the Prokaryotic Genome Annotation Pipeline (PGAP). The quality of assemblies was assessed with checkM software (Parks et al., 2015). Sequencing statistics of genome sequences are shown in Table 5.1.

Genomic analysis

Generation of bacterial genome data set

As of March 2023, the pubMLST database contained 47,651 *Neisseria* spp. genomes, of which 97.8% (46,603 genomes) belonged to relevant human pathogenic *Neisseria* species (*N. meningitidis and N. gonorrhoeae*), 1.4% (685 genomes) to *N. lactamica* and only 0.8 % (363 genomes) to other *Neisseria* species isolated from humans and other animals. In order to perform a comparative analysis and have the highest genomic diversity, we downloaded all commensal genomes as well as 57 genomes of *N. meningitidis, N. gonorrhoeae* and *N. lactamica*. Finally, our database was composed of 422 genomes including the two *N. bradyp* isolated from sloths.

Phylogenetic analysis and Average Nucleotide Identity (ANI)

A ribosomal multilocus sequence typing (rMLST) tree was constructed based on 53 rMLST loci of all 422 genomes extracted from pubMLST (Jolley et al., 2018). A maximum likelihood method was used with the general time reversible (GTR) model and 1000 bootstrap replicates running the IQ-TREE2 (Minh et al., 2020) program and visualized with *ggtree* package in R (Yu et al., 2017).

All 422 genome assemblies were used as input files for Prokka annotation program with its default settings (Seemann, 2014). The protein sequence outputs produced by Prokka were further used as inputs for panaroo (Tonkin-Hill et al., 2020), in order to identify groups of orthologous genes. Then the core genes were aligned with MAFFT aligner and the concatenated core alignment was used as input for tree reconstruction using IQ-TREE2. The *Neisseria* tree was rooted using *Moraxella osloensis* and *Moraxella catarrhalis* genomes as outgroups.

FastANI (Jain et al., 2018) was used to compute the pairwise average nucleotide identity (ANI) among all 422 bacterial genomes, and the *gheatmap* function from the R package *ggtree* (Yu et al., 2017) was utilized to create a heatmap representation of the results.

Pangenomic analysis and Functional annotation

Orthologous genes were grouped with panaroo (Tonkin-Hill et al., 2020) using sequence identity of > 70% and query coverage of > 70 %. The pangenome was represented on a heatmap using the presence/absence matrix of orthologous genes.

We used the Clusters of Orthologous Group (COG) database (Galperin et al., 2021) for functional annotation of all proteins predicted in *Neisseria* genomes. If a gene was assigned to more than one COG category for functional annotation analysis, it was defined as ‘ambiguous’ category. Graphical representation of results was elaborated with ggplot2.

5.3. RESULTS

Genomic sequencing of new putative species *Neisseria bradyp* isolated from brown-throated sloth

Two strains (327A and 3324A), named *Neisseria bradyp* as a result of this study, were independently isolated from the oral cavity of two healthy brown-throated sloths (*Bradypus variegatus*) in the PEFI- São Paulo, Brazil. The whole genome sequencing was performed in the Illumina platform obtaining draft assemblies of 55 and 82 contigs. The CG content of these isolates was around 52.5 %, and the genome size was around 2.4 Mbp (Table 5.1).

The quality of assemblies was assessed with checkM resulting in a completeness value of 100% for both genomes and contamination of less than 0.5%.

Table 5.1: Sequencing statistics of two *N. bradyp* isolates

Isolate	Specie	% GC	N° of contigs	N50	L50	Size (pb)	N° CDS	N° rRNA	N° tRNA	Completeness	contamination
327A	*Neisseria bradyp*	52.59	55	11	74,752	2,405,174	2253	3	59	100	0.13
3324A	*Neisseria bradyp*	52.53	82	14	61,072	2,412,187	2277	3	59	100	0.04

Phylogenomic analysis of *Neisseria* genus identifies distinct species groups

Phylogenetic analysis based on core genome alignment of 422 genomes (representing 33 different *Neisseria* species) generated discrete species groups with the new putative species *N. bradyp* isolated from sloths more related to *N. animalis* (guinea pig), and *N. chenwenguii* (plateau pika), while the most common human-associated *Neisseria* species (*N. gonorrhoeae*, *N. meningitidis*, *N. lactamica* and *N. polysaccharea*) appear distantly related to *N. bradyp* (Figure 5.1). The grouping observed in human-associated *Neisseria* species was consistent with phylogeny based on rMLST scheme (Figure A5.1). Another seven human-associated species isolated from African patients: *N. bergeri*, *N. blantyrii, N. uirgultaei, N.*

benedictiae, N. bassei, N. maigaei and *N. viridae* were also closely related to *N. meningitidis* (Figure 5.1). Within the core genome phylogeny, *N. polysaccharea* and *N. bergeri* formed a polyphyly with isolates descending from various ancestral groups suggesting the existence of different subspecies, this was also observed by Maiden and Harrison using rMLST scheme.

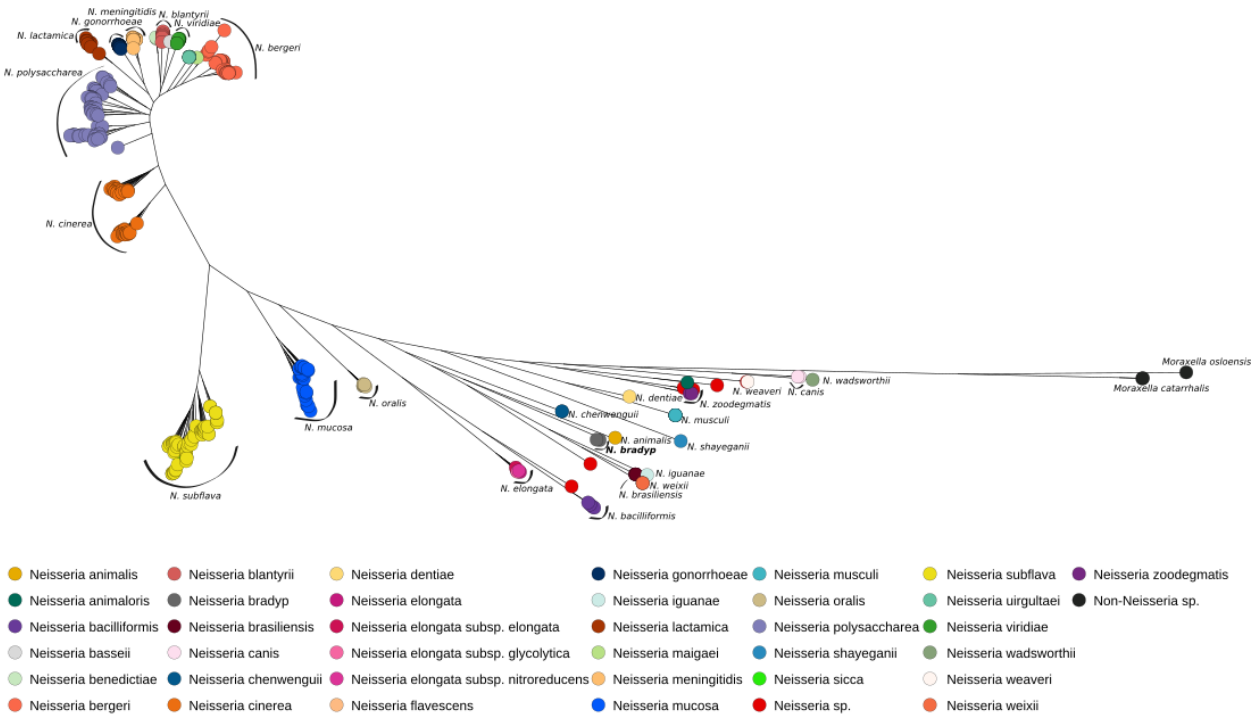

Figure 5.1. Phylogeny of the *Neisseria* genus. Maximum likelihood tree of *Neisseria* genus generated from core genome alignment from 422 publicly available genomes including two from novel *N. bradyp*. Tippoints are colored according to the labeled species provided on the strain sample in the PubMLST database. We used two outgroup species: *Moraxella osloensis* and *Moraxella catarrhalis.*

Average nucleotide identity (ANI) supports the assignation of *Neisseria bradyp* as a new *Neisseria* species

Calculation of ANI has been used as the gold standard in bacterial taxonomy and a threshold of 95–96% is recommended as the criterion for species assortment (Richter and Rosselló-Móra, 2009). We performed an ANI analysis between all genomes of our dataset using fastANI (Figure 5.2 and Table A5.1). The ANI value between our strains 327A and 3324A was 99.3 indicating they belong to the same species (*N. bradyp*). The ANI value

between the type strain 327A and other *Neisseria* species was 78.3–81.0%, which is lower than the threshold of 95% ANI at which isolates should be considered distinct species (Figure A5.2). The ANI scores between *N. bradyp* and closely related species *N. chenwenguii* and *N. animalis* are 80.7 and 81.0%, respectively, while ANI between *N. brady* and the group of *N. meningitidis* and *N. gonorrhoeae* is around 79.8%. ANI analysis shows that the genome sequence of *N. bradyp* is distinct from other close relatives in the genus.

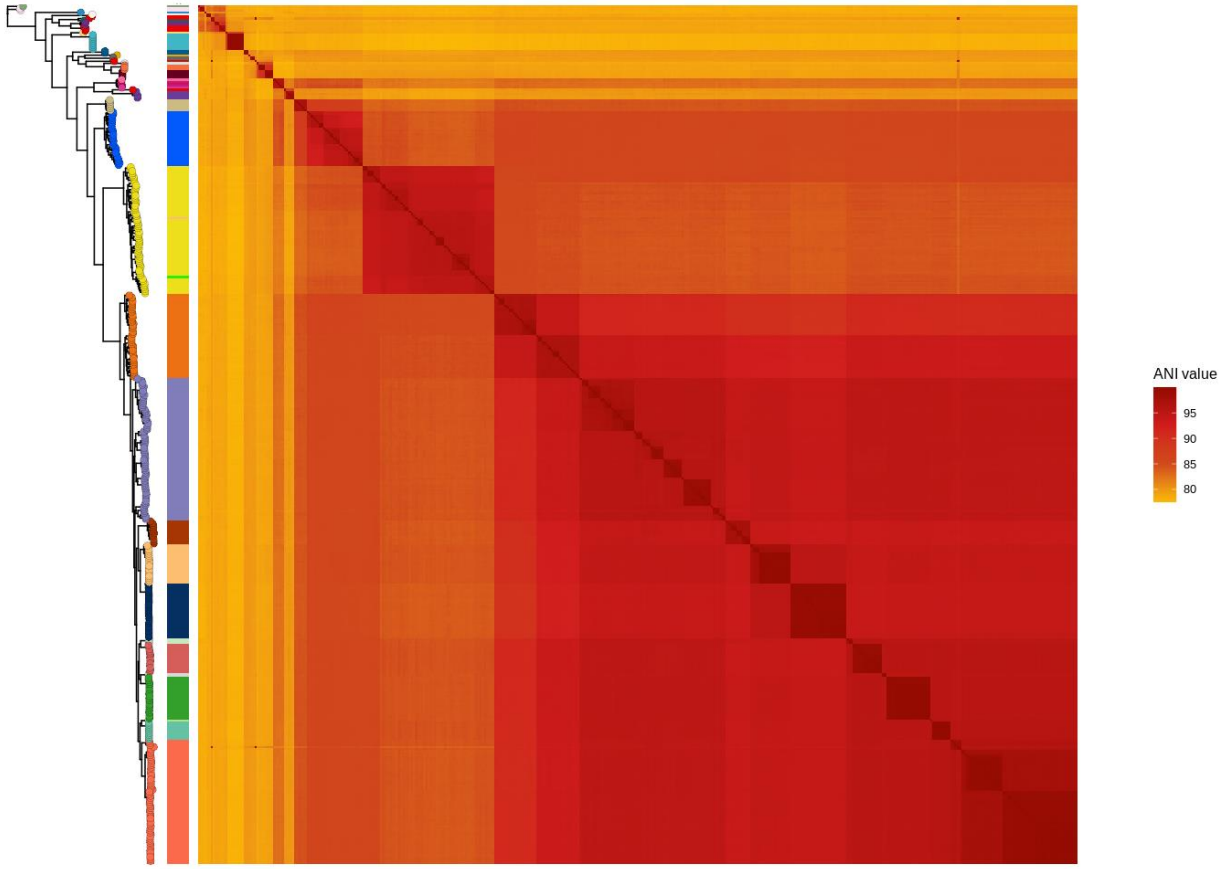

Figure 5.2. Average Nucleotide Analysis between *Neisseria* species. Maximum likelihood tree of *Neisseria* genus generated from the core genome alignment of 422 publicly available genomes including two from novel *N. bradyp* and a heatmap of pairwise average nucleotide identity (ANI) values between each pair of genomes. The species of each genome is indicated by the vertical color strip to the right of the tree and by the tip point. ANI values were calculated using FastANI and are represented as a color gradient as indicated in the ANI (%) key.

Pangenome of *Neisseria* genus and identification of exclusive genes in *N. bradyp* species

We calculated the genus pangenome using the panaroo pipeline. The pan-genome comprised 25,489 discrete genus groups (Figure 5.3), of which 253 were present in at least 99 percent of all genomes. The soft-core (present in 95 and 99 % of all genomes) was composed of 293 groups, whereas the cloud genes (present between 0 and 15 % of all genomes) were 22,676 groups. The heatmap in Figure 5.3 shows some clusters of genes that are differentially present in discrete species groups distributed in the phylogenetic tree. We identified 480 genes that were unique in both *N. bradyp* isolates and absent in other *Neisseria* members.

