

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

Instituto Aggeu Magalhães

FUNDAÇÃO OSWALDO CRUZ
INSTITUTO AGGEU MAGALHÃES
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOCÊNCIAS E BIOTECNOLOGIA EM
SAÚDE

João Luiz de Lemos Padilha Pitta

Estudo genômico de isolados de *Yersinia pestis* no Brasil

Recife

2023

João Luiz de Lemos Padilha Pitta

Estudo genômico de isolados de *Yersinia pestis* no Brasil

Tese de Doutorado em Biociências e Biotecnologia em Saúde apresentada ao Programa de Pós-graduação em Biociências e Biotecnologia em Saúde do Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, como requisito parcial para obtenção do grau de Doutor em Ciências.

Orientador: Dr. Antonio Mauro Rezende

Coorientador: Dra. Alzira Maria Paiva de Almeida

Recife

2023

Título do trabalho em inglês: Genomic study of Yersinia pestis isolates in Brazil.

P688e Pitta, João Luiz de Lemos Padilha.
Estudo genômico de isolados de Yersinia pestis do Brasil / João Luiz de Lemos Padilha Pitta. -- 2023.
100 p. : il.color.

Orientador: Antonio Mauro Rezende.
Coorientadora: Alzira Maria Paiva de Almeida.
Tese (Doutorado em Programa de Pós-graduação em Biociências e Biotecnologia em Saúde) - Fundação Oswaldo Cruz, Instituto Aggeu Magalhães, Recife, 2023.
Bibliografia: f. 76-86.

1. Yersinia pestis. 2. Peste. 3. Genômica. 4. Fatores de virulência.
I. Título.

CDU 616.98

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da Rede de Bibliotecas da Fiocruz com os dados fornecidos pelo(a) autor(a).

Bibliotecário responsável pela elaboração da ficha catalográfica: Adagilson Batista Bispo da Silva - CRB-1239
Biblioteca Luciana Borges Abrantes dos Santos

João Luiz de Lemos Padilha Pitta

Estudo genômico de isolados de *Yersinia pestis* no Brasil

Tese de Doutorado em Biociências e Biotecnologia em Saúde apresentada ao Programa de Pós-graduação em Biociências e Biotecnologia em Saúde do Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, como requisito para obtenção do grau de Doutor em Ciências.

Aprovado em: 22 de setembro de 2023.

BANCA EXAMINADORA

Dr. Antonio Mauro Rezende (Orientador)
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz

Dr. Christian Robson de Souza Reis (Titular Interno)
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz

Dr. Luiz Carlos Alves (Titular Interno)
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz

Dr. Túlio de Lima Campos (Titular Externo)
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz

Dra. Maria Paloma Silva Barros (Titular Externo)
Centro de Tecnologias Estratégicas do Nordeste

AGRADECIMENTOS

Muitas pessoas contribuíram das mais diversas formas para a realização deste trabalho, mas em especial faço os seguintes agradecimentos:

Aos meus orientadores, doutores Antonio Rezende e Alzira Almeida, por aceitarem o desafio de orientar um aluno vindo de outra área do conhecimento e pela enorme disponibilidade em me auxiliar ao longo de todo o processo do doutorado. A caminhada até a defesa foi certamente mais leve e repleta de aprendizado pois sempre pude contar com o conhecimento e responsabilidade dos meus orientadores.

A Beatriz Toscano, minha colega, amiga, namorada e maior sorte que a vida poderia me presentear. Esteve comigo desde meus primeiros passos na Fiocruz, antes mesmo que eu estivesse oficialmente na pós-graduação, gentilmente me cedendo material de estudo para a seleção do mestrado, auxiliando nas muitas dúvidas que tive nas disciplinas, sendo o ombro sempre disponível nos meus tombos, revisando dissertação, artigos, tese... Suas contribuições vão muito além do meu mestrado/doutorado e extrapolam para minha vida de uma forma que não cabe em palavras. Obrigado por fazer minha vida em todos os aspectos mais leve.

A Crhisllane Vasconcelos, a irmã que escolhi, por tudo que estudamos, discutimos e aprendemos juntos na vida que foi além dos muros do Aggeu Magalhães. Sou grato por seu apoio e presença incondicionais no meu mestrado e início do doutorado, o tanto que me ensinou, os tantos scripts que me ajudou a fazer ou refazer, as longas discussões dos meus projetos e por ser a melhor vizinha que já tive, mesmo que por pouco tempo.

Aldrey Galindo, pelas inúmeras contribuições revisando comigo por muitas madrugadas meus scripts. Quando os prazos apertaram e as coisas não estavam funcionando, pude contar com o vasto conhecimento técnico, paciência e boa conversa desse amigo e profissional extremamente qualificado.

Aos meus amigos da Bioinformática Filipe Zimmer, Alexandre Silva e Elverson Melo por sempre estarem dispostos a me ajudar na troca de informação de ferramentas de bioinformática, muitas vezes dando um norte para alguns problemas ao longo do doutorado.

A Matheus Bezerra, do Serviço de Referência Nacional em Peste da Fiocruz PE, pelas incontáveis contribuições científicas no artigo resultante deste projeto e o compartilhamento de conhecimento em peste.

Aos membros da banca examinadora que gentilmente aceitaram o convite para contribuir neste trabalho.

*“Os flagelos, na verdade, são uma coisa comum,
mas é difícil acreditar neles quando se abatem
sobre nós. Houve no mundo igual número de
pestes e de guerras. E, contudo, as pestes, assim
como as guerras, encontram sempre as pessoas
igualmente desprevenidas”*

Albert Camus (1947)

RESUMO

PITTA, João Luiz de Lemos Padilha. Estudo genômico de isolados de *Yersinia pestis* no Brasil. 2023. Tese (Doutorado em Biociências e Biotecnologia em Saúde) – Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2023.

Yersinia pestis, o agente etiológico da peste, tem uma história evolutiva recente e é considerada uma espécie geneticamente homogênea. A doença foi responsável por milhões de mortes ao longo de três pandemias e continua vitimando pessoas, principalmente na África e Américas. Atualmente o Brasil vive um período de silêncio epidemiológico, porém, anticorpos contra esta bactéria ainda são detectados em animais sentinelas, sugerindo atividade da infecção no ciclo silvestre. O presente estudo utilizou abordagens *in silico* para analisar 406 genomas de *Y. pestis* isolados no Brasil entre 1966 e 1997, pertencentes ao Serviço de Referência Nacional em Peste da Fiocruz/PE. Foram realizadas três análises principais, sendo estas análises de SNV, pangenoma e cgMLST, onde os resultados foram cruzados com os dados epidemiológicos. A análise do pangenoma agrupou as cepas em clados, baseando-se na presença ou ausência de genes. Cepas de quatro clados principais (C, E, G e H) foram destacadas pela ausência de genes, com ênfase nos genes associados a importantes fatores de virulência de *Y. pestis*. Destaca-se, entre esses genes ausentes, os associados a formação de biofilme, adesão celular, aquisição e transporte de ferro e ao sistema de secreção tipo 6 (T6SS). As cepas do clado G apresentaram o maior número de genes ausentes. O clado E apresentou significativa ausência de genes relacionados ao sistema de secreção T6SS, um sistema pouco estudado para patogenicidade de *Y. pestis*, enquanto o clado H demonstrou, predominantemente, a ausência de genes relacionados a plasmídeos. Juntos, esses resultados sugerem uma atenuação da virulência ao longo do tempo nessas cepas. Não foram encontrados genes associados à resistência aos antimicrobianos nos genomas das cepas analisadas. A análise cgMLST associou dados genômicos e epidemiológicos, destacando padrões evolutivos relacionados aos anos de isolamento e surtos de *Y. pestis* no Brasil. Três agrupamentos principais se destacam por terem cepas recuperadas de dois períodos distintos na Chapada do Araripe, sendo um com cepas do fim dos anos 60 até 1972 e outro que aparece após o breve período de silêncio epidemiológico a partir de 1973/74. Analisando esses agrupamentos de cepas por surtos, foi possível destacar um grupo de cepas do surto do município de Triunfo entre os anos de 1978 e 1979. Um outro agrupamento em evidência foi o das cepas na Chapada da Borborema, com todas as cepas isoladas do estado da Paraíba no período entre 1979 e 1982. A análise de SNV mostrou que as cepas desses surtos também tiveram acúmulo de variações acima da média para o conjunto de cepas analisadas, corroborando com a análise de cgMLST. Assim, com os resultados obtidos foi possível cruzar os dados genômicos e epidemiológicos, contribuindo para a compreensão da microevolução genética e da virulência em *Y. pestis* no Brasil, ressaltando a importância da utilização de dados genômicos em investigações epidemiológicas.