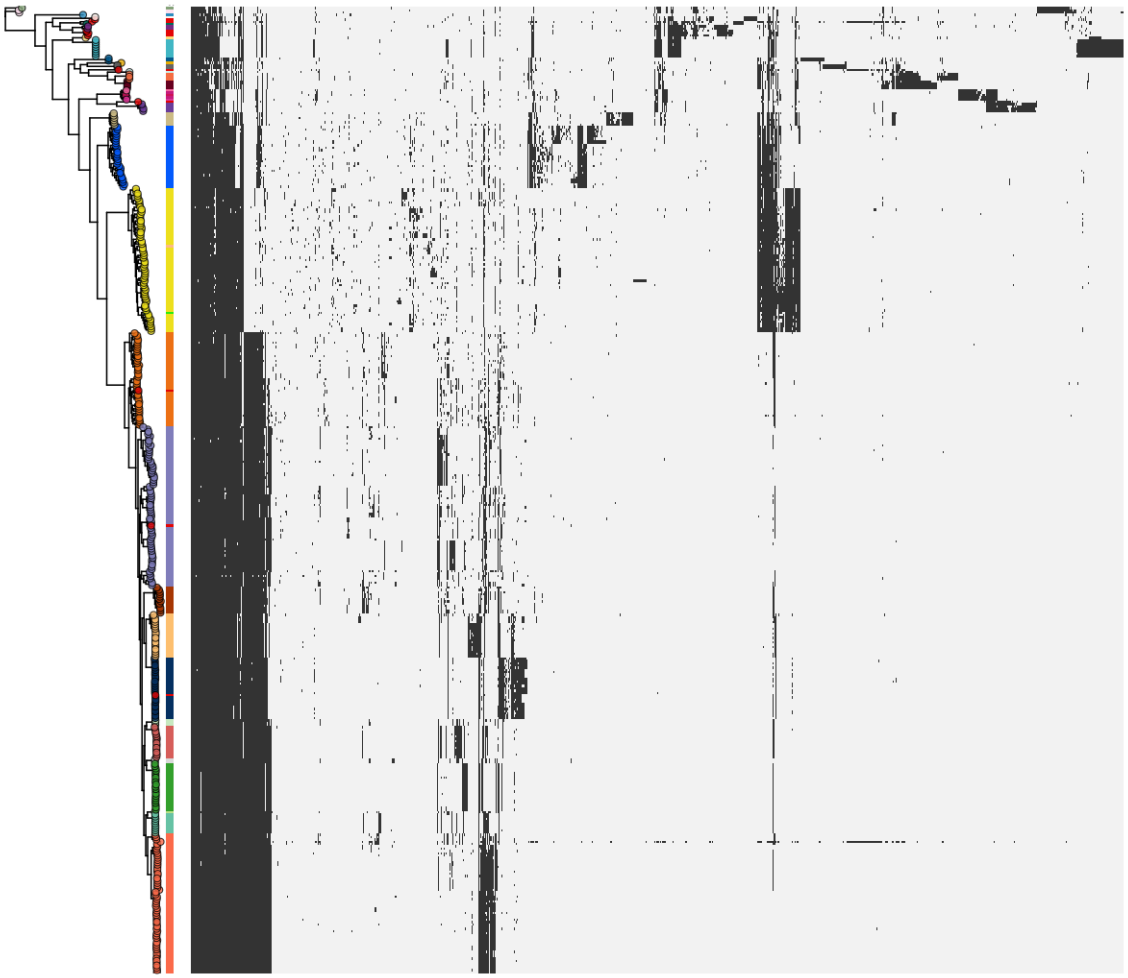

Figure 5.3: The Pangenome of *Neisseria*. Presence/Absence matrix of the 25489 genes in the *Neisseria* pangenome generated using panaroo, alongside the maximum likelihood tree in Figure 5.2. Black squares represent the presence, while white gray squares represent the absence of the gene

Comparison of functional annotation between *Neisseria bradyp* and other *Neisseria* members

The functional annotation using COG database of exclusive genes in *N. bradyp* revealed that these belonged mainly to G (Carbohydrate transport and metabolism), R (General function prediction only), M (Cell wall, membrane and envelope biogenesis), S (Function unknown) and K (Transcription) categories (Figure 5.4B).

We annotated the complete genome of five different *Neisseria* species including *N. meningitidis, N. animalis, N. iguanae, N. braziliensis and N. bradyp* using the COG database. A comparison of the number of genes in different categories reveals that there are no significant differences between the five *Neisseria* species (Figure 5.4A). In all five species, the four most representative categories were J (Translation), M (Cell wall, membrane and envelope biogenesis), E (Amino acid metabolism and transport) and C (Energy production and conversion). However, an important difference was observed: one hundred genes were annotated in the G (Carbohydrate transport and metabolism) category for *N. bradyp*; while 60, 70, 72 and 68 genes were annotated for *N. meningitidis, N. animalis, N. braziliensis* and *N. iguanae*, respectively.

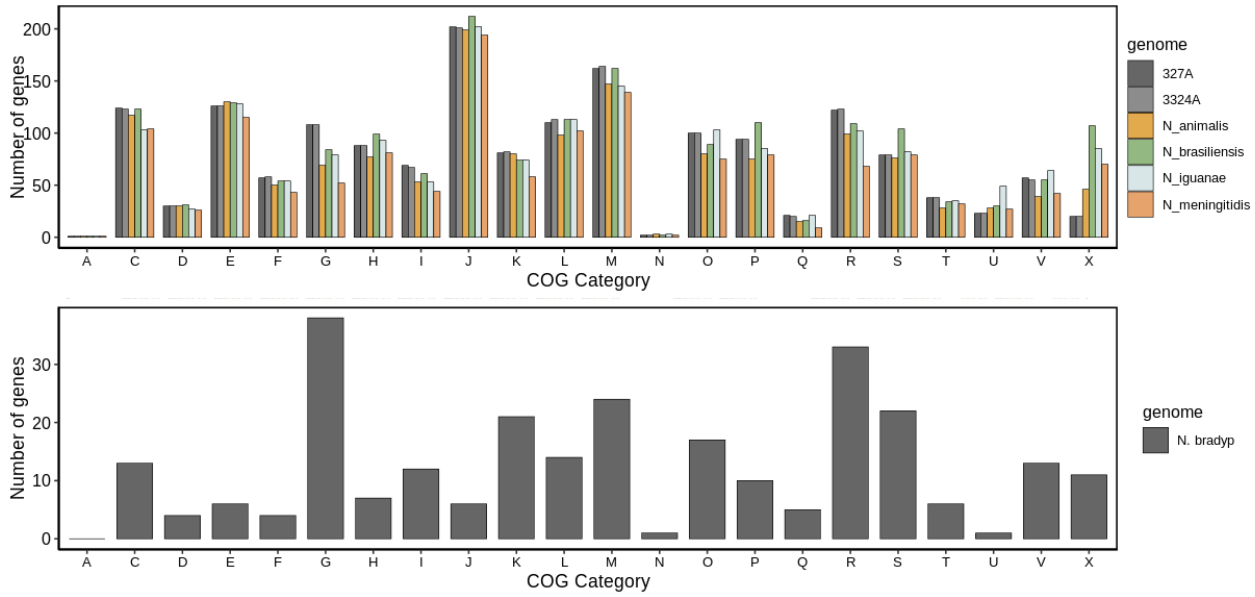

Figure 5.4. Functional annotation of *Neisseria bradyp*. Functional categories based on COG database of **A)** five different *Neisseria* species and **B)** Exclusive 480 genes in *N. bradyp* isolates.

Characterization of capsule locus in *Neisseria bradyp*

The polysaccharide capsule is an essential requisite for disease in almost all invasive meningococcal diseases (IMD) produced by *N. meningitidis* (Unkmeir et al., 2002). However, a very similar capsule structure was also recognized in many Non-pathogenic *Neisseria* (NPN) suggesting that it can be transmitted by horizontal gene transfer between *Neisseria* members (Clemence et al., 2018; Calder et al., 2020). Here we characterized the capsule locus of the novel *N. bradyp* and compared it with the capsule structure of *N. meningitidis* and *N. animalis* (Figure 5.5). The capsule locus in *N. meningitidis* is divided into several contiguous regions according to functionality: the genes required for capsule synthesis (Region A), export of the capsule (region B and C) and LPS synthesis (Region D and D'). In both *N. bradyp* isolates we identified the regions A, B, C and D with 68-80% of sequence identity compared to the same regions in *N. meningitidis*. Whereas regions A, C and D were contiguous, region B was found in a different contig. This may be due to genome assembly and a closed genome of *N. bradyp* is necessary to confirm the separation of these regions.

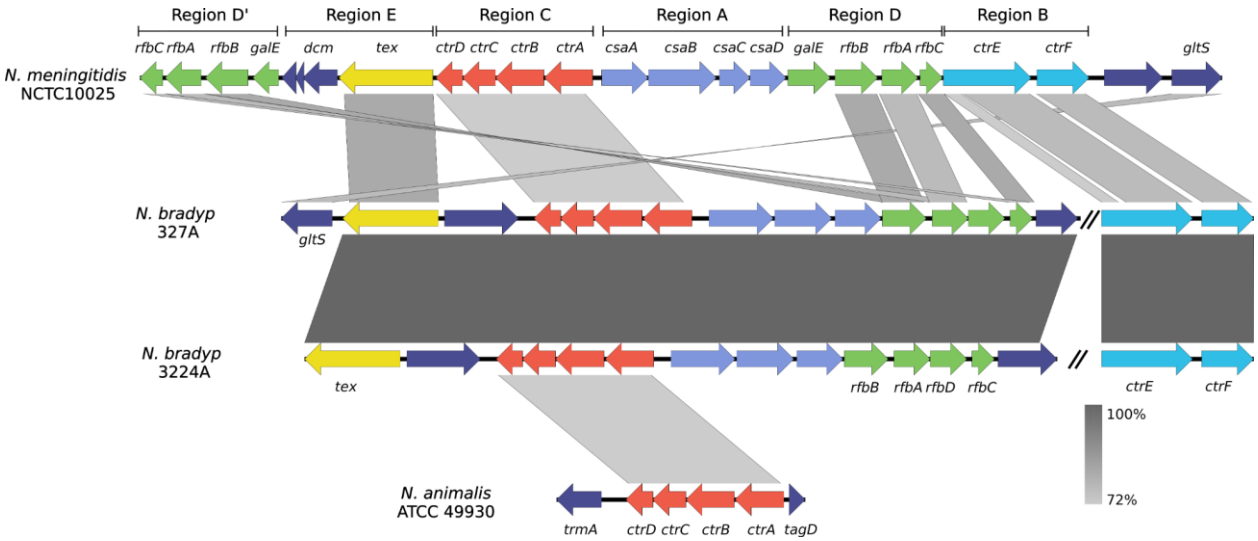

Figure 5.5: Arrangement of cps locus in *N. bradyp*. The *cps* sequence structure of 327A and 3224A isolates relative to *N. meningitidis* and *N. animalis* reference genomes. Arrows represent genes, which are color-coded by previously defined *cps* regions. Gray rectangles represent percentage nucleotide identity; darker shading indicates higher identity.

5.4. DISCUSSION

Historically, the most representative and studied species of *Neisseria* have been *N. meningitidis* and *N. gonorrhoeae* due to their impact on human health. However, the diversity of Non-pathogenic *Neisseria* (NPN), especially in other animal hosts, has received little attention. In recent years, a considerable number of new *Neisseria* species isolated from animals of different taxonomic classes have been described: *N. iguanae* from iguana (Barrett et al., 1994); *N. weixii* and *N. chenwengui* from Tibetan plateau pikas (Zhang et al., 2019a, 2019b); *N. animalis* from guinea pig, *N. zalophi* from sea lion (Volokhov et al., 2018); *N. animaloris, N. zoodegmatis* and *N. canis* from dogs (Cantas et al., 2011; Merlino et al., 2021); *N. musculi* from wild mouse (Weyand et al., 2016); *N. dentiae* from cow (Sneath and Barrett, 1996). In underrepresented human populations: *N. bergeri*, *N. blantyrii, N. uirgultaei, N. benedictiae, N. bassei, N. maigaei* and *N. viridae* from African human populations and *N. braziliensis* from Brazilian patients were also reported recently (Diallo et al., 2019; Mustapha et al., 2020). These studies reveal that the diversity within the genus *Neisseria* is greater when

we explore a wide range of host animals. Here we describe a new species of *Neisseria* isolated from brown-throated sloths (*Bradypus variegatus*) in Brazil that we name *Neisseria bradyp* representing the first species of *Neisseria* isolated from this host.

We used phylogenetic methods, ANI score, and accessory genome variation to characterize the new *Neisseria* species at the genomic level following a previous approach (Diallo et al., 2019). The comparative genomic analysis was carried out using broad diverse genomic information available for the *Neisseria* genus (420 genomes from 35 different *Neisseria spp.*) in the GenBank database (Jolley et al., 2018). The core-genome and rMLST phylogenies discretely group isolates of the same species and most human-restricted *Neisseria* spp. are more closely related to each other than to *Neisseria* spp. from other hosts, this distribution is consistent with previous studies using genomic data (Bennett et al., 2014; Maiden and Harrison, 2016). Phylogenetically, both *N. bradyp* isolates from sloths are more related to *N. animalis* from guinea pig and *N. chenwenguii* from Tibetan plateau pika, and this was confirmed with ANI analysis showing that the ANI value between *N. bradyp* and the two close species have 80.7 and 81.0% of ANI value. Taking into account the limit of 95% to differentiate species, these results support the novel species status for our isolates (Richter and Rosselló-Móra, 2009; Jain et al., 2018). *N. bradyp* represents the first *Neisseria* species isolated from common sloths. A new species isolated from Brazilian patients have previously been reported (Mustapha et al., 2020); although both *N. bradyp* and *N. braziliensis* were isolated from the same country, at the phylogenetic level it was as distant from *N. bradyp* as the rest of *Neisseria* species from humans.

A pangenome approach was used to study the accessory genome of *N. bradyp* and identify variations in genetic content in comparison to other *Neisseria* species. The core genome (present in at least 95% of genomes) of *Neisseria* composed of 546 groups represents 2.1 % of the pangenome of the genus (25,489 groups). The accessory genome variation shows gene clusters that are differentially present in different *Neisseria* lineages (Figure 5.3). A total of 480 genes were found exclusively in two *N. bradyp* isolates and absent

in the rest of the *Neisseria* species. The functional annotation of these genes reveals that the most represented categories were related to carbohydrate metabolism and transport (G) and Cell wall, membrane and envelope biogenesis (M). However, since most commensal species are underrepresented, it limits our ability to compare presence/absence patterns between species.

The polysaccharide capsule has been considered an important virulence factor only for pathogenic *N. meningitidis* (Unkmeir et al., 2002), while in other non-pathogenic species, the capsule can serve as a factor of adaptation to the host (Marri et al., 2010). The structure of the capsular locus in *N. bradyp* shows genes for transport (Region B and C) and translocation (Region D) homologous to those found in *N. meningitidis* with the observation that Region B (*crtF* and *ctrG*) are not contiguous to rest of regions (Figure 5.5); although this may be due to assembly. The putative synthesis region (region A) in *N. bradyp* differed from that in *N. meningitidis*, but we identified glycosyltransferase genes that may play a role in its synthesis. It was previously reported that the capsular structure of *N. subflava* is most similar to that of *N. meningitidis*, as it contains four contiguous regions: A, B, C, and D (Clemence et al., 2018, 2019). While other *Neisseria* species, like *N. bradyp*, have at least homologous B and C regions, the synthesis genes are not identical, although they likely serve the same function (Clemence et al., 2018). Notably, several Gram-negative bacteria also exhibit variations in their capsule structure among different isolates, such as *E. coli*, which has more than 80 distinct capsular structures (Roberts, 1996). The primary source of diversity in the capsular locus among *Neisseria* species lies in region A, and this diversity may be linked to *Neisseria's* ability to colonize different hosts or niches.

5.5. CONTRIBUTIONS

This work was carried out in collaboration with the Laboratory of Microbiology and Molecular Biology of the Parque Zoológico de São Paulo. Led by the researchers: Iris Lima

Gonzalez and Patricia Locosque Ramos who performed the sampling, isolation and microbiological characterization of Neisseria sp. strains.

This work also involved the collaboration and supervision of Prof. Aline Maria Da Silva from the IQ-USP.

Genome sequencing was done by the Joint Genome Institute, thanks to a recommendation by Professor Barny Whitman, from the University of Georgia, USA.

In this chapter, my primary contributions include: (i) bioinformatics data analyses, (ii) results interpretation and (iii) draft the manuscript.

Our intention is to prepare a manuscript based on the results of this chapter and submit it for publication in the months to follow.

6. CONCLUSION

In this dissertation, different comparative genomic approaches were applied to characterize and study the genomic variation of three bacterial species: *Salmonella* Typhimurium from guinea pigs in Peru and *Kerstersia gyiorum* and *Neisseria* sp. from brown-throated sloths in Brazil.

First, the results support the presence of two distant ST19 clusters, one of which was a monophasic variant, both lineages carried a virulent plasmid pSLT and a limited number of antimicrobial-resistant genes. Phylogenetically, the guinea pig isolates were found to be closely related to Peruvian human isolates. This information is significant because it suggests that these lineages can infect both humans and guinea pigs, involving zoonotic transmission.

Second, utilizing newly available genomic data, I pursued the objective of investigating the temporal evolution of *S.* Typhimurium in guinea pigs in Peru. The analysis unveiled the emergence of an exclusive *S.* Typhimurium lineage in Peru (B6 lineage), which appeared in the latter half of the 20th century and coincided with the recent intensification of guinea pig production. B6 lineage carries loss-of-function mutations in genes critical for host-cell interaction, potentially influencing virulence, pathogenicity, and host preference. The high prevalence of B6 isolates in guinea pigs from Peru suggests that B6 lineage has a preference for infecting guinea pigs. Future research will focus on *in-vivo* and *in-vitro* experiments to assess whether this novel *S.* Typhimurium lineage exhibits altered virulence in guinea pigs and humans.

Third, *Kestersia gyiorum* genomes from brown-throated sloths and humans were compared. The phylogenetic and pangenome analyses revealed a structured population of *K. gyiorum* according to hosts suggesting that these populations are host-specialized and possess a divergent evolutionary history. Differences in functional gene assignments indicate variations in metabolic capabilities among lineages from different hosts. An open pangenome was determined for the species in concordance with *K. gyiorum* versatility despite few mobile

genetic elements identified which may be due to the high diversity of defense mechanism systems. This study represents the first comparative analysis of *K. gyiorum* and provides new insight into mechanisms that drive the diversification of the species in different hosts.

Fourth, phylogenomic and ANI analyses demonstrate the uniqueness and continuity of *Neisseria bradyp* isolates compared to other 32 species of *Neisseria* suggesting a novel species and the first characterized in brown-throated sloths. Additionally, *N. bradyp* displayed exclusive gene content and capsular structure when compared to other *Neisseria* species, this repertoire may play a role in the host specificity of *Neisseria bradyp*.

Finally, the results presented in these four chapters highlight the importance of comparative genomics in studying the emergence of novel pathogenic variants in livestock, the underlying genetic mechanisms that drive host specialization and the exploration of the diversity and evolution of bacteria inhabiting wild hosts.

7. APPENDIX

7.1. Appendix Figures

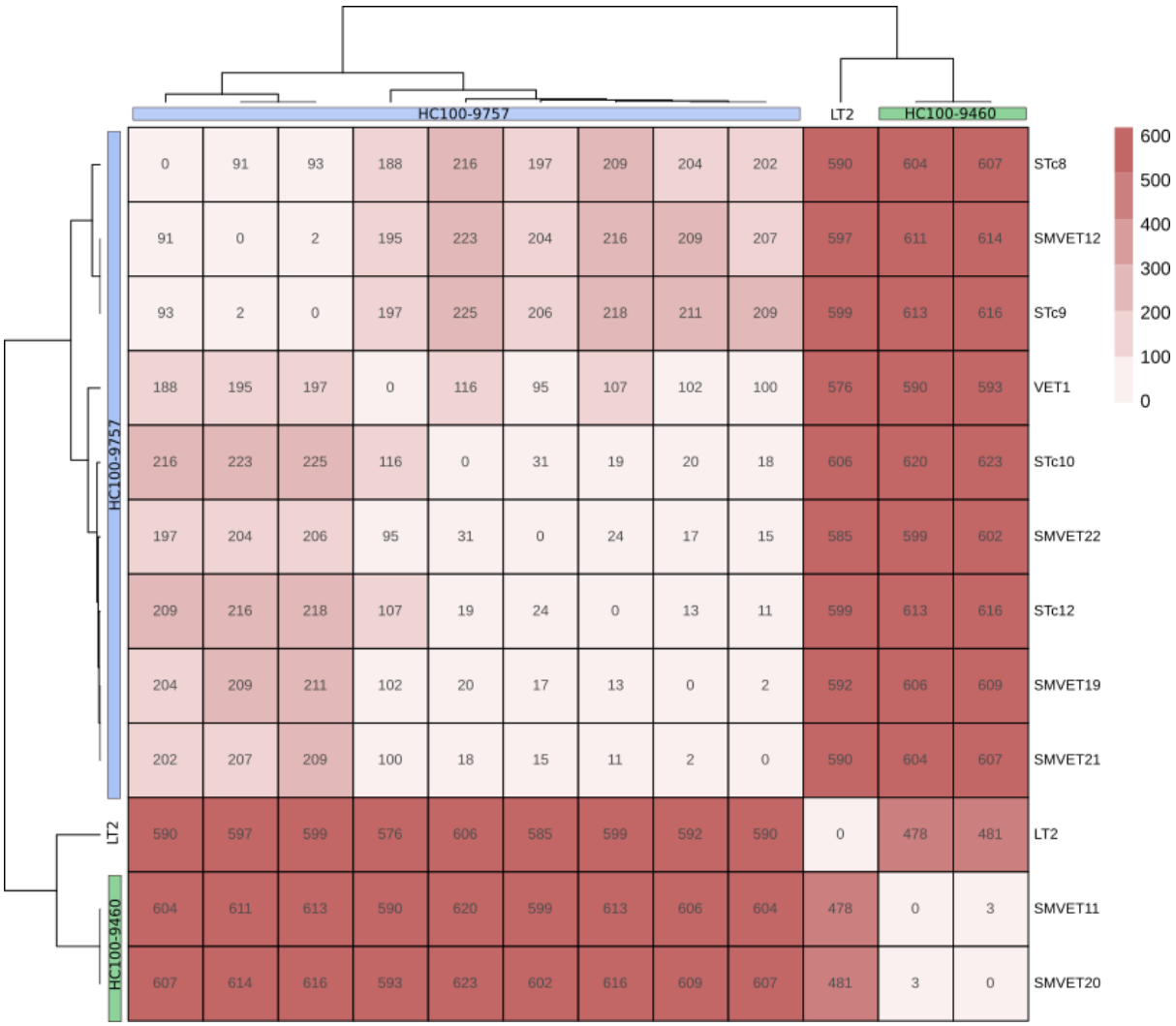

Figure A2.1. SNP distance matrix between *S.* Typhimurium genomes from guinea pigs and LT2 reference strain. Color gradients indicate the number of SNPs between isolates from lowest to highest, as shown in the legend, while the number within each square represents the number of SNPs between two genomes. Dendrograms were constructed using hierarchical clustering in the pheatmap package.

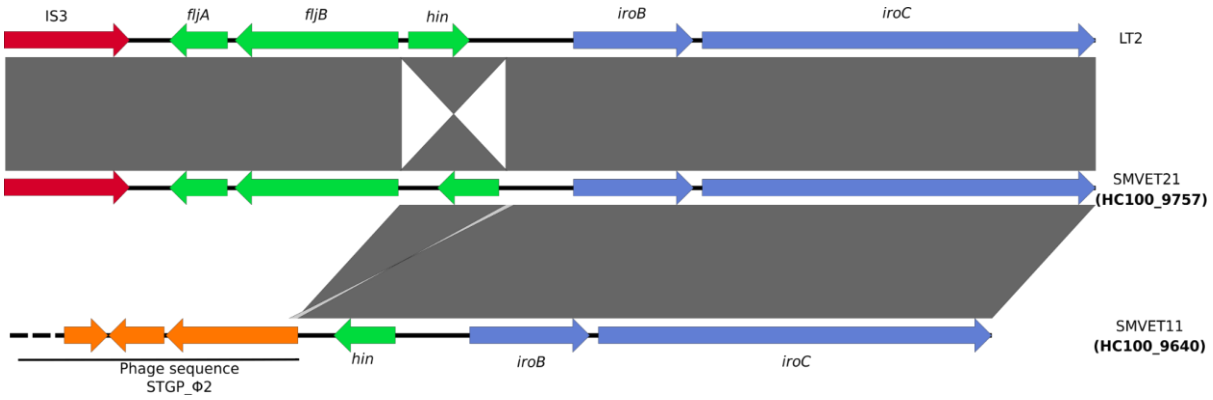

Figure A2.2. Scheme that represents gene deletion patterns in the fljAB operon. From above, the intact structure of the *fljAB* operon of the LT2 strain, followed by the intact structure of the *fjlAB* operon of the SMVET21 strain belonging to the H100_9757 cluster. Below, the *fljAB* operon deleted by the STGP_Φ2 phage of the monophasic strain SMVET11 belonging to the H100_9460 cluster.

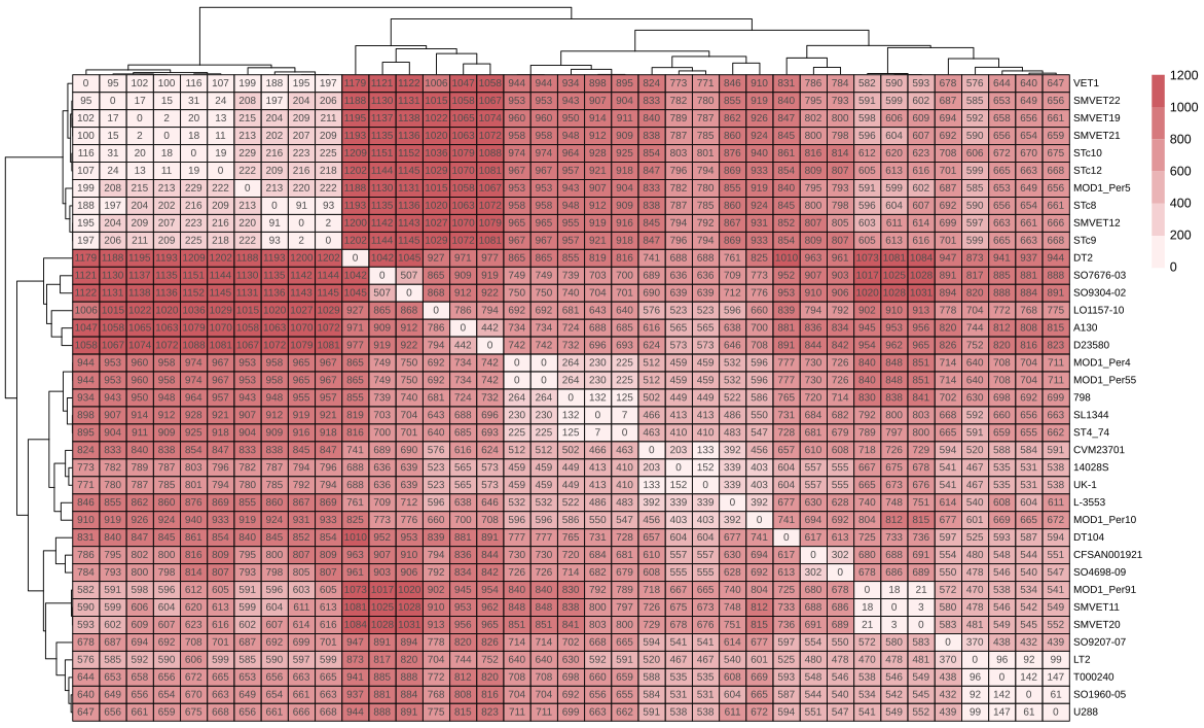

Figure A2.3. SNP distance matrix of 37 *S.* Typhimurium genomes used in this work. Color gradients indicate the number of SNPs between isolates from lowest to highest, as shown in the legend, while the number within each square represents the number of SNPs between two genomes.