Palavras-chave: *Yersinia pestis*; peste; genômica comparativa; fatores de virulência

ABSTRACT

PITTA, João Luiz de Lemos Padilha. Genomic study of *Yersinia pestis* isolates in Brazil. 2023. Thesis (Doctorate in Biosciences and Biotechnology in Health) – Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2023.

Yersinia pestis, the etiological agent of plague, has a recent evolutionary history and is considered a genetically homogeneous species. The disease was responsible for millions of deaths throughout three pandemics and continues to victimize people, mainly in Africa and the Americas. Brazil is currently experiencing a period of epidemiological silence, however, antibodies against this bacterium are still detected in sentinel animals, suggesting infection activity in the sylvatic cycle. The present study used *in silico* approaches to analyze 406 genomes of *Y. pestis* isolated in Brazil between 1966 and 1997, belonging to the National Reference Service of Pest of Fiocruz/PE. Three main analyses were performed, these being SNV, pangenome, and cgMLST analyses, in which the results were cross-referenced with epidemiological data. Pangenome analysis grouped strains into clades, based on the presence or absence of genes. Strains from four main clades (C, E, G and H) were highlighted by the absence of genes, with an emphasis on genes associated with important virulence factors of *Y. pestis*. Among these missing genes, those associated with biofilm formation, cell adhesion, iron acquisition and transport and the type 6 secretion system (T6SS) stand out. Clade G strains had the highest number of missing genes. Clade E showed a significant absence of genes related to the T6SS secretion system, a system that has been minimally studied in relation to the pathogenicity of *Y. pestis*, while clade H demonstrated, predominantly, the absence of genes related to plasmids. Together, these results suggest an attenuation of virulence over time in these strains. No genes associated with antimicrobial resistance were found in the genomes of the strains analyzed. The cgMLST analysis associated genomic and epidemiological data, highlighting evolutionary patterns related to the years of *Y. pestis* isolation and outbreaks in Brazil. Three main clusters stand out for having strains recovered from two distinct periods in Chapada do Araripe, one with strains from the late 1960s to 1972 and the other appearing after the brief period of epidemiological silence from 1973/74. Analyzing these strains by outbreaks, it was possible to highlight a group of strains from the outbreak in the municipality of Triunfo between the years 1978 and 1979. Another grouping in evidence was that of the strains in Chapada da Borborema, with all strains isolated from the state of Paraíba in the period between 1979 and 1982. The SNV analysis showed that the strains from these outbreaks also had an accumulation of variations above the average for the set of strains analyzed, corroborating the cgMLST analysis. Thus, with the results obtained, it was possible to cross-reference genomic and epidemiological data, contributing to the understanding of genetic microevolution and virulence in *Y. pestis* in Brazil, highlighting the importance of using genomic data in epidemiological investigations.

Keywords: *Yersinia pestis*; comparative genomics; virulence factors

LISTA DE ILUSTRAÇÕES

Figura 1 - Focos de Peste no Brasil	17
Figura 2 - Distribuição das Variedades Geográficas de <i>Y. pestis</i> segundo Devignat (1951) ..	21
Figura 3 - Esquema de funcionamento do <i>pipeline</i> NASP	41
Figura 4 - Esquema de funcionamento do <i>pipeline</i> Roary	44
Figura 5 - Esquema de funcionamento da ferramenta Chewbbaca	46
Figura 6 - Árvore baseada em SNVs da matriz de bestsnp	51
Figura 7 - Árvore baseada em SNVs da matriz de missingdata	52
Figura 8 - Análise de efeito de SNVs em proteínas	53
Figura 9 - Distribuição entre genes do núcleo e acessórios	54
Figura 10 - Árvore baseada no genoma <i>core</i>	56
Figura 11 - Árvore demonstrando presença e ausência de genes	57
Figura 12 - Árvore filogenética com clados destacados de acordo com presença e ausência de genes	58
Figura 13 - Filogenia baseada na presença ou ausência de gente associada aos dados epidemiológicos	60
Figura 14 - Fatores de virulência agrupados por clado e <i>p-valor</i>	62
Figura 15 - Comparação da distribuição de genes entre Roary e ChewBBACA	63
Figura 16 - Agrupamentos de cepas baseados na análise gene-a-gene, por ano	64
Figura 17 - Agrupamentos de cepas baseados na análise gene-a-gene em relação aos surtos ..	66

LISTA DE ABREVIATURAS E SIGLAS

ANI	Average Nucleotide Identity
BAM	Binary Alignment Map
cgMLST	core genome Multi Locus Sequence Typing
CSV	Comma-Separated Values
DST	Doenças Sexualmente Transmissíveis
GbG	Gene by Gene
GSEA	Gene Set Enrichment Analysis
kb	Kilobase
MLST	Multi Locus Sequence Typing
NGS	Next Generation Sequencing
OMS	Organização Mundial de Saúde
ORF	Open Reading Frame
Pb	Pares de base
PCR	Polymerase Chain Reaction
PHEIC	Public Health Emergency of International Concern
SAM	Sequence Alignment/Map
SNV	Single Nucleotide Variations
SRP	Serviço de Referência de Peste
T3SS	Type 3 Secretion System
WFCC	World Federation Culture Collection
wgMLST	Whole-genome MLST
WGS	Whole genome sequencing
XML	Extensible Markup Language
Yop	<i>Yersinia</i> outer membrane proteins

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	16
2.1	HISTÓRICO DA PESTE NO MUNDO	16
2.2	O GÊNERO <i>YERSINIA</i>	19
2.3	<i>Yersinia pestis</i> : AGENTE ETIOLÓGICO E DIAGNÓSTICO	20
2.4	ASPECTOS CLÍNICOS DA PESTE	22
2.5	TRANSMISSÃO E VETORES DA PESTE	24
2.6	TRATAMENTO DA PESTE	26
2.7	RESISTÊNCIA ANTIMICROBIANA POR <i>Y. pestis</i>	27
2.8	FATORES DE VIRULÊNCIA DA <i>Y. pestis</i>	28
2.9	EVOLUÇÃO GENÉTICA DA <i>Y. pestis</i>	30
2.10	GENÔMICA COMPARATIVA E DIVERSIDADE BACTERIANA	32
3	JUSTIFICATIVA	35
4	PERGUNTA CONDUTORA	36
5	HIPÓTESE	37
6	OBJETIVOS	38
6.1	OBJETIVO GERAL	38
6.2	OBJETIVOS ESPECÍFICOS	38
7	MATERIAIS E MÉTODOS	39
7.1	ISOLADOS DO ESTUDO	39
7.2	ANÁLISE DO SEQUENCIAMENTO DOS ISOLADOS	40
7.3	ABORDAGEM DE MONTAGEM POR REFERÊNCIA E ANÁLISE DE SNVS	40
7.3.1	Pipeline para mapeamento e análise de SNVs	40
7.3.2	Análise das montagens dos genomas de <i>Y. pestis</i>	42
7.3.3	Filogenia baseada em SNVs	42
7.3.4	Análise de efeitos dos SNVs	43
7.4	ABORDAGEM DE MONTAGEM DE NOVO E PANGENÔMICA	43
7.4.1	Montagem de novo e anotação dos genomas	43
7.4.2	Análise pangenômica	44
7.4.3	Análise estatística de presença e ausência dos genes na pangenômica	45
7.5	ANÁLISE GENÔMICA POR CGMLST	45

7.6	ANÁLISE FILOGENÉTICA ENTRE OS ISOLADOS DE <i>Y. pestis</i>	47
7.7	PESQUISA POR GENES DE RESISTÊNCIA A ANTIMICROBIANOS E FATORES DE VIRULÊNCIA	47
7.8	ANÁLISE DE GENES ANOTADOS COMO HIPOTÉTICOS	48
8	CONSIDERAÇÕES ÉTICAS	49
9	RESULTADOS	50
10	DISCUSSÃO	67
11	CONCLUSÃO	75
	REFERÊNCIAS	76
	APÊNDICE A - DADOS DO SEQUENCIAMENTO E MONTAGEM DAS CEPAS DE <i>Y. pestis</i> DA FIOCRUZ/CYP	87
	APÊNDICE B - CEPAS DE BASE DE DADOS PÚBLICA UTILIZADAS NO ESQUEMA DO cgMLST	96
	APÊNDICE C – ARTIGOS CIENTÍFICOS PUBLICADOS	98

1 INTRODUÇÃO

A peste é a infecção que tem como agente causador a bactéria *Yersinia pestis*, um bacilo aeróbico não móvel, Gram-negativo, que já devastou populações humanas ao longo de suas três principais pandemias registradas na era cristã. A primeira pandemia, conhecida como a peste de Justiniano, devastou o mundo civilizado entre 542 e 602 d.C. A segunda pandemia de peste, comumente conhecido como a Peste Negra, começou na Ásia e chegou à Europa e ao norte da África, persistindo do século XIV ao século XVI e tirando a vida de um terço da população europeia neste período. A última pandemia teve início na China em 1894 e foi propagada para outras partes do mundo por meio das navegações, deixando focos que persistem até os dias atuais principalmente na África e nas Américas.