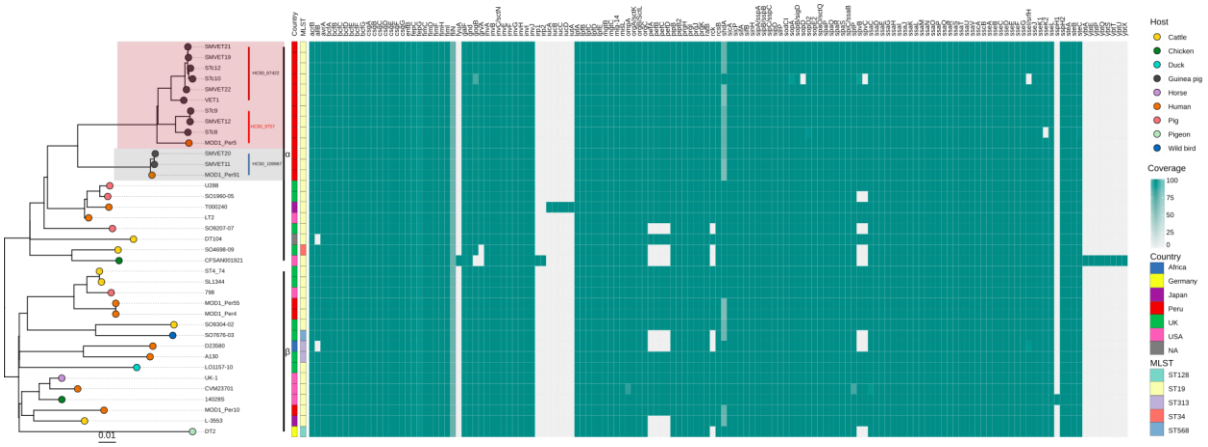

Figure A2.4. Repertoire of virulence genes in *S.* Typhimurium strains. The green square indicates the presence of the gene while the white indicates absence.

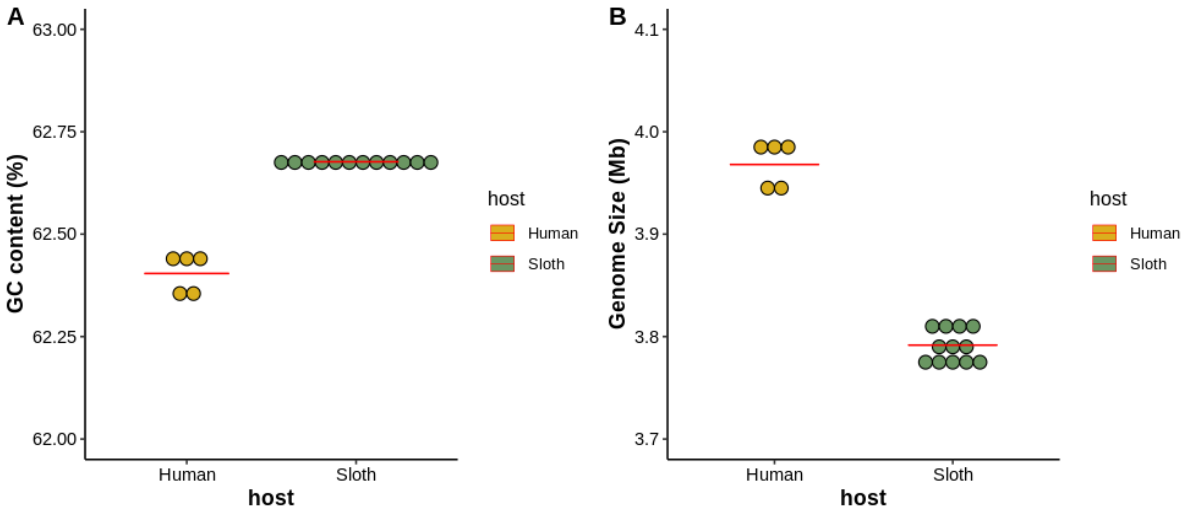

Figure A4.1. Differences in genome size and GC content between genomes from human and sloth origin

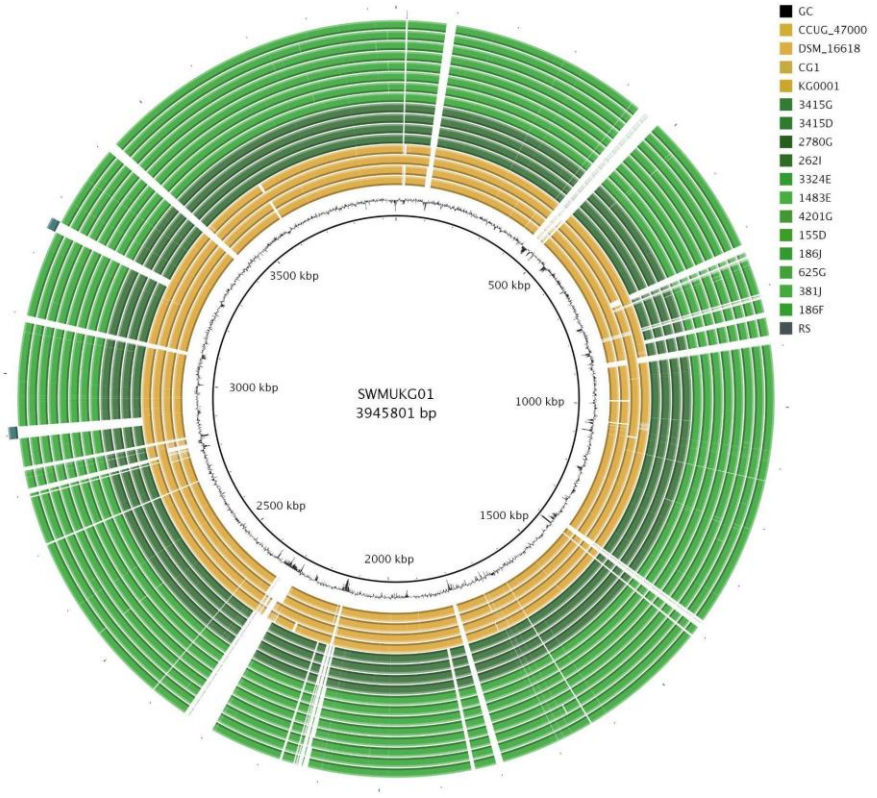

Figure A4.2. Identification of specific genomic regions in *K. gyiorum* from human origin. Blast alignment of all 16 genomes of *K. gyiorum* reveal 2 genetic regions (RH01 and RH02) present in human isolates but absent in sloth isolates

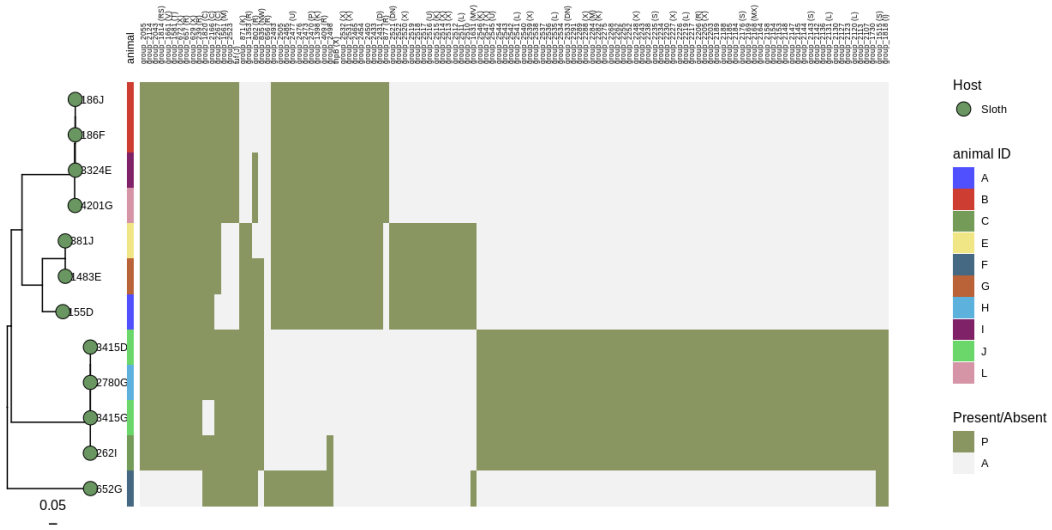

Figure A4.3. Heatmap showing gene content diversity in 12 *K. gyiorum* isolated from brown-throated sloths in Brazil

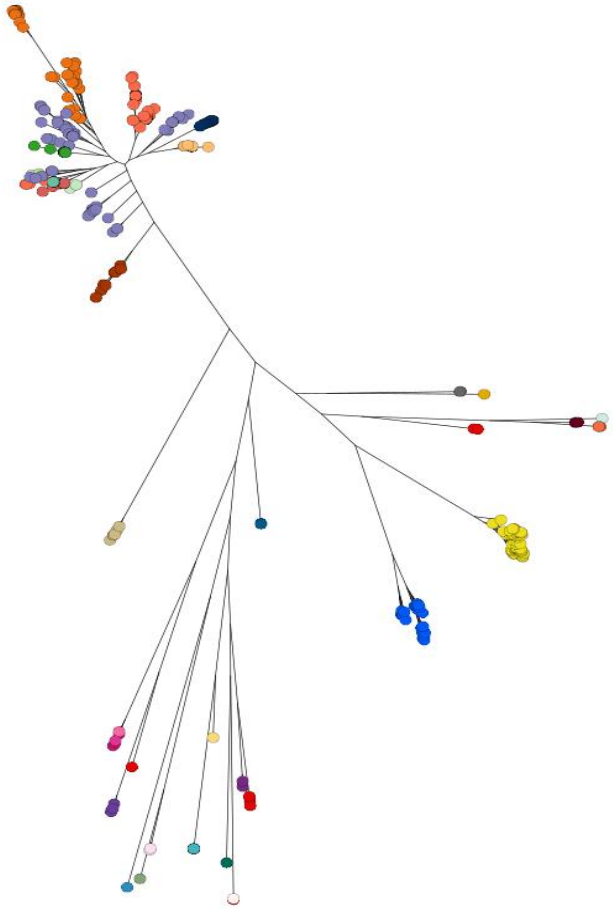

Figure A5.1. Phylogeny of the genus *Neisseria* reconstructed with 51 concatenated rps genes (rMLST)

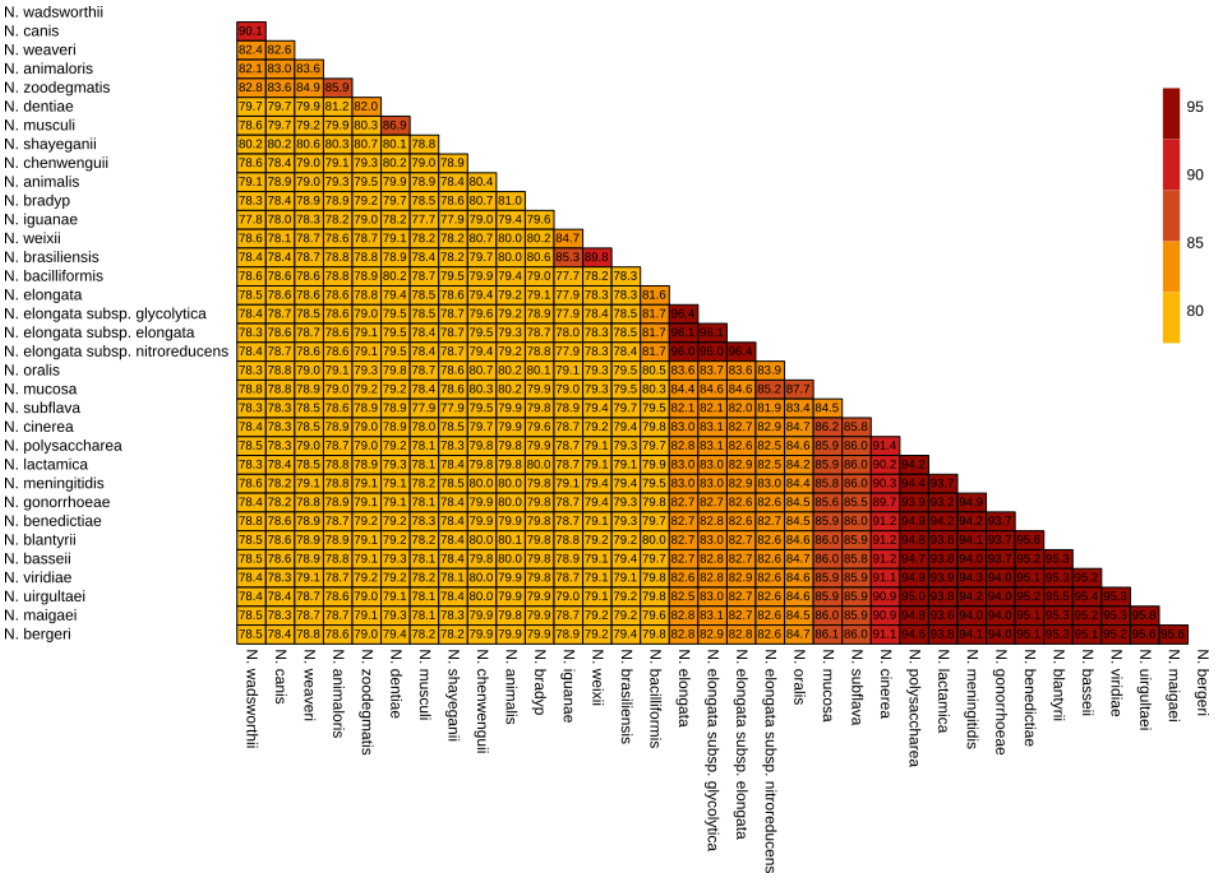

Figure A5.2. Heatmap of ANI pairwise comparison between complete genomes of 33 Neisseria species including novel N. bradyp. ANI values were calculated using FastANI and are represented as a color gradient as indicated in the ANI (%) key

7.2. Appendix Tables

To access the appendix tables, please use the following link, as some of them are quite lengthy.

https://github.com/Heraud04/MasterDissertation

8. REFERENCES

Akbar, N., Siddiqui, R., Sagathevan, K. A., and Khan, N. A. (2019). Gut bacteria of animals/pests living in polluted environments are a potential source of antibacterials. *Appl. Microbiol. Biotechnol.* 103, 3955–3964. doi: 10.1007/s00253-019-09783-2.

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. doi: 10.1093/nar/gks406.

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12, 402. doi: 10.1186/1471-2164-12-402.

Aljahdali, N. H., Foley, S. L., Han, J., Sanad, Y. M., Nayak, R., and Khajanchi, B. K. (2020). Whole-Genome Sequences of 66 Incompatibility Group FIB Plasmid-Carrying Salmonella enterica Serovar Typhimurium Isolates from Food Animal Sources. *Microbiol. Resour. Announc.* 9, e01435-19. doi: 10.1128/MRA.01435-19.

Almeida, F., Seribelli, A. A., Medeiros, M. I. C., Rodrigues, D. D. P., de MelloVarani, A., Luo, Y., et al. (2018). Phylogenetic and antimicrobial resistance gene analysis of Salmonella Typhimurium strains isolated in Brazil by whole genome sequencing. *PloS One* 13, e0201882. doi: 10.1371/journal.pone.0201882.

Almuzara, M. N., Barberis, C. M., Traglia, G. M., Ordoñez, A. M., Famiglietti, A. M. R., Ramirez, M. S., et al. (2012). Isolation of Kerstersia gyiorum from a Patient with Cholesteatomatous Chronic Otitis Media. *J. Clin. Microbiol.* 50, 3809–3811. doi: 10.1128/JCM.02051-12.

Andrews S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Argimón, S., Yeats, C. A., Goater, R. J., Abudahab, K., Taylor, B., Underwood, A., et al. (2021). A global resource for genomic predictions of antimicrobial resistance and surveillance of Salmonella Typhi at pathogenwatch. *Nat. Commun.* 12, 2879. doi: 10.1038/s41467-021-23091-2.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387.

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. doi: 10.1038/s41579-021-00650-4.

Arrieta-Gisasola, A., Atxaerandio-Landa, A., Garrido, V., Grilló, M. J., Martínez-Ballesteros, I., Laorden, L., et al. (2020). Genotyping Study of Salmonella 4,[5],12:i:- Monophasic Variant of Serovar Typhimurium and Characterization of the Second-Phase Flagellar Deletion by Whole Genome Sequencing. *Microorganisms* 8, 2049. doi: 10.3390/microorganisms8122049.

Ashton, P. M., Owen, S. V., Kaindama, L., Rowe, W. P. M., Lane, C. R., Larkin, L., et al. (2017). Public health surveillance in the UK revolutionises our understanding of the invasive Salmonella Typhimurium epidemic in Africa. *Genome Med.* 9, 92. doi: 10.1186/s13073-017-0480-7.

Bakkeren, E., Gül, E., Huisman, J. S., Steiger, Y., Rocker, A., Hardt, W.-D., et al. (2022). Impact of horizontal gene transfer on emergence and stability of cooperative virulence in Salmonella Typhimurium. *Nat. Commun.* 13, 1939. doi: 10.1038/s41467-022-29597-7.

Banerji, S., Simon, S., Tille, A., Fruth, A., and Flieger, A. (2020). Genome-based Salmonella serotyping as the new gold standard. *Sci. Rep.* 10, 4333. doi: 10.1038/s41598-020-61254-1.

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. doi: 10.1089/cmb.2012.0021.
Barrett, S. j., Schlater, L. K., Montali, R. j., and Sneath, P. h. a. (1994). A new species of Neisseria from iguanid lizards, Neisseria iguanae sp. nov. *Lett. Appl. Microbiol.* 18, 200–202. doi: 10.1111/j.1472-765X.1994.tb00846.x.
Bauer, A. W., Kirby, W. M., Sherris, J. C., and Turck, M. (1966). Antibiotic susceptibility testing by a standardized single disk method. *Am. J. Clin. Pathol.* 45, 493–496.
Bawn, M., Alikhan, N.-F., Thilliez, G., Kirkwood, M., Wheeler, N. E., Petrovska, L., et al. (2020). Evolution of Salmonella enterica serotype Typhimurium driven by anthropogenic selection and niche adaptation. *PLOS Genet.* 16, e1008850. doi: 10.1371/journal.pgen.1008850.
Bennett, J. S., Bratcher, H. B., Brehony, C., Harrison, O. B., and Maiden, M. C. J. (2014). "The Genus Neisseria," in *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer), 881–900. doi: 10.1007/978-3-642-30197-1_241.
Bennett, J. S., Jolley, K. A., Earle, S. G., Corton, C., Bentley, S. D., Parkhill, J., et al. (2012). A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus Neisseria. *Microbiol. Read. Engl.* 158, 1570–1580. doi: 10.1099/mic.0.056077-0.
Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170.
Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. Ø. (2016). Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proceedings of the National Academy of Sciences of the United States of America, 113(26), E3801–E3809. https://doi.org/10.1073/pnas.1523199113
Branchu, P., Bawn, M., and Kingsley, R. A. (2018). Genome Variation and Molecular Epidemiology of Salmonella enterica Serovar Typhimurium Pathovariants. *Infect. Immun.* 86, e00079-18. doi: 10.1128/IAI.00079-18.
Brickman, T. J., and Armstrong, S. K. (2005). Bordetella AlcS Transporter Functions in Alcaligin Siderophore Export and Is Central to Inducer Sensing in Positive Regulation of Alcaligin System Gene Expression. *J. Bacteriol.* 187, 3650–3661. doi: 10.1128/JB.187.11.3650-3661.2005.
Bush, N. G., Diez-Santos, I., Abbott, L. R., and Maxwell, A. (2020). Quinolones: Mechanism, Lethality and Their Contributions to Antibiotic Resistance. *Molecules* 25, 5662. doi: 10.3390/molecules25235662.
Cai, R., Wang, G., Le, S., Wu, M., Cheng, M., Guo, Z., et al. (2019). Three Capsular Polysaccharide Synthesis-Related Glucosyltransferases, GT-1, GT-2 and WcaJ, Are Associated With Virulence and Phage Sensitivity of Klebsiella pneumoniae. *Front. Microbiol.* 10. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01189 [Accessed November 21, 2022].
Calchi, A. C., Vultão, J. G., Alves, M. H., Yogui, D. R., Desbiez, A. L. J., De Santi, M., et al. (2020). Ehrlichia spp. and Anaplasma spp. in Xenarthra mammals from Brazil, with evidence of novel 'Candidatus Anaplasma spp.' *Sci. Rep.* 10, 12615. doi: 10.1038/s41598-020-69263-w.
Calder, A., Menkiti, C. J., Çağdaş, A., Lisboa Santos, J., Streich, R., Wong, A., et al. (2020). Virulence genes and previously unexplored gene clusters in four commensal Neisseria spp. isolated from the human throat expand the neisserial gene repertoire. *Microb. Genomics* 6, mgen000423. doi: 10.1099/mgen.0.000423.
Cantas, H., Pekarkova, M., Kippenes, H. S., Brudal, E., and Sorum, H. (2011). First Reported Isolation of Neisseria canis from a Deep Facial Wound Infection in a Dog▽ . *J. Clin. Microbiol.* 49, 2043–2046. doi: 10.1128/JCM.02610-10.
Cantlay, J. C., Ingram, D. J., and Meredith, A. L. (2017). A Review of Zoonotic Infection Risks Associated with the Wild Meat Trade in Malaysia. *EcoHealth* 14, 361–388. doi:

10.1007/s10393-017-1229-x.
Carattoli, A., and Hasman, H. (2020). PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). *Methods Mol. Biol. Clifton NJ* 2075, 285–294. doi: 10.1007/978-1-4939-9877-7_20.
Carattoli, A., Villa, L., Fortini, D., and García-Fernández, A. (2021). Contemporary IncI1 plasmids involved in the transmission and spread of antimicrobial resistance in Enterobacteriaceae. *Plasmid* 118, 102392. doi: 10.1016/j.plasmid.2018.12.001.
Carhuaricra Huaman, D. E., Luna Espinoza, L. R., Rodríguez Cueva, C. L., Duran Gonzales, C. G., Rosadio Alcántara, R. H., Setubal, J. C., et al. (2022). Genomic Characterization of Salmonella Typhimurium Isolated from Guinea Pigs with Salmonellosis in Lima, Peru. *Microorganisms* 10, 1726. doi: 10.3390/microorganisms10091726.
Carro, L., Peix, Á., and Velázquez, E. (2021). "The Taxonomy of Bacteria in the Genomic Era," in *Developmental Biology in Prokaryotes and Lower Eukaryotes*, eds. T. G. Villa and T. de Miguel Bouzas (Cham: Springer International Publishing), 289–309. doi: 10.1007/978-3-030-77595-7_12.
Carroll, L. M., Wiedmann, M., den Bakker, H., Siler, J., Warchocki, S., Kent, D., et al. (2017). Whole-Genome Sequencing of Drug-Resistant Salmonella enterica Isolates from Dairy Cattle and Humans in New York and Washington States Reveals Source and Geographic Associations. *Appl. Environ. Microbiol.* 83, e00140-17. doi: 10.1128/AEM.00140-17.
Casjens, S. (1998). The Diverse and Dynamic Structure of Bacterial Genomes. *Annu. Rev. Genet.* 32, 339–377. doi: 10.1146/annurev.genet.32.1.339.
Catenacci, L. S., Ferreira, M., Martins, L. C., De Vleeschouwer, K. M., Cassano, C. R., Oliveira, L. C., et al. (2018). Surveillance of Arboviruses in Primates and Sloths in the Atlantic Forest, Bahia, Brazil. *EcoHealth* 15, 777–791. doi: 10.1007/s10393-018-1361-2.
Centers for Disease Control and Prevention (2018). National Salmonella surveillance annual report, 2016. Available at: https://www.cdc.gov/nationalsurveillance/salmonella-surveillance.html [Accessed August 3, 2022].
Chauca Francia, L. J. (2013). Crianza del cuy (Cavia porcellus) y su impacto en el desarrollo rural. *Inst. Nac. Innov. Agrar.* Available at: https://repositorio.inia.gob.pe/handle/20.500.12955/440 [Accessed August 31, 2023].
Chauca L. (1997). Producción de cuyes (Cavia porcellus) -. Rome Available at: https://www.fao.org/3/w6562s/w6562s00.htm [Accessed June 2, 2022].
Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516.
Chun, J., and Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.* 64, 316–324. doi: 10.1099/ijs.0.054171-0.
Clemence, M. E. A., Harrison, O. B., and Maiden, M. C. J. (2019). Neisseria meningitidis has acquired sequences within the capsule locus by horizontal genetic transfer. *Wellcome Open Res.* 4, 99. doi: 10.12688/wellcomeopenres.15333.2.
Clemence, M. E. A., Maiden, M. C. J., and Harrison, O. B. (2018). Characterization of capsule genes in non-pathogenic Neisseria species. *Microb. Genomics* 4, e000208. doi: 10.1099/mgen.0.000208.
CLSI (2021). Performance Standards for Antimicrobial Disk Susceptibility Tests, 31th Edition. Available at: https://clsi.org/standards/products/microbiology/documents/m02/.
Coenye, T., Vancanneyt, M., Cnockaert, M. C., Falsen, E., Swings, J., and Vandamme, P. (2003). Kerstersia gyiorum gen. nov., sp. nov., a novel Alcaligenes faecalis-like organism isolated from human clinical samples, and reclassification of Alcaligenes denitrificans Rüger and Tan 1983 as Achromobacter denitrificans comb. nov. *Int. J. Syst. Evol. Microbiol.* 53, 1825–1831. doi: 10.1099/ijs.0.02609-0.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43, e15. doi: 10.1093/nar/gku1196.

Csárdi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal* Complex Systems. Available at: https://zenodo.org/record/8240644 [Accessed September 11, 2023].

Cuenca-Arias, P., Montaño, L. A., Villarreal, J. M., and Wiesner, M. (2020). Molecular and phenotypic characterization of Salmonella Typhimurium monophasic variant (1,4,[5],12:i:-) from Colombian clinical isolates. *Biomed. Rev. Inst. Nac. Salud* 40, 722–733. doi: 10.7705/biomedica.5417.

Dalmutt, A. C., Moreno, L. Z., Gomes, V. T. M., Cunha, M. P. V., Barbosa, M. R. F., Sato, M. I. Z., et al. (2020). Characterization of bacterial contaminants of boar semen: identification by MALDI-TOF mass spectrometry and antimicrobial susceptibility profiling. *J. Appl. Anim. Res.* 48, 559–565. doi: 10.1080/09712119.2020.1848845.

Delsuc, F., Metcalf, J. L., Wegener Parfrey, L., Song, S. J., González, A., and Knight, R. (2014). Convergence of gut microbiomes in myrmecophagous mammals. *Mol. Ecol.* 23, 1301–1317. doi: 10.1111/mec.12501.

Deng, Y., He, L., Chen, S., Zheng, H., Zeng, Z., Liu, Y., et al. (2011). F33:A−:B− and F2:A−:B− Plasmids Mediate Dissemination of rmtB-blaCTX-M-9 Group Genes and rmtB-qepA in Enterobacteriaceae Isolates from Pets in China▽ . *Antimicrob. Agents Chemother.* 55, 4926–4929. doi: 10.1128/AAC.00133-11.

Diallo, K., MacLennan, J., Harrison, O. B., Msefula, C., Sow, S. O., Daugla, D. M., et al. (2019). Genomic characterization of novel Neisseria species. *Sci. Rep.* 9, 13742. doi: 10.1038/s41598-019-50203-2.