A bactéria *Y. pestis* foi introduzida no Brasil durante a última pandemia, em 1899, por ratos e pulgas presentes em navios comerciais a vapor no porto de Santos, no estado de São Paulo. Devido ao transporte e comércio de mercadorias, a peste se espalhou pelo interior do país por meio de estradas, ferrovias e transporte fluvial, atingindo inicialmente vilas do interior. Em algumas regiões, a peste conseguiu estabelecer novos focos, mantendo um ciclo sustentável em roedores silvestres dos ecossistemas locais, causando muitos surtos em paisagens rurais dentro do bioma da Caatinga. Esses surtos ocorreram em vários estados da região Nordeste do Brasil e no nordeste de Minas Gerais. Além disso, fora da Caatinga, há uma área menor de transmissão da peste no bioma da Mata Atlântica, no Vale do Rio Doce em Minas Gerais e na Serra dos Órgãos no estado do Rio de Janeiro. Devido ao impacto e número de casos de peste no Brasil, o governo brasileiro em parceria com a OMS financiou a partir do ano 1966 o Projeto Piloto de Peste (PPP) no município de Exu - PE, desenvolvendo o programa de pesquisa da Chapada do Araripe até 1974. A coleção de peste foi transferida em 1982 para o Instituto Aggeu Magalhães recebendo novos isolados até 1997, ano em que a última cepa foi coletada nas ações de vigilância.

A peste atinge primordialmente roedores, geralmente transmitida por pulgas infectadas, e ocasionalmente pode infectar outros mamíferos como os seres humanos, sendo estes hospedeiros incidentais quando penetram no ecossistema de circulação da doença. Apesar de ser considerada por muitos como uma doença do passado, dado os longos períodos de quiescência, ela ainda persiste atualmente em diversos hospedeiros, reservatórios e diversos focos silvestres em vários países do mundo, sendo atualmente considerada uma doença reemergente.

A transmissão da peste ocorre mais frequentemente por meio da picada de pulgas infectadas, mas também pode ser transmitida através de escoriações na pele decorrentes da manipulação de animais contaminados, pela ingestão de piolhos, pulgas ou carne de animais e pela inalação de aerossóis produzidos por pacientes ou animais doentes. Devido à sua alta taxa de letalidade e ao seu potencial epidêmico, a peste é classificada pela organização mundial de saúde como “emergência de saúde pública de importância internacional”, sendo no Brasil uma doença de notificação compulsória nacional e que exige vigilância permanente dos focos.

A forma clínica mais comum da doença é a peste bubônica, assim denominada porque os pacientes apresentam nódulos linfático dolorido e inchado, conhecido como bubão, sendo os nódulos femorais e inguinais os mais frequentes. A peste septicêmica é a segunda forma clínica mais comum e a terceira forma, mais rara, é a peste pneumônica. Em todas essas formas clínicas a peste apresenta altas taxas de mortalidade, a menos que o diagnóstico e tratamento com antibióticos seja iniciado precocemente. A alta taxa de letalidade da peste justifica inclusive a possibilidade do uso de seu agente etiológico como arma biológica para ataques de bioterrorismo.

Alterações climáticas e ambientais combinadas à capacidade da *Y. pestis* em sofrer modificações genômicas, podem contribuir para o ressurgimento de surtos. Assim, a peste precisa ser vista como um problema atual e receber atenção contínua, com ações permanentes de monitoramento das áreas focais e esforços para aumentar o entendimento sobre a virulência e os mecanismos de disseminação do seu agente etiológico, visando à detecção precoce da fonte da doença e a determinação das melhores estratégias de controle na ocorrência de um surto.

Devido aos avanços tecnológicos e a consequente redução dos custos, sequenciamento de genoma completo (WGS – *Whole Genome Sequencing*) está cada vez mais comum nos laboratórios. Esse tipo de abordagem gera uma grande quantidade de dados por análise, o que impossibilita muitos estudos de usar os métodos mais tradicionais de laboratório. Para contornar esse problema, vários métodos computacionais estão em contante desenvolvimento e sendo cada dia mais utilizados, possibilitando a análise de dados em larga escala. Além de softwares para análises específicas, a automatização dessas análises em larga escala através de *pipelines* (processos ou softwares encadeados para processar várias etapas do estudo), processamento paralelo, computação em nuvem e outros métodos computacionais estão auxiliando os estudos genômicos nos últimos anos.

Dado o exposto, o presente estudo utilizou abordagens computacionais e analisou os 406 genomas de *Y. pestis* já sequenciados e pertencentes à coleção de culturas de *Y. pestis* do Serviço de Referência Nacional em Peste da FIOCRUZ PE (Fiocruz/CYP), que conta com cepas isoladas de diversas fontes em focos da doença no Nordeste brasileiro entre 1966 e 1997, quando o último caso foi detectado no Brasil.

As informações obtidas da análise do sequenciamento dos genomas das cepas brasileiras de *Y. pestis* permitiram estudar a diversidade temporal, geográfica e genética durante as epidemias e períodos endêmicos no Brasil, evidenciando o perfil de virulência dessas cepas e associando dados genéticos aos dados epidemiológicos.