Diard, M., Bakkeren, E., Lentsch, V., Rocker, A., Bekele, N. A., Hoces, D., et al. (2021). A rationally designed oral vaccine induces immunoglobulin A in the murine gut that directs the evolution of attenuated Salmonella variants. *Nat. Microbiol.* 6, 830–841. doi: 10.1038/s41564-021-00911-1.

Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R., and Wilson, D. J. (2018). Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* 46, e134. doi: 10.1093/nar/gky783.

Dill-McFarland, K. A., Weimer, P. J., Pauli, J. N., Peery, M. Z., and Suen, G. (2016). Diet specialization selects for an unusual and simplified gut microbiota in two- and three-toed sloths. *Environ. Microbiol.* 18, 1391–1402. doi: 10.1111/1462-2920.13022.

Domingo-Sananes, M. R., and McInerney, J. O. (2021). Mechanisms That Shape Microbial Pangenomes. *Trends Microbiol.* 29, 493–503. doi: 10.1016/j.tim.2020.12.004.

Dróżdż, M., Małaszczuk, M., Paluch, E., and Pawlak, A. (2021). Zoonotic potential and prevalence of Salmonella serovars isolated from pets. *Infect. Ecol. Epidemiol.* 11, 1975530. doi: 10.1080/20008686.2021.1975530.

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214. doi: 10.1186/1471-2148-7-214.

Duran Gonzalez, C., Luna Espinoza, L., Carhuaricra Huamán, D., Salvatierra Rodríguez, G., Rosadio Alcántara, R., Maturrano Hernández, L., et al. (2021). Evaluación de factores de virulencia en cepas de Salmonella Typhimurium aisladas de cuyes (Cavia porcellus) enfermos y sanos. *Rev. Investig. Vet. Perú* 32. doi: 10.15381/rivep.v32i5.21331.

Durrant, M. G., Li, M. M., Siranosian, B. A., Montgomery, S. B., and Bhatt, A. S. (2020). A Bioinformatic Analysis of Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. *Cell Host Microbe* 27, 140-153.e9. doi: 10.1016/j.chom.2019.10.022.

Ellabaan, M. M. H., Munck, C., Porse, A., Imamovic, L., and Sommer, M. O. A. (2021). Forecasting the dissemination of antibiotic resistance genes across bacterial genomes. *Nat. Commun.* 12, 2435. doi: 10.1038/s41467-021-22757-1.

European Centre for Disease Prevention and Control (2017). The European Union summary

report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. *EFSA J.* 15, e05077. doi: 10.2903/j.efsa.2017.5077.
Fang, X., Ahmad, I., Blanka, A., Schottkowski, M., Cimdins, A., Galperin, M. Y., et al. (2014). GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. *Mol. Microbiol.* 93, 439–452. doi: 10.1111/mmi.12672.
Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J. G., Haendiges, J., Haft, D. H., et al. (2021). AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* 11, 12728. doi: 10.1038/s41598-021-91456-0.
Fenske, G. J., Thachil, A., McDonough, P. L., Glaser, A., and Scaria, J. (2019). Geography Shapes the Population Genomics of Salmonella enterica Dublin. *Genome Biol. Evol.* 11, 2220–2231. doi: 10.1093/gbe/evz158.
Fernandes, M., Nóbrega Carneiro, C., Villada Rosales, A. M., Grilo, M., Ramiro, Y., Cunha, E., et al. (2022). Antimicrobial resistance and virulence profiles of Enterobacterales isolated from two-finger and three-finger sloths (Choloepus hoffmanni and Bradypus variegatus) of Costa Rica. *PeerJ* 10, e12911. doi: 10.7717/peerj.12911.
Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 269, 496–512. doi: 10.1126/science.7542800.
Foster-Nyarko, E., and Pallen, M. J. (2022). The microbial ecology of Escherichia coli in the vertebrate gut. *FEMS Microbiol. Rev.* 46, fuac008. doi: 10.1093/femsre/fuac008.
Fournier, J. B., Knox, K., Harris, M., and Newstein, M. (2015). Family Outbreaks of Nontyphoidal Salmonellosis following a Meal of Guinea Pigs. *Case Rep. Infect. Dis.* 2015, e864640. doi: 10.1155/2015/864640.
Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., et al. (1995). The Minimal Gene Complement of Mycoplasma genitalium. *Science* 270, 397–404. doi: 10.1126/science.270.5235.397.
Fu, Y., Smith, J. C., Shariat, N. W., M'ikanatha, N. M., and Dudley, E. G. (2022). Evidence for common ancestry and microevolution of passerine-adapted Salmonella enterica serovar Typhimurium in the UK and USA. *Microb. Genomics* 8, 000775. doi: 10.1099/mgen.0.000775.
Gal-Mor, O., Boyle, E. C., and Grassl, G. A. (2014). Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ. *Front. Microbiol.* 5, 391. doi: 10.3389/fmicb.2014.00391.
Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 49, D274–D281. doi: 10.1093/nar/gkaa1018.
Garcia, C., Hinostroza, N., Astocondor, L., Ochoa, T., and Jacobs, J. (2019). Characterization of ESBL-Producing Salmonella enterica Serovar Infantis Infection in Humans, Lima, Peru. *Am. J. Trop. Med. Hyg.* 101, 746–748. doi: 10.4269/ajtmh.19-0213.
Gardy, J. L., and Loman, N. J. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* 19, 9–20. doi: 10.1038/nrg.2017.88.
Gensollen, T., Iyer, S. S., Kasper, D. L., and Blumberg, R. S. (2016). How colonization by microbiota in early life shapes the immune system. *Science* 352, 539–544. doi: 10.1126/science.aad9378.
Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., et al. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9, 37. doi: 10.1186/s40168-020-00990-y.
Gupta, A. K., Nayduch, D., Verma, P., Shah, B., Ghate, H. V., Patole, M. S., et al. (2012). Phylogenetic characterization of bacteria in the gut of house flies (Musca domestica L.). *FEMS Microbiol. Ecol.* 79, 581–593. doi: 10.1111/j.1574-6941.2011.01248.x.
Han, J., Lynne, A. M., David, D. E., Tang, H., Xu, J., Nayak, R., et al. (2012). DNA sequence analysis of plasmids from multidrug resistant Salmonella enterica serotype

Heidelberg isolates. *PloS One* 7, e51160. doi: 10.1371/journal.pone.0051160.
Havelaar, A. H., Kirk, M. D., Torgerson, P. R., Gibb, H. J., Hald, T., Lake, R. J., et al. (2015). World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. *PLOS Med.* 12, e1001923. doi: 10.1371/journal.pmed.1001923.
Hayssen, V. (2010). Bradypus variegatus (Pilosa: Bradypodidae). *Mamm. Species* 42, 19–32. doi: 10.1644/850.1.
Hendriksen, S. W. M., Orsel, K., Wagenaar, J. A., Miko, A., and van Duijkeren, E. (2004). Animal-to-Human Transmission of Salmonella Typhimurium DT104A Variant. *Emerg. Infect. Dis.* 10, 2225–2227. doi: 10.3201/eid1012.040286.
Heydecke, A., Andersson, B., Holmdahl, T., and Melhus, Å. (2013). Human wound infections caused by Neisseria animaloris and Neisseria zoodegmatis, former CDC Group EF-4a and EF-4b. *Infect. Ecol. Epidemiol.* 3, 20312. doi: 10.3402/iee.v3i0.20312.
Hoffman, C. L., Gonyar, L. A., Zacca, F., Sisti, F., Fernandez, J., Wong, T., et al. (2019). Bordetella pertussis Can Be Motile and Express Flagellum-Like Structures. *mBio* 10, e00787-19. doi: 10.1128/mBio.00787-19.
Holt, K. E., Thomson, N. R., Wain, J., Langridge, G. C., Hasan, R., Bhutta, Z. A., et al. (2009). Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. *BMC Genomics* 10, 36. doi: 10.1186/1471-2164-10-36.
Huamán, M., Pérez, C., Rodríguez, J., Killerby, M., Lovón, S., Chauca, L., et al. (2020). Caracterización genética y patrones de resistencia antimicrobiana en cepas de Salmonella enterica subsp. enterica serovar Typhimurium en cuyes de crianza intensiva. *Rev. Investig. Vet. Perú* 31. Available at: http://www.scielo.org.pe/scielo.php?script=sci_abstract&pid=S1609-91172020000100016&lng=es&nrm=iso&tlng=es [Accessed June 3, 2022].
Hurtado, R., Barh, D., Weimer, B. C., Viana, M. V. C., Profeta, R., Sousa, T. J., et al. (2022). WGS-Based Lineage and Antimicrobial Resistance Pattern of Salmonella Typhimurium Isolated during 2000–2017 in Peru. *Antibiotics* 11, 1170. doi: 10.3390/antibiotics11091170.
Hurtado, R., Carhuaricra, D., Soares, S., Viana, M. V. C., Azevedo, V., Maturrano, L., et al. (2018). Pan-genomic approach shows insight of genetic divergence and pathogenic-adaptation of Pasteurella multocida. *Gene* 670, 193–206. doi: 10.1016/j.gene.2018.05.084.
Ido, N., Lee, K., Iwabuchi, K., Izumiya, H., Uchida, I., Kusumoto, M., et al. (2014). Characteristics of Salmonella enterica Serovar 4,[5],12:i:- as a Monophasic Variant of Serovar Typhimurium. *PLoS ONE* 9, e104380. doi: 10.1371/journal.pone.0104380.
Ikhimiukor, O. O., Odih, E. E., Donado-Godoy, P., and Okeke, I. N. (2022). A bottom-up view of antimicrobial resistance transmission in developing countries. *Nat. Microbiol.* 7, 757–765. doi: 10.1038/s41564-022-01124-w.
Innamorati, K. A., Earl, J. P., Aggarwal, S. D., Ehrlich, G. D., and Hiller, N. L. (2020). “The Bacterial Guide to Designing a Diversified Gene Portfolio,” in *The Pangenome: Diversity, Dynamics and Evolution of Genomes*, eds. H. Tettelin and D. Medini (Cham: Springer International Publishing), 51–87. doi: 10.1007/978-3-030-38281-0_3.
Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 5114. doi: 10.1038/s41467-018-07641-9.
Jamshidi, A., Kalidari, G. a., and Hedayati, M. (2010). Isolation and Identification of Salmonella Enteritidis and Salmonella Typhimurium from the Eggs of Retail Stores in Mashhad, Iran Using Conventional Culture Method and Multiplex Pcr Assay. *J. Food Saf.* 30, 558–568. doi: 10.1111/j.1745-4565.2010.00225.x.
Jia, T., Chang, W.-S., Marcelino, V. R., Zhao, S., Liu, X., You, Y., et al. (2022). Characterization of the Gut Microbiome and Resistomes of Wild and Zoo-Captive Macaques. *Front. Vet. Sci.* 8. Available at:

https://www.frontiersin.org/article/10.3389/fvets.2021.778556 [Accessed May 3, 2022].
Johnson, T. J., Shepard, S. M., Rivet, B., Danzeisen, J. L., and Carattoli, A. (2011). Comparative genomics and phylogeny of the IncI1 plasmids: a common plasmid type among porcine enterotoxigenic Escherichia coli. *Plasmid* 66, 144–151. doi: 10.1016/j.plasmid.2011.07.003.
Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3, 124. doi: 10.12688/wellcomeopenres.14826.1.
Jones, B. A., Grace, D., Kock, R., Alonso, S., Rushton, J., Said, M. Y., et al. (2013). Zoonosis emergence linked to agricultural intensification and environmental change. *Proc. Natl. Acad. Sci. U. S. A.* 110, 8399–8404. doi: 10.1073/pnas.1208059110.
Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., et al. (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–993. doi: 10.1038/nature06536.
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. doi: 10.1093/molbev/mst010.
Key, F. M., Posth, C., Esquivel-Gomez, L. R., Hübler, R., Spyrou, M. A., Neumann, G. U., et al. (2020). Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. *Nat. Ecol. Evol.* 4, 324–333. doi: 10.1038/s41559-020-1106-9.
Kim, Y., Gu, C., Kim, H. U., and Lee, S. Y. (2020). Current status of pan-genome analysis for pathogenic bacteria. *Curr. Opin. Biotechnol.* 63, 54–62. doi: 10.1016/j.copbio.2019.12.001.
Kingsley, R. A., Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. A., et al. (2009). Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res.* 19, 2279–2287. doi: 10.1101/gr.091017.109.
Kirkwood, M., Vohra, P., Bawn, M., Thilliez, G., Pye, H., Tanner, J., et al. (2021). Ecological niche adaptation of Salmonella Typhimurium U288 is associated with altered pathogenicity and reduced zoonotic potential. *Commun. Biol.* 4, 1–15. doi: 10.1038/s42003-021-02013-4.
Kolenda, R., Ugorski, M., and Grzymajlo, K. (2019). Everything You Always Wanted to Know About Salmonella Type 1 Fimbriae, but Were Afraid to Ask. *Front. Microbiol.* 10. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01017 [Accessed September 4, 2023].
Kruse, H., Kirkemo, A.-M., and Handeland, K. (2004). Wildlife as Source of Zoonotic Infections. *Emerg. Infect. Dis.* 10, 2067–2072. doi: 10.3201/eid1012.040707.
Lan, R., Reeves, P. R., and Octavia, S. (2009). Population structure, origins and evolution of major Salmonella enterica clones. *Infect. Genet. Evol.* 9, 996–1005. doi: 10.1016/j.meegid.2009.04.011.
Langridge, G. C., Fookes, M., Connor, T. R., Feltwell, T., Feasey, N., Parsons, B. N., et al. (2015). Patterns of genome evolution that have accompanied host adaptation in Salmonella. *Proc. Natl. Acad. Sci.* 112, 863–868. doi: 10.1073/pnas.1416707112.
Lanzas, C., Warnick, L. D., James, K. L., Wright, E. M., Wiedmann, M., and Gröhn, Y. T. (2010). Transmission Dynamics of a Multidrug-Resistant Salmonella Typhimurium Outbreak in a Dairy Farm. *Foodborne Pathog. Dis.* 7, 467–474. doi: 10.1089/fpd.2009.0411.
Leon, I. M., Lawhon, S. D., Norman, K. N., Threadgill, D. S., Ohta, N., Vinasco, J., et al. (2018). Serotype Diversity and Antimicrobial Resistance among Salmonella enterica Isolates from Patients at an Equine Referral Hospital. *Appl. Environ. Microbiol.* 84, e02829-17. doi: 10.1128/AEM.02829-17.
Leverstein-van Hall, M. A., Dierikx, C. M., Cohen Stuart, J., Voets, G. M., van den Munckhof, M. P., van Essen-Zandbergen, A., et al. (2011). Dutch patients, retail chicken meat

and poultry share the same ESBL genes, plasmids and strains. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 17, 873–880. doi: 10.1111/j.1469-0691.2011.03497.x.