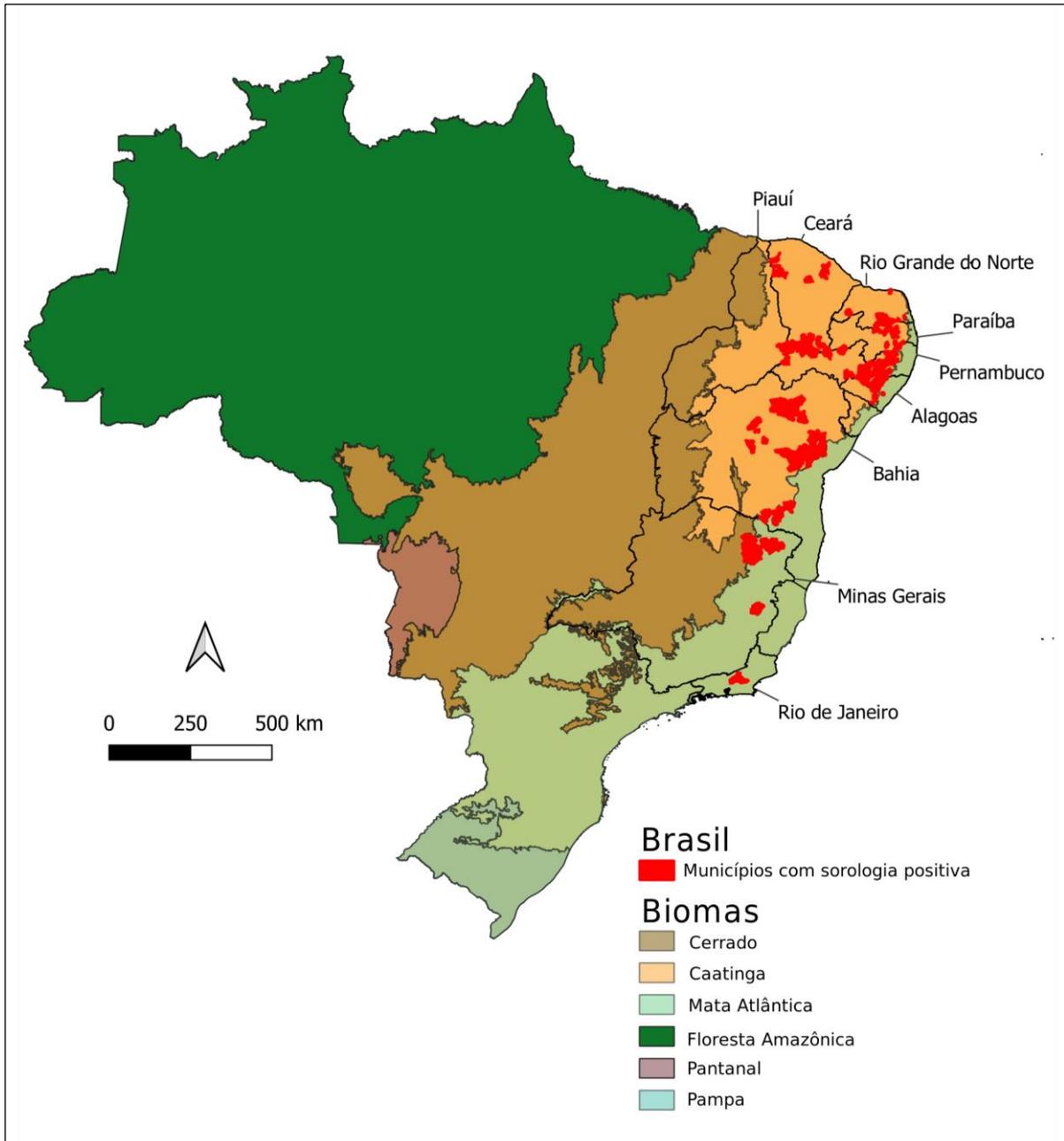
2 REFERENCIAL TEÓRICO

2.1 HISTÓRICO DA PESTE NO MUNDO

A peste é uma doença infecciosa que tem como agente etiológico a bactéria *Yersinia pestis* (Barbieri *et al.*, 2020; Bramanti *et al.*, 2016; Demeure *et al.*, 2019). Esta doença é conhecida desde a antiguidade, constando em registros já no Antigo Testamento, no II Livro de Samuel, provavelmente originada no Planalto Central Asiático. Através dos séculos, diversas pandemias de peste devastaram as populações humanas, com o registro de ao menos três grandes pandemias durante a era Cristã. A primeira pandemia, denominada Peste de Justiniano, devastou o mundo civilizado entre 542 e 602 d.C. Na segunda pandemia, conhecida como Peste Negra, estima-se 25 milhões de vítimas fatais somente na Europa entre os séculos XIV e XVII. Já a terceira pandemia, ou Pandemia Contemporânea, iniciou-se na China em 1894, e de Hong Kong dispersou-se por meio das navegações, deixando vários focos ao redor do mundo que persistem até os dias atuais (Barbieri *et al.*, 2020; Bramanti *et al.*, 2016; Butler, 2013; Demeure *et al.*, 2019; Perry; Fetherston, 1997).

A peste foi introduzida no Brasil no decorrer da última pandemia, em 1899, através do porto de Santos no Estado de São Paulo, trazida em navio por ratos e pulgas (Fernandes *et al.*, 2021; Tavares *et al.*, 2012; Vogler *et al.*, 2019). A partir da introdução no país, a zoonose espalhou-se inicialmente pelos portos do Rio de Janeiro e Fortaleza (1900) e continuou seguindo por outras cidades litorâneas atingindo o estado de Pernambuco e Rio Grande do Sul em 1902, Pará e Maranhão em 1903, Bahia em 1904, Paraná, Espírito Santo e Sergipe em 1906 (Tavares *et al.*, 2012). A partir de então, graças ao transporte e comércio de bens, a peste dispersou-se através de rotas comerciais terrestres, fluviais e ferrovias, instalando-se entre os roedores silvestres de áreas rurais em uma grande área do país que vai da região do Nordeste brasileiro, a partir do estado do Ceará, até a região sudeste no estado de Minas Gerais, além de uma área isolada no estado do Rio de Janeiro (Fernandes *et al.*, 2021; Giles; Peterson; Almeida, 2011; Vogler *et al.*, 2019). Na Figura 1 é possível visualizar os focos brasileiros da doença.

Figura 1 - Focos de Peste no Brasil



Fonte: O autor

Legenda: Distribuição dos casos de peste no Brasil confirmados por cultura bacteriológica e sorologia entre 1966 e 1997. Cada cor representa diferentes biomas brasileiros. Pontos vermelhos no mapa marcam os focos locais onde foram registrados os casos.

Os focos brasileiros atualmente estão localizados nos estados de Alagoas, Bahia, Ceará, Minas Gerais, Paraíba, Pernambuco, Piauí e Rio Grande do Norte. Além disso, existe no estado do Rio de Janeiro um foco isolado na serra dos Órgãos (Tavares et al., 2020; Vogler et al., 2019).

A história da peste no Brasil pode ser dividida em quatro fases. A primeira é a fase portuária, onde ocorreu a introdução da doença em 1899 e sua manutenção nos portos e seus entornos nos primeiros anos do século XX. A segunda fase é a urbana, caracterizada pela interiorização da peste através das ferrovias e das rodovias a partir de 1907, atingindo cidades, vilas e povoados. A terceira é a fase rural, onde o controle da zoonose nas cidades levou à invasão de fazendas e sítios a partir da década de 1930. Por fim temos a fase silvestre na qual ocorre o assentamento da peste aos habitats brasileiros constituindo os focos naturais (Giles et al., 2011; Tavares et al., 2012).

Apesar dos longos períodos quiescentes, a peste vem apresentando novos surtos em várias partes do mundo. Na Argélia, a peste ressurgiu em 2003, com 18 casos e uma morte, após ausência de registros da doença por 57 anos (Bertherat *et al.*, 2007). Em 2008, neste mesmo país, em um foco rural distinto, quatro pessoas foram contaminadas e uma morreu de pneumonia (Bitam *et al.*, 2010). Após 25 anos sem registros, a doença reapareceu também na Líbia em 2009, afetando cinco pessoas, com uma vítima fatal (Cabanel *et al.*, 2013). Nos EUA, em 2010, foram registrados dois casos no Oregon, que estava livre da doença desde 1995 (Centers for Disease Control and Prevention (CDC), 2011). Em dezembro de 2016, a OMS noticiou um surto que estava ocorrendo em Madagascar, em um distrito localizado fora dos limites de uma área já conhecida como endêmica para a doença. Desde 1950 não havia relatos de casos nesta região (World Health Organization (WHO), 2017). No Brasil, embora casos de peste não tenham sido registrados nos últimos anos, a bactéria *Y. pestis* permanece firmemente arraigada em seus focos naturais, impossibilitando a sua erradicação (Almeida *et al.*, 2020; Tavares *et al.*, 2012, 2020).

A crença de que a peste já foi erradicada, e que ficou restrita a um passado distante da humanidade consiste em um grave erro, visto que a bactéria continua ativa e provocando mortes em todo o mundo. Os dados mais recentes apontam registros de infecção em várias partes do mundo, sendo o continente africano, especialmente em Madagascar, o local com a maior incidência de casos. Entre 2013 e 2018 foram registrados 2.323 casos apenas em Madagascar, mas também existem registros de 410 casos na República Democrática do Congo, 36 casos na Tanzânia e 22 em Uganda (Bertherat, 2019; Vallès *et al.*, 2020). Neste mesmo período alguns casos foram reportados na Ásia, tendo China e Mongólia reportado cinco casos em cada país, além das Américas onde 40 casos foram registrados no Peru e 3 na Bolívia. Mesmo em países mais desenvolvidos, com maior controle e melhores condições

sanitárias, casos são registrados, como os 40 casos registrados nos Estados Unidos ainda no período entre 2013 e 2018 (Bertherat, 2019; Vallès *et al.*, 2020).

2.2 O GÊNERO *YERSINIA*

O gênero *Yersinia* pertence a ordem Enterobacteriales e à família Yersiniaceae. Três espécies desse gênero possuem destaques como patógenos em humanos e animais: *Yersinia pestis*, o agente causador da peste, *Yersinia pseudotuberculosis* e *Yersinia enterocolitica* (biotipos 1B, 2, 3, 4 e 5), responsáveis pela Yersinose entérica (Guern *et al.*, 2020; Nguyen *et al.*, 2019). Além das três espécies de destaque, a nova espécie *Yersinia wautersii*, caracterizada em 2014, é reportada como “supostamente patogênica” (Savin *et al.*, 2014). Esta nova espécie foi designada a partir da análise de cepas da espécie *Y. pseudotuberculosis* através das técnicas denominadas de *Multi Locus Sequence Typing* (MLST) e *Average Nucleotide Identity* (ANI). Além da caracterização como nova espécie, alguns isolados de humanos apresentaram genes de virulência nos plasmídeos pYV e pVM82 (Guern *et al.*, 2020; Savin *et al.*, 2014). O gênero ainda possui duas espécies patogênicas apenas para animais: *Yersinia ruckeri*, agente causador da doença Enterite da Boca Vermelha (EBV) que acomete principalmente peixes salmonídeos, e *Yersinia entomophaga*, conhecida por causar doença em besouros na Nova Zelândia (*Costelytra zealandica*) (Guern *et al.*, 2020; Hurst *et al.*, 2015). Além disso, 14 espécies não patogênicas foram descritas desde os anos 1980: *Yersinia aldovae*, *Yersinia aleksiciae*, *Yersinia bercovieri*, *Yersinia frederiksenii*, *Yersinia intermedia*, *Yersinia kristensenii*, *Yersinia massiliensis*, *Yersinia mollaretii*, *Yersinia nurmii*, *Yersinia pekkanenii*, *Yersinia rohdei*, *Yersinia similis*, *Yersinia hibernica* e *Yersinia canariae* (Guern *et al.*, 2020).

Em uma análise recente, 38 cepas pertencentes ao Laboratório de Referência Nacional de *Yersinia* (YNRL) em Paris, França, isoladas de humanos e animais, foram avaliadas utilizando core genome multilocus sequence typing (cgMLST) (Savin *et al.*, 2019), e se mostraram como pertencentes a seis novas espécies do gênero *Yersinia*: *Yersinia artesianae*; *Yersinia proxima*; *Yersinia alsatica*; *Yersinia vastinensis*; *Yersinia thracica* e *Yersinia occitânica* (Guern *et al.*, 2020).

2.3 *Yersinia pestis*: AGENTE ETIOLÓGICO E DIAGNÓSTICO

A bactéria causadora da peste, *Y. pestis*, é um bacilo Gram-negativo, aeróbio ou anaeróbio facultativo, imóvel e não formador de esporos, crescendo em meios comuns (BHI: *Brain Heart Infusion* e BAB: *Blood Agar Base*), à temperatura ótima de 28 °C e pH entre 7,4 e 7,6, e é destruída rapidamente pela luz solar ou temperaturas acima de 40 °C. Pode permanecer viável em materiais como fezes de pulgas, escarros e tecidos protegidos (como a medula óssea), principalmente, quando em baixas temperaturas (Demeure *et al.*, 2019).

A primeira caracterização de *Y. pestis* foi realizada pelo médico franco-suíço Alexandre Yersin em Hong Kong no ano de 1894. Alexandre Yersin pertencia ao instituto Pasteur de Paris, França, e contribuiu também para o entendimento do ciclo de transmissão da *Y. pestis* envolvendo o roedor *Rattus rattus* e seu ectoparasita, uma pulga da espécie *Xenopsylla cheopis* (Barbieri *et al.*, 2020). A bactéria foi inicialmente nomeada *Bacterium pestis* até 1900, quando foi renomeada *Bacillus pestis*. Após 1923 foi classificada dentro do gênero *Pasteurella*, e denominada de *Pasteurella pestis*, sendo reclassificada como *Yersinia pestis* após o desmembramento do gênero *Pasteurella* em *Yersinia* e *Francisella* em 1974 (Perry; Fetherston, 1997).

A manutenção da infecção durante os períodos inter-epidêmicos pode ser explicada pela capacidade da *Y. pestis* de sobreviver por períodos prolongados no solo das galerias dos roedores, possibilitando a contaminação destes animais (Ayyadurai *et al.*, 2008). Existem várias hipóteses para explicar essa capacidade da *Y. pestis*, entre elas a sobrevivência em protozoários e a formação de biofilme na superfície de nematóides (Yang *et al.*, 2023). A notável aptidão da *Y. pestis* em permanecer viável nos mais diversos focos e em uma ampla variedade de condições climáticas e ecológicas está relacionada à sua capacidade de sofrer modificações genéticas em seu genoma, tais como aquisição e perda de genes, mutações pontuais (*Single Nucleotide Variations* – SNVs) e rearranjos (Bolotin; Hershberg, 2015).

As cepas de *Y. pestis* são bastante homogêneas quando analisadas por métodos fenotípicos. Apresentam apenas um sorotipo ou sorovar, um fagotipo ou fagovar e quatro biotipos ou biovars, caracterizados pela habilidade de fermentar glicerol e arabinose e reduzir nitrato a nitrito (Zhou *et al.*, 2004). Originalmente baseado em dados históricos e no conhecimento sobre os focos naturais da peste, Devignat (Devignat, 1951) identificou três biovars clássicos pela associação entre cepas responsáveis pelas pandemias de peste e as reações do nitrato e glicerol, que o levou a propor a seguinte nomenclatura: Antiqua

(glicerol+, nitrato+), referente à Pandemia de Justiniano; Medievalis (glicerol+, nitrato-), referente à Peste Negra; e Orientalis (glicerol-, nitrato+), referente à última pandemia. O esquema de classificação segundo Devignat (1951) pode ser visualizado na Figura 2.

Microtus, o quarto biovar, foi isolado de roedores dos focos da China e classificadas como cepas glicerol+, nitrato- e arabinose-. No seu genoma, observa-se um plasmídeo específico (pCRY), que codifica um sistema de secreção tipo IV, além de rearranjos genômicos no plasmídeo pMT1 e a presença de pseudogenes (3,5%) (Song, 2004).

Figura 2 - Distribuição das Variedades Geográficas de *Y. pestis* segundo Devignat (1951)

Variedades	Nitrato	Glicerol	Distribuição Geográfica
Antiqua	+	+	África (Zaire, Quênia, Tanzânia); China (Mandchúria, Mongólia)
Medievalis	-	+	Rússia, Kurdistão
Orientalis	+	-	Ásia (Índia, Birmânia, Viet-Nam, Tailândia, Camboja); África (África do Sul, Madagascar); América (Estados Unidos, Bolívia, Brasil, Equador, Perú)



G⁺ N⁺ **G⁺ N⁻** **G⁻ N⁺**
Antiqua *Medievalis* *Orientalis*

Fonte: Almeida et al. (2002).

Das provas fenotípicas as novas metodologias como WGS, outras classificações ou subtipagens têm sido propostas (Anisimov *et al.*, 2004; Vogler et al., 2016). Dependendo do esquema de classificação utilizado têm sido identificados vários grupos intraespécies (subespécies, ecótipos e/ou biovars) que têm todos diferentes distribuições geográficas e hospedeiros primários (Anisimov *et al.*, 2004).

Dada a alta letalidade da peste, o diagnóstico precoce, além de medidas preventivas como a vigilância epidemiológica, é extremamente importante para salvar a vida do paciente.

O diagnóstico em uma região tida como indene para peste pode ser difícil se baseado apenas no reconhecimento dos sintomas. Um exemplo disso é na forma bubônica, onde pode ocorrer a confusão com doenças sexualmente transmissíveis (DST) como sífilis ou outras infecções como tuberculose e até neoplasias, podendo, porém, ter o diagnóstico diferencial quando ocorre a presença de linfangite. Mesmo em regiões endêmicas, o diagnóstico da forma pneumônica pode ser ainda mais desafiador pois, no estágio inicial e sem a presença de bubão, os sintomas podem ser inespecíficos e às vezes detectados apenas radiologicamente (Barbieri *et al.*, 2020; Butler, 2014; Tavares *et al.*, 2020).

O padrão ouro para o diagnóstico da peste continua sendo o isolamento da *Y. pestis* em cultura, porém esse é um processo laborioso e demorado (no mínimo quatro dias) para o início precoce do tratamento (Barbieri *et al.*, 2020; Butler, 2013; Vallès *et al.*, 2020). Para obter o diagnóstico rápido, e conseqüentemente o início precoce do tratamento, técnicas moleculares como PCR (*Polymerase Chain Reaction*) ou MALDI-TOF MS (*Matrix-assisted Laser Desorption Ionization–time of Flight Mass Spectrometry*) podem ser utilizadas no lugar das técnicas bacteriológicas tradicionais, mesmo quando as bactérias não estão mais viáveis (Barbieri *et al.*, 2020; Butler, 2013; Leal; Almeida, 1999). No caso da PCR, os genes *cafI* e *pla* são tradicionais alvos para investigação de peste (Higgins *et al.*, 1998; Hinnebusch; Schwan, 1993; Loïez *et al.*, 2003; Rahalison *et al.*, 2000) porém existem também abordagens baseadas em PCR (Multiplex-PCR) que visam o diagnóstico e caracterização de patogenicidade com iniciadores (primers) para genes de virulência característicos de *Y. pestis* como *cafI* e *pla* (Leal; Almeida, 1999).

2.4 ASPECTOS CLÍNICOS DA PESTE

A peste é conhecida por apresentar três formas clínicas principais, sendo elas a peste bubônica, peste septicêmica e pneumônica (Vallès *et al.*, 2020). Existem, porém, outras formas conhecidas além das três principais como, por exemplo, a peste benigna, a peste meningea e a peste faríngea (Barbieri *et al.*, 2020).