Levin, D., Raab, N., Pinto, Y., Rothschild, D., Zanir, G., Godneva, A., et al. (2021). Diversity and functional landscapes in the microbiota of animals in the wild. *Science* 372, eabb5352. doi: 10.1126/science.abb5352.

Li, M., Havelaar, A. H., Hoffmann, S., Hald, T., Kirk, M. D., Torgerson, P. R., et al. (2019a). Global disease burden of pathogens in animal source foods, 2010. *PLOS ONE* 14, e0216545. doi: 10.1371/journal.pone.0216545.

Li, S., He, Y., Mann, D. A., and Deng, X. (2021). Global spread of Salmonella Enteritidis via centralized sourcing and international trade of poultry breeding stocks. *Nat. Commun.* 12, 5109. doi: 10.1038/s41467-021-25319-7.

Li, Y., Tang, M., Wang, G., Li, C., Chen, W., Luo, Y., et al. (2019b). Genomic characterization of Kerstersia gyiorum SWMUKG01, an isolate from a patient with respiratory infection in China. *PLOS ONE* 14, e0214686. doi: 10.1371/journal.pone.0214686.

Liu, G., Tang, C. M., and Exley, R. M. (2015). Non-pathogenic Neisseria: members of an abundant, multi-habitat, diverse genus. *Microbiology* 161, 1297–1312. doi: 10.1099/mic.0.000086.

Lord, E., Collins, C., deFrance, S., LeFebvre, M. J., Pigière, F., Eeckhout, P., et al. (2020). Ancient DNA of Guinea Pigs (Cavia spp.) Indicates a Probable New Center of Domestication and Pathways of Global Distribution. *Sci. Rep.* 10, 8901. doi: 10.1038/s41598-020-65784-6.

Madge, D. S. (1969). Effect of antibiotics on intestinal absorption in guinea-pigs. *Comp. Biochem. Physiol.* 30, 295–307. doi: 10.1016/0010-406X(69)90812-3.

Mageiros, L., Méric, G., Bayliss, S. C., Pensar, J., Pascoe, B., Mourkas, E., et al. (2021). Genome evolution and the emergence of pathogenicity in avian Escherichia coli. *Nat. Commun.* 12, 765. doi: 10.1038/s41467-021-20988-w.

Maiden, M. C. J., and Harrison, O. B. (2016). Population and Functional Genomics of Neisseria Revealed with Gene-by-Gene Approaches. *J. Clin. Microbiol.* 54, 1949–1955. doi: 10.1128/JCM.00301-16.

Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., et al. (2010). The global burden of nontyphoidal Salmonella gastroenteritis. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 50, 882–889. doi: 10.1086/650733.

Marcelo M, G., Rosadio A, R., Chero O, A., Díaz O, G., Ciprian C, A., and Maturrano H, L. (2017). Identificación de Salmonella Enteritidis y Salmonella Typhimurium en cuyes mediante la técnica de PCR múltiple. *Rev. Investig. Vet. Perú* 28, 411–417. doi: 10.15381/rivep.v28i2.13074.

Marri, P. R., Paniscus, M., Weyand, N. J., Rendón, M. A., Calton, C. M., Hernández, D. R., et al. (2010). Genome sequencing reveals widespread virulence gene exchange among human Neisseria species. *PloS One* 5, e11835. doi: 10.1371/journal.pone.0011835.

Matsuura S., A., Morales C., S., Calle E., S., and Ara G., M. (2010). Susceptibilidad a antibacterianos in vitro de Salmonella enterica aislada de cuyes de crianza familiar-comercial en la provincia de Carhuaz, Áncash. *Rev. Investig. Vet. Perú* 21, 93–99. Available at: http://www.scielo.org.pe/scielo.php?script=sci_abstract&pid=S1609-91172010000100014&lng=es&nrm=iso&tlng=es [Accessed May 9, 2022].

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci.* 110, 3229–3236. doi: 10.1073/pnas.1218525110.

McInerney, J. O., McNally, A., and O'Connell, M. J. (2017). Why prokaryotes have pangenomes. *Nat. Microbiol.* 2, 1–5. doi: 10.1038/nmicrobiol.2017.40.

McInerney, J. O., Whelan, F. J., Domingo-Sananes, M. R., McNally, A., and O'Connell, M. J. (2020). "Pangenomes and Selection: The Public Goods Hypothesis," in *The Pangenome: Diversity, Dynamics and Evolution of Genomes*, eds. H. Tettelin and D.

Medini (Cham: Springer International Publishing), 151–167. doi: 10.1007/978-3-030-38281-0_7.
Mejía, D., Salvatierra, G., Maximiliano, J., Rímac, R., Carhuaricra, D., Almeyda, M., et al. (2019). Expresión de citoquinas Th1 (IL-2, IL-12, IFN-γ, TNF-α), Th2 (IL-4, IL-10, TGF-β) y Th17 (IL-17) en linfocitos circulantes de cuyes inoculados con una cepa de campo de Salmonella Typhimurium. *Rev. Investig. Vet. Perú* 30, 1750–1761. doi: 10.15381/rivep.v30i4.17188.
Merlino, J., Gray, T., Beresford, R., Baskar, S. R., Gottlieb, T., and Birdsall, J. (2021). Wound infection caused by Neisseria zoodegmatis, a zoonotic pathogen: a case report. *Access Microbiol.* 3, 000196. doi: 10.1099/acmi.0.000196.
Micoli, F., Rondini, S., Alfini, R., Lanzilao, L., Necchi, F., Negrea, A., et al. (2018). Comparative immunogenicity and efficacy of equivalent outer membrane vesicle and glycoconjugate vaccines against nontyphoidal Salmonella. *Proc. Natl. Acad. Sci.* 115, 10428–10433. doi: 10.1073/pnas.1807655115.
Miller, E. A., Elnekave, E., Flores-Figueroa, C., Johnson, A., Kearney, A., Munoz-Aguayo, J., et al. (2020). Emergence of a Novel Salmonella enterica Serotype Reading Clonal Group Is Linked to Its Expansion in Commercial Turkey Production, Resulting in Unanticipated Human Illness in North America. *mSphere* 5, e00056-20. doi: 10.1128/mSphere.00056-20.
Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015.
Moore, E. R. B., Mihaylova, S. A., Vandamme, P., Krichevsky, M. I., and Dijkshoorn, L. (2010). Microbial systematics and taxonomy: relevance for a microbial commons. *Res. Microbiol.* 161, 430–438. doi: 10.1016/j.resmic.2010.05.007.
Morales, E. (1994). The Guinea Pig in the Andean Economy: From Household Animal to Market Commodity. *Lat. Am. Res. Rev.* 29, 129–142. doi: 10.1017/S0023879100035561.
Mottawea, W., Duceppe, M.-O., Dupras, A. A., Usongo, V., Jeukens, J., Freschi, L., et al. (2018). Salmonella enterica Prophage Sequence Profiles Reflect Genome Diversity and Can Be Used for High Discrimination Subtyping. *Front. Microbiol.* 9. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2018.00836 [Accessed June 3, 2022].
Mourkas, E., Taylor, A. J., Méric, G., Bayliss, S. C., Pascoe, B., Mageiros, L., et al. (2020). Agricultural intensification and the evolution of host specialism in the enteric pathogen Campylobacter jejuni. *Proc. Natl. Acad. Sci.* 117, 11018–11028. doi: 10.1073/pnas.1917168117.
Moya A. L. (2019). Prevalencia de salmonelosis en cuyes (Cavia porcellus) procedentes de granjas del centro poblado Huancaquito Alto - Virú - La Libertad. *Univ. Priv. Antenor Orrego* Thesis. Available at: https://repositorio.upao.edu.pe/handle/20.500.12759/5577 [Accessed August 6, 2022].
Muñoz-García, C. I., Sánchez-Montes, S., Villanueva-García, C., Romero-Callejas, E., Díaz-López, H. M., Gordillo-Chávez, E. J., et al. (2019). The role of sloths and anteaters as Leishmania spp. reservoirs: a review and a newly described natural infection of Leishmania mexicana in the northern anteater. *Parasitol. Res.* 118, 1095–1101. doi: 10.1007/s00436-019-06253-6.
Mushegian, A. R., and Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10268–10273. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC38373/ [Accessed April 23, 2023].
Mustapha, M. M., Lemos, A. P. S., Griffith, M. P., Evans, D. R., Marx, R., Coltro, E. S. F., et al. (2020). Two Cases of Newly Characterized Neisseria Species, Brazil. *Emerg. Infect. Dis.* 26, 366–369. doi: 10.3201/eid2602.190191.

Néron, B., Littner, E., Haudiquet, M., Perrin, A., Cury, J., and Rocha, E. P. C. (2022). IntegronFinder 2.0: Identification and Analysis of Integrons across Bacteria, with a Focus on Antibiotic Resistance in Klebsiella. *Microorganisms* 10, 700. doi: 10.3390/microorganisms10040700.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300.

Ogawa, Y., Lee, S.-T., Kasahara, K., Koizumi, A., Chihara, Y., Nakano, R., et al. (2016). A first case of isolation of Kerstersia gyiorum from urinary tract. *J. Infect. Chemother. Off. J. Jpn. Soc. Chemother.* 22, 265–267. doi: 10.1016/j.jiac.2015.11.003.

Okoro, C. K., Barquist, L., Connor, T. R., Harris, S. R., Clare, S., Stevens, M. P., et al. (2015). Signatures of Adaptation in Human Invasive Salmonella Typhimurium ST313 Populations from Sub-Saharan Africa. *PLoS Negl. Trop. Dis.* 9, e0003611. doi: 10.1371/journal.pntd.0003611.

Ortiz G., T., L., J. R. L., Aybar E., M., Angulo-Tisoc, J., Ramos D., D., Ortiz G., T., et al. (2022). Determinación del periodo de retiro de enrofloxacina en cuyes (Cavia porcellus). *Rev. Investig. Vet. Perú* 33. doi: 10.15381/rivep.v33i4.23339.

Özdemir, C., Akçelik, N., Neslihan Özdemir, F., Evcili, İ., Kahraman, T., Gürsel, İ., et al. (2021). The role of bcsE gene in the pathogenicity of Salmonella. *Pathog. Dis.* 79, ftab037. doi: 10.1093/femspd/ftab037.

Parker, C. T., Huynh, S., Alexander, A., Oliver, A. S., and Cooper, K. K. (2021). Genomic Characterization of Salmonella typhimurium DT104 Strains Associated with Cattle and Beef Products. *Pathogens* 10, 529. doi: 10.3390/pathogens10050529.

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. doi: 10.1101/gr.186072.114.

Pence, M. A., Sharon, J., McElvania Tekippe, E., Pakalniskis, B. L., Ford, B. A., and Burnham, C.-A. D. (2013). Two Cases of Kerstersia gyiorum Isolated from Sites of Chronic Infection. *J. Clin. Microbiol.* 51, 2001–2004. doi: 10.1128/JCM.00829-13.

Petrovska, L., Mather, A. E., AbuOun, M., Branchu, P., Harris, S. R., Connor, T., et al. (2016). Microevolution of Monophasic Salmonella Typhimurium during Epidemic, United Kingdom, 2005–2010. *Emerg. Infect. Dis.* 22, 617–624. doi: 10.3201/eid2204.150531.

Pires, B. M. F. B., de Oliveira, B. G. R. B., de Oliveira, F. P., de Almeida Silva, K. de C. F., Del Peloso, P. F., Ferreira-Carvalho, B. T., et al. (2020). Ciprofloxacin-Resistant Kerstersia gyiorum Isolated From a Chronic Wound in Brazil: A Case Report. *Wound Manag. Prev.* 66, 42–45.

Poole, K. (2000). Efflux-Mediated Resistance to Fluoroquinolones in Gram-Negative Bacteria. *Antimicrob. Agents Chemother.* 44, 2233–2241. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC90051/ [Accessed September 18, 2023].

Pulford, C. V., Perez-Sepulveda, B. M., Canals, R., Bevington, J. A., Bengtsson, R. J., Wenner, N., et al. (2021). Stepwise evolution of Salmonella Typhimurium ST313 causing bloodstream infection in Africa. *Nat. Microbiol.* 6, 327–338. doi: 10.1038/s41564-020-00836-1.

Quino, W., Hurtado, C. V., Meza, A. M., Zamudio, M. L., Gavilan, R. G., Quino, W., et al. (2020). Patrones de resistencia a los antimicrobianos en serovares de Salmonella enterica en Perú, 2012-2015. *Rev. Chil. Infectol.* 37, 395–401. doi: 10.4067/S0716-10182020000400395.

Rabsch, W., Andrews, H. L., Kingsley, R. A., Prager, R., Tschäpe, H., Adams, L. G., et al. (2002). Salmonella enterica serotype Typhimurium and its host-adapted variants. *Infect. Immun.* 70, 2249–2255. doi: 10.1128/IAI.70.5.2249-2255.2002.

Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2, vew007. doi: 10.1093/ve/vew007.

Razavi, M., Kristiansson, E., Flach, C.-F., and Larsson, D. G. J. (2020). The Association between Insertion Sequences and Antibiotic Resistance Genes. *mSphere* 5, e00418-20. doi: 10.1128/mSphere.00418-20.

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. doi: 10.1038/s41559-018-0617-0.