As características clínicas da peste são parcialmente dependentes da via de contaminação. A forma mais comum de contaminação é através da picada de pulgas, que inoculam o bacilo intradermicamente, sendo esta rota mais associada à peste bubônica (Barbieri *et al.*, 2020; Vallès *et al.*, 2020). Além da inoculação através da pele, a *Y. pestis* pode infectar o hospedeiro por contato com a conjuntiva ocular ou por meio da mucosa do

aparelho respiratório, através da aspiração de gotículas expelidas por pacientes ou animais infectados. É possível ainda que a infecção ocorra através do aparelho digestivo por meio da deglutição de tecidos de animais infectados. O bacilo inoculado por via da conjuntiva ocular leva ao desenvolvimento da peste septicêmica, enquanto a inoculação por via intratraqueal está mais associada à peste pneumônica (Barbieri *et al.*, 2020).

A peste bubônica é a forma mais comum, e recebe este nome devido a formação de bubões, que são nódulos linfáticos doloridos e inchados (Barbieri *et al.*, 2020; Butler, 2013). Apesar de ter como principal porta de entrada a picada de pulga na pele, as lesões causadas pela picada não são bem descritas e raramente são mencionadas na literatura nos casos mais recentes, embora lesões cutâneas descritas como carbúnculos tivessem grande importância no diagnóstico da peste especialmente na segunda pandemia histórica (Barbieri *et al.*, 2020). O período de incubação ocorre entre dois e seis dias, e geralmente é seguido pelo aparecimento abrupto dos sintomas como febre alta, calafrio, cefaleia, dores generalizadas, mialgia, confusão mental, taquicardia entre outros (Butler, 2013). Os bubões comumente surgem no terceiro dia após a infecção, e podem ter tamanhos entre 1 e 10 cm, deixando a pele distendida e avermelhada e podendo assumir formato arredondado ou ovular. Os nódulos femorais e inguinais são os mais comuns, correspondendo à 31% e 24% dos casos respectivamente, enquanto ainda são reportados casos com nódulos axilares em 22% dos casos e, menos comuns, cervicais em 9% dos casos (Barbieri *et al.*, 2020).

Segunda forma mais comum, a peste septicêmica se difere inicialmente da forma bubônica por, frequentemente, não apresentar inchaços ganglionares visíveis. (Barbieri *et al.*, 2020). Pode-se desenvolver a partir do contato direto de secreções ou sangue infectado com lesões na pele ou com a conjuntiva ocular. Esta forma pode decorrer também secundariamente ao não tratamento correto de outras formas da doença. A forma septicêmica normalmente apresenta sintomas como febre muito alta (42°C ou mais), aceleração do pulso, hipotensão arterial, dispneia, hemorragia cutânea e até de órgãos internos e pode evoluir rapidamente para coma e morte em dois ou três dias (Butler, 2014).

Por fim, a terceira forma mais comum da peste, chamada de pneumônica, pode surgir como uma evolução secundária das outras formas ou pelo contato com aerossóis de pacientes infectados. Esta forma ainda requer atenção pela possibilidade do uso da peste como arma biológica para bioterrorismo (Butler, 2013; Vallès *et al.*, 2020). Esta forma tem um período de incubação de dois a três dias e tem uma rápida evolução para febre alta (acima dos 40°C), calafrios, arritmia, hipotensão, náusea, e confusão mental entre outros sintomas. Os sintomas

pulmonares no início são discretos ou ausentes, mas evoluem para forte dor torácica, insuficiência respiratória, cianose e expectoração. A expectoração pode ser aquosa, espumosa ou hemorrágica, sendo em ambos os casos repleta de bacilos (Butler, 2014).

2.5 TRANSMISSÃO E VETORES DA PESTE

A peste é uma zoonose que atinge principalmente roedores, mas também outros mamíferos incluindo humanos, sendo estes hospedeiros incidentais. O ciclo bioecológico da peste é complexo e envolve várias fases, fatores ambientais, hospedeiros e vetores (Vallès *et al.*, 2020). Os hospedeiros exercem uma função de reservatório, que pode ter função primária ou secundária, sendo a principal diferença entre esses tipos o fato de que hospedeiros primários são pouco suscetíveis à infecção enquanto os secundários são susceptíveis e geralmente morrem da doença (Brasil, 2008).

Os ciclos bioecológicos da peste podem ser divididos em quatro: Enzoótico, epizoótico, urbano e pneumônico. No ciclo enzoótico a transmissão ocorre entre hospedeiros primários, por exemplo por meio da contaminação do solo das galerias (Barbieri *et al.*, 2020; Fernandes *et al.*, 2020; Vallès *et al.*, 2020). Já no ciclo epizoótico ocorre a transmissão entre hospedeiros secundários que frequentemente morrem em decorrência da doença. Após a morte, as pulgas buscam outros hospedeiros, mantendo o ciclo de infecções, e alternativamente a infecção pode ocorrer pela ingestão do animal morto por outros animais carnívoros (Barbieri *et al.*, 2020). O ciclo urbano ocorre quando um hospedeiro infectado no ciclo epizoótico entra em contato com potenciais hospedeiros urbanos, e as pulgas infectadas transmitem para roedores comensais e humanos. Os humanos infectados, por sua vez desenvolvem peste bubônica que pode evoluir para peste pneumônica e assim dando início ao ciclo pneumônico, ocorrendo a transmissão direta entre humanos pela inalação de aerossóis infectados (Brasil, 2008).

Existem mais de duzentas espécies de mamíferos susceptíveis à infecção por *Y. pestis*, sendo a maioria composta por roedores, e as espécies resistentes são consideradas fontes da bactéria (Barbieri *et al.*, 2020; Gage; Kosoy, 2005; Mahmoudi *et al.*, 2021; Perry; Fetherston, 1997). Os roedores são mamíferos que tem por principais características uma estrutura dentária com ausência de caninos e incisivos sem raiz, que crescem indefinidamente. Eles pertencem a ordem *Rodentia*, sendo essa a maior ordem da classe *Mammalia*, e suas mais de duas mil espécies correspondem a cerca de 40% dos mamíferos existentes, habitando

basicamente qualquer ambiente terrestre e sendo muito adaptáveis a diversas condições climáticas, terrenos e altitudes, apresentando muitas adaptações fisiológicas. Nos focos de peste brasileiros, os principais roedores infectados que foram capturados nas ações de vigilância foram das espécies *Necromys lasiurus* e *Rattus rattus*, correspondendo por cerca de 79% dos achados (Fernandes *et al.*, 2020).

Os principais vetores da peste bubônica entre roedores são as pulgas. Estes ectoparasitas pertencem à ordem *Siphonaptera*, que possui mais de três mil espécies (Barbieri *et al.*, 2020; Gage; Kosoy, 2005). Os adultos de ambos os gêneros desses insetos são hematófagos e, apesar da predominância em roedores, também tem como hospedeiros outros mamíferos e aves. Morfologicamente possuem coloração castanha, corpo achatado, três pares de patas e não possuem asas (Brasil, 2008). No Brasil, as principais espécies de pulgas que foram encontradas infectadas nas ações de vigilância e controle da peste são dos gêneros *Polygenis*, *Xenopsylla*, *Pulex* e *Ctenocephalides* (Fernandes *et al.*, 2020)

Atualmente são descritos dois mecanismos principais de transmissão da peste por pulgas. O primeiro mecanismo de transmissão foi descrito por Bacot e Martin em 1914 (Bacot; Martin, 1914) que ocorre pela multiplicação do bacilo no estômago do hospedeiro, preenchendo uma porção do tubo digestivo da pulga chamada de proventrículo, um fenômeno conhecido como bloqueio do proventrículo e que é dependente de biofilme (Hinnebusch; Jarrett; Bland, 2017). O esforço realizado para se alimentar faz com que as pulgas “bloqueadas” regurgitem o conteúdo do proventrículo causando a inoculação do bacilo. O biofilme bloqueia a pulga e ela não consegue se alimentar, passando seus últimos dias fazendo repetidas tentativas de alimentação e conseqüentemente aumentando o potencial de transmissão (Hinnebusch; Jarrett; Bland, 2017). O outro mecanismo é conhecido como transmissão de fase inicial (Hinnebusch; Jarrett; Bland, 2017). As pulgas possuem a capacidade de processar rapidamente o sangue do qual se alimentou, sendo este um mecanismo comum de “limpeza” contra patógenos, porém *Y. pestis* possui mecanismos para evitar ser excretada pela pulga (Hinnebusch; Jarrett; Bland, 2017; Hinnebusch; Perry; Schwan, 1996). Após a infecção, a pulga tem o potencial de transmitir o bacilo na próxima alimentação ao longo da semana seguinte. Esse tipo de transmissão foi inicialmente chamado de transmissão em massa devido ao fato da doença raramente se desenvolver se não houver um grupo de pulgas se alimentando do mesmo hospedeiro simultaneamente (Hinnebusch; Jarrett; Bland, 2017). Já o termo transmissão de fase inicial é mais recente e enfatiza o fato

que o pico de transmissão ocorre logo na alimentação seguinte da pulga nos dias que se seguem após a infecção (Eisen *et al.*, 2006).

2.6 TRATAMENTO DA PESTE

O tratamento precoce para a peste é de suma importância dada a alta taxa de letalidade da doença. Após a descoberta e caracterização da *Y. pestis*, Alexandre Yersin passou a trabalhar em um antissoro em colaboração com outros pesquisadores do Instituto Pasteur em 1895. O primeiro tratamento com este antissoro foi realizado em 1896 em 23 pacientes chineses e foi capaz de reduzir a taxa de mortalidade para apenas 9% (Barbieri *et al.*, 2020). O uso deste tratamento foi reportado como responsável pela redução da mortalidade de 70% para 35% na década seguinte ao seu desenvolvimento com base em dados da Ásia, África e América do Sul (Butler, 2014; Meyer *et al.*, 1952).

Dado os graves efeitos colaterais e a proteção apenas de curto prazo do antissoro para *Y. pestis*, progressivamente o tratamento foi substituído pelas novas drogas desenvolvidas. Sulfonamidas foram utilizadas pela primeira vez no tratamento da peste em 1938, em pacientes da África Oriental e Índia (Anisimov; Amoako, 2006; Barbieri *et al.*, 2020) e a injeção de sulfapiridina foi reportada como responsável pela redução da mortalidade para apenas 24% em paciente de Madagascar em 1940 (Barbieri *et al.*, 2020). Atualmente as sulfonamidas são tidas como drogas de segunda linha e utilizado apenas na ausência de drogas mais potentes (Butler, 2014).

O padrão-ouro no tratamento da *Y. pestis* é a estreptomicina, sendo particularmente eficaz no tratamento da forma pneumônica da peste (Anisimov; Amoako, 2006; Barbieri *et al.*, 2020; Valles *et al.*, 2020). Apesar da taxa de sucesso no tratamento e redução da mortalidade com a estreptomicina, casos reportados de alta toxicidade renal de paciente e fetos em mulheres grávidas levaram a busca de alternativas à estreptomicina (Butler, 2014). A gentamicina foi uma das drogas desenvolvidas e, exceto em pacientes em estágio terminal, pode ser utilizada alternativamente à estreptomicina especialmente para crianças ou gestantes, evitando os efeitos colaterais da estreptomicina nesses grupos (Barbieri *et al.*, 2020; Butler, 2014). As drogas mais novas em uso como alternativa ao padrão-ouro são as tetraciclina, como doxiciclina, sendo indicadas em casos não complicados (Barbieri *et al.*, 2020; Valles *et al.*, 2020). Já para casos em que ocorrem complicações com infecções em espaços tissulares,

como meningite, onde as drogas previamente mencionadas não atingem nível terapêutico, é possível utilizar o cloranfenicol (Anisimov; Amoako, 2006; Valles *et al.*, 2020).

Uma classe mais nova de drogas em estudo para o tratamento da peste são as fluoroquinolonas, que se mostraram efetivas em estudos com animais (Bonacorsi *et al.*, 1994; Butler, 2014; Russell *et al.*, 1996). Apesar de ainda precisar de estudos comparativos em humanos, as fluoroquinolonas já foram utilizadas com sucesso em um paciente nos Estados Unidos e cinco pacientes na Argélia (Butler, 2014; Kuberski; Robinson; Schurgin, 2003; Raoult *et al.*, 2013).

Por fim, pode-se citar também terapias alternativas aos antimicrobianos como a terapia por bacteriófagos. Os fagos são extremamente espécie-específicos, não são tóxicos para eucariotos e o processo de replicação termina por matar a bactéria alvo. A terapia com fagos já foi utilizada no passado com sucesso, utilizando em 1925 um fago muito virulento para *Y. pestis* isolado de ratos da Indochina em 1920 (Anisimov; Amoako, 2006). Com o desenvolvimento das drogas antimicrobianas, a terapia por fagos caiu em desuso ao longo dos anos, mas novos estudos estão surgindo utilizando terapia por fagos derivado aos eventos de resistência bacteriana reportados para *Y. pestis* (Yuan *et al.*, 2020).

2.7 RESISTÊNCIA ANTIMICROBIANA POR *Y. pestis*

Resistência bacteriana aos antimicrobianos é um problema global e ocorre devido a conhecida capacidade das bactérias de transferir a resistência através de plasmídeos via conjugação para cepas não resistentes (Ditchburn; Hodgkins, 2019; Galimand *et al.*, 1997; Galimand; Carniel; Courvalin, 2006). A *Y. pestis* é sensível a maioria dos antimicrobianos com exceção à penicilina e derivados, tendo a estreptomicina como antimicrobiano de primeira escolha para tratamento e drogas como cloranfenicol e tetraciclina como alternativas. Além disso, tetraciclina e sulfonamidas são recomendadas na profilaxia (Galimand *et al.*, 1997).

Em 1995 duas cepas resistentes de *Y. pestis* foram reportadas em Madagascar, sendo uma multirresistente a todas as drogas recomendadas para o tratamento da peste (Galimand *et al.*, 1997; Galimand; Carniel; Courvalin, 2006). A cepa 17/95 foi obtida de amostra clínica de um paciente do sexo masculino de 16 anos e possuía um plasmídeo de 150 kb chamado pIP1202 que apresentou resistência a oito antimicrobianos usados tanto para o tratamento quanto para profilaxia da peste, além de também apresentar resistência a drogas alternativas

como ampicilina e espectinomicina (Ditchburn; Hodgkins, 2019; Galimand *et al.*, 1997; Galimand; Carniel; Courvalin, 2006). A segunda cepa resistente, chamada de 16/95, também foi obtida em amostra clínica e foi encontrado um plasmídeo de 40 kb, chamado de pIP1203, que apresentava resistência apenas à estreptomicina, porém permanecia susceptível à outras drogas (Ditchburn; Hodgkins, 2019; Galimand *et al.*, 1997; Galimand; Carniel; Courvalin, 2006). Pesquisas mais recentes também demonstraram que uma mutação no gene *rpsL* conferiu resistência à estreptomicina em uma amostra coletada na China (Dai *et al.*, 2021).

Acredita-se que origem da resistência antimicrobiana das cepas de Madagascar ocorreu por transferência conjugativa (Chanteau *et al.*, 2000; Guiyoule, 2001). Em mamíferos, o bacilo infecta a derme, o tecido linfoide e o sangue, locais normalmente estéreis, o que dificulta a troca conjugativa pois seria necessária a presença de uma segunda infecção por outra bactéria simultaneamente. Portanto este tipo de troca genética é mais provável no trato digestivo, não estéril, de um ectoparasita (pulga) vetor da *Y. pestis* (Hinnebusch *et al.*, 2002). O aumento do nível de resistência a antimicrobianos em Madagascar levou o governo a pressionar por pesquisas de alternativas para o tratamento da peste, fazendo da terapia por bacteriófagos a principal alternativa (Ditchburn; Hodgkins, 2019). Além da resistência adquirida por plasmídeo, um recente estudo mostrou que uma mutação no gene *rplS* conferiu resistência à estreptomicina em *Y. pestis* (Dai *et al.*, 2021).

2.8 FATORES DE VIRULÊNCIA DA *Y. pestis*

Sabe-se que a *Y. pestis* é uma bactéria muito virulenta, e que acarreta alta letalidade em humanos. Apesar dessa alta virulência, é conhecido o fato que a *Y. pestis* divergiu evolutivamente da *Yersinia pseudotuberculosis*, sendo este último um patógeno relativamente benigno (Perry; Fetherston, 1997; Spyrou *et al.*, 2018; Susat *et al.*, 2020). O ciclo de vida complexo do bacilo, que precisa sobreviver, crescer e multiplicar em organismos tão distintos como ectoparasitas e mamíferos, indicam que fatores de virulência podem ser determinantes pois afetam os hospedeiros por meio da aderência, modulação de resposta imunológica, interrupção do metabolismo celular ou obtenção nutrientes essenciais (Perry; Fetherston, 1997). Para muitos patógenos uma rápida resposta do sistema imune inato diminui o progresso da infecção, enquanto patógenos com grande habilidade de evasão ou supressão da resposta inata, como *Y. pestis*, podem causar a morte do hospedeiro antes de uma efetiva resposta do sistema imune adaptativo (Montminy *et al.*, 2006).