Richter, A. M., Povolotsky, T. L., Wieler, L. H., and Hengge, R. (2014). Cyclic-di-GMP signalling and biofilm-related properties of the Shiga toxin-producing 2011 German outbreak Escherichia coli O104:H4. *EMBO Mol. Med.* 6, 1622–1637. doi: 10.15252/emmm.201404309.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* 106, 19126–19131. doi: 10.1073/pnas.0906412106.

Roach, M. J., McNair, K., Giles, S. K., Inglis, L., Pargin, E., Roux, S., et al. (2021). Philympics 2021: Prophage Predictions Perplex Programs. 2021.06.03.446868. doi: 10.1101/2021.06.03.446868.

Roberts, I. S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* 50, 285–315. doi: 10.1146/annurev.micro.50.1.285.

Robertson, S., Burakoff, A., Stevenson, L., Tompkins, B., Patel, K., Tolar, B., et al. (2018). Notes from the Field: Recurrence of a Multistate Outbreak of Salmonella Enteritidis Infections Linked to Contact with Guinea Pigs — Eight States, 2015–2017. *Morb. Mortal. Wkly. Rep.* 67, 1195–1196. doi: 10.15585/mmwr.mm6742a6.

Rocha, E. P. C., and Bikard, D. (2022). Microbial defenses against mobile genetic elements and viruses: Who defends whom from what? *PLOS Biol.* 20, e3001514. doi: 10.1371/journal.pbio.3001514.

Rodrigues, G. L., Panzenhagen, P., Ferrari, R. G., Dos Santos, A., Paschoalin, V. M. F., and Conte-Junior, C. A. (2020). Frequency of Antimicrobial Resistance Genes in Salmonella From Brazil by in silico Whole-Genome Sequencing Analysis: An Overview of the Last Four Decades. *Front. Microbiol.* 11, 1864. doi: 10.3389/fmicb.2020.01864.

Rouli, L., Merhej, V., Fournier, P.-E., and Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 7, 72–85. doi: 10.1016/j.nmni.2015.06.005.

Ruan, Z., Yu, Y., and Feng, Y. (2020). The global dissemination of bacterial infections necessitates the study of reverse genomic epidemiology. *Brief. Bioinform.* 21, 741–750. doi: 10.1093/bib/bbz010.

Salipante, S. J., and Hall, B. G. (2003). Determining the limits of the evolutionary potential of an antibiotic resistance gene. *Mol. Biol. Evol.* 20, 653–659. doi: 10.1093/molbev/msg074.

Salvatierra R, G., Rimac B, R., Chero O, A., Reyna W, I., Rosadio A, R., and Maturrano H, L. (2018). Resistencia antimicrobiana y genotipificación de cepas de Salmonella Typhimurium aisladas de cuyes (Cavia porcellus) provenientes de granjas de producción intensiva de la ciudad de Lima, Perú. *Rev. Investig. Vet. Perú* 29, 319–327. doi: 10.15381/rivep.v29i1.14089.

Sánchez-Macías, D., Barba-Maggi, L., Morales-delaNuez, A., and Palmay-Paredes, J. (2018). Guinea pig for meat production: A systematic review of factors affecting the production, carcass and meat quality. *Meat Sci.* 143, 165–176. doi: 10.1016/j.meatsci.2018.05.004.

Sant'Ana, F. J. F. de, Batista, J. D. S., Blume, G. R., Sonne, L., and Barros, C. S. L. de (2020). Fatal disseminated toxoplasmosis in a brown-throated sloth (Bradypus variegatus) from Northern Brazil - Case report. *Acta Vet. Hung.* 68, 285–288. doi: 10.1556/004.2020.00049.

Schwartzman, J. A., Lebreton, F., Salamzade, R., Martin, M. J., Schaufler, K., Urhan, A., Abeel, T., Camargo, I. L. B. C., Sgardioli, B. F., Prichula, J., Frazzon, A. P. G., Van Tyne, D., Treinish, G., Innis, C. J., Wagenaar, J. A., Whipple, R. M., Manson, A. L.,

Earl, A. M., & Gilmore, M. S. (2023). Global diversity of enterococci and description of 18 novel species. bioRxiv : the preprint server for biology, 2023.05.18.540996. https://doi.org/10.1101/2023.05.18.540996
Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153.
Seib, K. L., Zhao, X., and Rappuoli, R. (2012). Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clin. Microbiol. Infect.* 18, 109–116. doi: 10.1111/j.1469-0691.2012.03939.x.
Shen, W., Chen, Y., Wang, N., Wan, P., Peng, Z., Zhao, H., et al. (2022). Seasonal variability of the correlation network of antibiotics, antibiotic resistance determinants, and bacteria in a wastewater treatment plant and receiving water. *J. Environ. Manage.* 317, 115362. doi: 10.1016/j.jenvman.2022.115362.
Sheppard, S. K., Guttman, D. S., and Fitzgerald, J. R. (2018). Population genomics of bacterial host adaptation. *Nat. Rev. Genet.* 19, 549–565. doi: 10.1038/s41576-018-0032-z.
Sherry, N. L., Horan, K. A., Ballard, S. A., Gonçalves da Silva, A., Gorrie, C. L., Schultz, M. B., et al. (2023). An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. *Nat. Commun.* 14, 60. doi: 10.1038/s41467-022-35713-4.
Sneath, P. H., and Barrett, S. J. (1996). A new species of Neisseria from the dental plaque of the domestic cow, Neisseria dentiae sp. nov. *Lett. Appl. Microbiol.* 23, 355–358. doi: 10.1111/j.1472-765x.1996.tb00207.x.
Solano, C., García, B., Valle, J., Berasain, C., Ghigo, J.-M., Gamazo, C., et al. (2002). Genetic analysis of Salmonella enteritidis biofilm formation: critical role of cellulose. *Mol. Microbiol.* 43, 793–808. doi: 10.1046/j.1365-2958.2002.02802.x.
Song, C., Wang, B., Tan, J., Zhu, L., and Lou, D. (2017). Discovery of tauroursodeoxycholic acid biotransformation enzymes from the gut microbiome of black bears using metagenomics. *Sci. Rep.* 7, 45495. doi: 10.1038/srep45495.
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, e00356-13. doi: 10.1128/mBio.00356-13.
STACKEBRANDT, E., and GOEBEL, B. M. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *Int. J. Syst. Evol. Microbiol.* 44, 846–849. doi: 10.1099/00207713-44-4-846.
Stevens, M. P., Humphrey, T. J., and Maskell, D. J. (2009). Molecular insights into farm animal and zoonotic Salmonella infections. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 2709–2723. doi: 10.1098/rstb.2009.0094.
Stevens, M. P., and Kingsley, R. A. (2021a). Salmonella pathogenesis and host-adaptation in farmed animals. *Curr. Opin. Microbiol.* 63, 52–58. doi: 10.1016/j.mib.2021.05.013.
Stevens, M. P., and Kingsley, R. A. (2021b). Salmonella pathogenesis and host-adaptation in farmed animals. *Curr. Opin. Microbiol.* 63, 52–58. doi: 10.1016/j.mib.2021.05.013.
Stewart, R. D., Auffret, M. D., Warr, A., Walker, A. W., Roehe, R., and Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* 37, 953–961. doi: 10.1038/s41587-019-0202-3.
Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039.
Suminda, G. G. D., Bhandari, S., Won, Y., Goutam, U., Kanth Pulicherla, K., Son, Y.-O., et al. (2022). High-throughput sequencing technologies in the detection of livestock pathogens, diagnosis, and zoonotic surveillance. *Comput. Struct. Biotechnol. J.* 20, 5378–5392. doi: 10.1016/j.csbj.2022.09.028.
Sun, H., Wan, Y., Du, P., and Bai, L. (2020). The Epidemiology of Monophasic Salmonella Typhimurium. *Foodborne Pathog. Dis.* 17, 87–97. doi: 10.1089/fpd.2019.2676.

Syberg-Olsen, M. J., Garber, A. I., Keeling, P. J., McCutcheon, J. P., and Husnik, F. (2022). Pseudofinder: Detection of Pseudogenes in Prokaryotic Genomes. *Mol. Biol. Evol.* 39, msac153. doi: 10.1093/molbev/msac153.

Tassinari, E., Bawn, M., Thilliez, G., Charity, O., Acton, L., Kirkwood, M., et al. (2020). Whole-genome epidemiology links phage-mediated acquisition of a virulence gene to the clonal expansion of a pandemic Salmonella enterica serovar Typhimurium clone. *Microb. Genomics* 6, mgen000456. doi: 10.1099/mgen.0.000456.

Tesson, F., Hervé, A., Mordret, E., Touchon, M., d'Humières, C., Cury, J., et al. (2022). Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* 13, 2561. doi: 10.1038/s41467-022-30269-9.

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006.

The CRyPTIC Consortium (2022). Genome-wide association studies of global Mycobacterium tuberculosis resistance to 13 antimicrobials in 10,228 genomes identify new resistance mechanisms. *PLOS Biol.* 20, e3001755. doi: 10.1371/journal.pbio.3001755.

Thorpe, H. A., Booton, R., Kallonen, T., Gibbon, M. J., Couto, N., Passet, V., et al. (2022). A large-scale genomic snapshot of Klebsiella spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings. *Nat. Microbiol.* 7, 2054–2067. doi: 10.1038/s41564-022-01263-0.

Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W., and Corander, J. (2018). RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res.* 3, 93. doi: 10.12688/wellcomeopenres.14694.1.

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21, 180. doi: 10.1186/s13059-020-02090-4.

Unkmeir, A., Kämmerer, U., Stade, A., Hübner, C., Haller, S., Kolb-Mäurer, A., et al. (2002). Lipooligosaccharide and Polysaccharide Capsule: Virulence Factors of Neisseria meningitidis That Determine Meningococcal Interaction with Human Dendritic Cells. *Infect. Immun.* 70, 2454–2462. doi: 10.1128/IAI.70.5.2454-2462.2002.

van der Bij, A. K., and Pitout, J. D. D. (2012). The role of international travel in the worldwide spread of multiresistant Enterobacteriaceae. *J. Antimicrob. Chemother.* 67, 2090–2100. doi: 10.1093/jac/dks214.

Van Puyvelde, S., Gasperini, G., Biggel, M., Phoba, M.-F., Raso, M. M., de Block, T., et al. (2022). Genetic and Structural Variation in the O-Antigen of Salmonella enterica Serovar Typhimurium Isolates Causing Bloodstream Infections in the Democratic Republic of the Congo. *mBio* 13, e00374-22. doi: 10.1128/mbio.00374-22.

Vandamme, P., De Brandt, E., Houf, K., and De Baere, T. (2012). Kerstersia similis sp. nov., isolated from human clinical samples. *Int. J. Syst. Evol. Microbiol.* 62, 2156–2159. doi: 10.1099/ijs.0.037887-0.

Viana, D., Comos, M., McAdam, P. R., Ward, M. J., Selva, L., Guinane, C. M., et al. (2015). A single natural nucleotide mutation alters bacterial pathogen host-tropism. *Nat. Genet.* 47, 361–366. doi: 10.1038/ng.3219.

Volokhov, D. V., Amselle, M., Bodeis-Jones, S., Delmonte, P., Zhang, S., Davidson, M. K., et al. (2018). Neisseria zalophi sp. nov., isolated from oral cavity of California sea lions (Zalophus californianus). *Arch. Microbiol.* 200, 819–828. doi: 10.1007/s00203-018-1499-x.

Wan, X., McLaughlin, R. W., Zhou, J., Hao, Y., Zheng, J., and Wang, D. (2016). Isolation of culturable aerobic bacteria and evidence of Kerstersia gyiorum from the blowhole of captive Yangtze finless porpoises. *Antonie Van Leeuwenhoek* 109, 1167–1175. doi: 10.1007/s10482-016-0713-6.

Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W., and Zheng, S.-S. (2015). Application of metagenomics in the human gut microbiome. *World J. Gastroenterol. WJG* 21, 803–814. doi: 10.3748/wjg.v21.i3.803.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., et al. (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int. J. Syst. Evol. Microbiol.* 37, 463–464. doi: 10.1099/00207713-37-4-463.

Welch, R. A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc. Natl. Acad. Sci.* 99, 17020–17024. doi: 10.1073/pnas.252529799.

Weyand, N. J., Ma, M., Phifer-Rixey, M., Taku, N. A., Rendón, M. A., Hockenberry, A. M., et al. (2016). Isolation and characterization of Neisseria musculi sp. nov., from the wild house mouse. *Int. J. Syst. Evol. Microbiol.* 66, 3585–3593. doi: 10.1099/ijsem.0.001237.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. 1st ed. Springer New York, NY Available at: https://doi.org/10.1007/978-0-387-98141-3 [Accessed August 6, 2022].

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330.

Yen, S., and Johnson, J. S. (2021). Metagenomics: a path to understanding the gut microbiome. *Mamm. Genome* 32, 282–296. doi: 10.1007/s00335-021-09889-x.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628.

Yue, M., Han, X., Masi, L. D., Zhu, C., Ma, X., Zhang, J., et al. (2015). Allelic variation contributes to bacterial host specificity. *Nat. Commun.* 6, 8754. doi: 10.1038/ncomms9754.

Zhang, G., Yang, J., Lai, X.-H., Jin, D., Pu, J., Bai, X., et al. (2019a). Neisseria weixii sp. nov., isolated from rectal contents of Tibetan Plateau pika (Ochotona curzoniae). *Int. J. Syst. Evol. Microbiol.* 69, 2305–2311. doi: 10.1099/ijsem.0.003466.

Zhang, G., Yang, J., Lai, X.-H., Lu, S., Jin, D., Pu, J., et al. (2019b). Neisseria chenwenguii sp. nov. isolated from the rectal contents of a plateau pika (Ochotona curzoniae). *Antonie Van Leeuwenhoek* 112, 1001–1010. doi: 10.1007/s10482-019-01234-2.

Zhang, S., den Bakker, H. C., Li, S., Chen, J., Dinsmore, B. A., Lane, C., et al. (2019c). SeqSero2: Rapid and Improved Salmonella Serotype Determination Using Whole-Genome Sequencing Data. *Appl. Environ. Microbiol.* 85, e01746-19. doi: 10.1128/AEM.01746-19.