Os fatores de virulência da *Y. pestis* são complexos e recaem em um arsenal de regulação refinada, codificados por genes presentes tanto em plasmídeos quanto no cromossomo. Tipicamente as cepas de *Y. pestis* apresentam três plasmídeos associados a virulência (pYV/pCD1, pPla/pPCP1 e pFra/pMT1) e uma região no cromossomo chamada de “Ilha de patogenicidade” (Butler, 2013; Demeure *et al.*, 2019). Entre os fatores de virulência da *Y. pestis* os que estão melhor caracterizados são as Yops (Yersinia outer membrane proteins ou Proteínas de Membrana Externa de *Yersinia*) que servem como efetores da bactéria e são secretados por um sistema conhecido como T3SS (Type 3 Secretion System) (Demeure *et al.*, 2019; Nauth *et al.*, 2018).

O primeiro processo para uma efetiva atuação das Yops é o estabelecimento da adesão celular da bactéria à célula hospedeira e alguns genes do cromossomo da *Y. pestis* possuem papéis fundamentais nesse processo. Um claro exemplo é a proteína de membrana externa Ail, que é codificada pelo gene *ail* presente no cromossomo e que atua no processo de adesão celular, entrega de Yop e resistência (Thomson; Plecha; Krukoniš, 2019). Outro importante fator promotor na secreção das Yops codificado por gene cromossomal é o pH 6 *antigen*, que inibe a fagocitose da *Y. pestis* por macrófagos e participa na mediação de adesão a células epiteliais (Bao *et al.*, 2013). Genes do cromossomo são ainda responsáveis pela captação de metais essenciais como ferro e zinco, um mecanismo de proteção da bactéria que ocorre através de secreção de sideróforo yersiniabactina (Ybt) (Fetherston *et al.*, 2010).

Os plasmídeos associados à *Y. pestis* também atuam em várias frentes no decorrer da doença. O gene *pla*, codificado pelo plasmídeo de 9.5kb chamado pPla/pPCP1, participa na codificação de proteína externa da membrana com papel fundamental de adesão, invasão, fibrinolítica e de coagulase (Demeure *et al.*, 2019). Estudos apontam também o papel do gene *pla* na clivagem de vários substratos de células hospedeiras como plasminogênio, componente do complemento C3 e ligantes Fas/FasL (Zimble *et al.*, 2016). A proteína Pla também é encontrada junto com outras proteínas como Ail e F1 em vesículas na membrana externa, que pode servir para a distribuição e entrega desses fatores de virulência (Demeure *et al.*, 2019). O antígeno F1 é codificado a partir do gene *caf*, este por sua vez pertencente a outro plasmídeo associado à virulência chamado pFra/pMT1, tendo o antígeno F1 participação na inibição de adesão para evitar a absorção por macrófagos e servindo portando como mecanismo de escape (Demeure *et al.*, 2019; Liu; Chen; Galván, 2006). Por fim, o plasmídeo pYV/pCD1 está associado da codificação das Yops, tendo essas papéis fundamentais na infecção, como a inibição da Rho GTPases e rompimento do citoesqueleto de actina para inibir a fagocitose,

reduzir a produção de citocinas pró-inflamatórias e induzir morte celular (Trosky; Liverman; Orth, 2008).

2.9 EVOLUÇÃO GENÉTICA DA *Y. pestis*

A peste é uma doença que aflige a humanidade a mais de 5000 anos de acordo com documentos históricos e científicos, evoluindo geneticamente ao longo deste período para se adaptar a diversos hospedeiros, vetores e ambientes (Barbieri *et al.*, 2020; Rasmussen *et al.*, 2015; Spyrou *et al.*, 2018). Estudos genéticos de cepas da era do bronze (aproximadamente 3300 AC) mostram por exemplo que a capacidade de transmissão por pulgas é uma evolução relativamente recente (entre 3 e 4 mil anos), e que essas cepas não eram capazes de causar a peste bubônica (Rasmussen *et al.*, 2015; Spyrou *et al.*, 2018).

Um dos primeiros estudos de DNA bacteriano ancestral foi conduzido em 1998, onde a polpa dentária de restos mortais de humanos que viveram entre os séculos 16 e 18, escavados na França, foi utilizada para fazer a extração de um fragmento de DNA com 133 bp do gene *rpoB* de *Y. pestis* (Drancourt *et al.*, 1998). A polpa dentária foi utilizada devido a sua estrutura que permite proteção à contaminantes externos, mantendo a morfologia de células sanguíneas intactas que podem incluir patógenos como *Y. pestis* (Spyrou *et al.*, 2019). Os estudos evoluíram e passaram a buscar outras sequências genéticas específicas para *Y. pestis* como os genes cromossômicos *rpoB* e *glpD* (Drancourt *et al.*, 1998; Tran *et al.*, 2011) e os genes plasmidiais como *cafI* e *pla* (Keller *et al.*, 2019; Namouchi *et al.*, 2018).

Os genomas mais antigos de *Y. pestis* já sequenciados são de cepas que datam de períodos anteriores as pandemias históricas e foram sequenciados a partir de ossadas do final do período neolítico da idade do bronze (Barbieri *et al.*, 2020). A cepa ancestral mais antiga que se conhece (Gok2) divergiu de todas as outras cepas de *Y. pestis* a mais de 5700 anos, e as cepas da era do bronze pertencem a uma linhagem independente que surgiu na Eurásia (Rascovan *et al.*, 2019). A segunda divergência ocorreu a 4000 anos na Eurásia dando origem a várias linhagens que podem ser ancestrais às responsáveis pelas pandemias históricas (Spyrou *et al.*, 2018). Nos genomas ancestrais é observada a ausência de uma região de 19kb no plasmídeo pMT1 associado ao gene de virulência *ymt*, que codifica uma fosfolipase D relacionada à sobrevivência e a replicação da *Y. pestis* no proventrículo da pulga (Rascovan *et al.*, 2019; Rasmussen *et al.*, 2015).

Outras cepas da idade do bronze apresentaram ausência do gene *ympt1.66c* associado à sobrevivência do bacilo dentro do macrófago, e consequentemente do progresso da peste bubônica (Yu *et al.*, 2020). A ausência do gene *ymt* dificulta também a transmissão do bacilo através da pulga, uma vez que este gene tem um papel importante na formação de biofilme no estágio final da transmissão, o que pode explicar a ausência histórica de grandes epidemias neste período (Barbieri *et al.*, 2020).

Estudos genéticos de esqueletos dos séculos 6 a 8, de vários países da Europa, confirmaram a *Y. pestis* como responsável pela primeira pandemia e mostraram uma grande diversidade genética nessas cepas (Keller *et al.*, 2019). Foi observado nesses genomas deleções nos genes *pheA* e *celB*, e em cepas do final da primeira pandemia foi observada a deleção de genes de virulência *mgtB* e *mgtC* que podem estar associados ao fim dessa pandemia (Barbieri *et al.*, 2020; Keller *et al.*, 2019). Ainda não existe consenso sobre a origem da primeira pandemia histórica pois existem dados genéticos e históricos que se contradizem, apontando possíveis origens tão distintas como a China (Damgaard *et al.*, 2018) e o Egito (Green *et al.*, 2014).

Atualmente existem 56 genomas sequenciados de *Y. pestis* que datam da segunda pandemia histórica, mais conhecida como peste negra. Esses genomas foram identificados filogeneticamente como uma linhagem única de *Y. pestis* responsável por esta pandemia, dando suporte a hipótese de um único ponto de entrada da peste na Europa com a propagação por vias terrestre e marítimas. (Bos *et al.*, 2016; Namouchi *et al.*, 2018; Spyrou *et al.*, 2019). Esses genomas trazem evidências de microevoluções da bactéria dentro desta linhagem no fim da segunda pandemia dando origem a várias cepas europeias entre os séculos 15 e 18 (Spyrou *et al.*, 2019). Nos genomas ancestrais foram constatadas as ausências dos genes *mgtB* e *mgtC*, porém estes mesmos genes estão presentes nas cepas da segunda pandemia e esses genomas da segunda pandemia tem pouca diferença dos genomas modernos (Barbieri *et al.*, 2020). Uma diferença, porém, das cepas da primeira e segunda pandemias em relação às cepas atuais do biovar Orientalis é que as cepas atuais não possuem uma ilha genômica de 15kb correspondente à DFR4 (difference region 4), e esta diferença do perfil de DFR reflete a origem do biovar Orientalis (Radnedge *et al.*, 2002). Esta região deletada inclui genes possivelmente de virulência como *ccm2A*, o qual tem um papel na infecção humana pouco compreendido (Barbieri *et al.*, 2020).

No Brasil sabe-se que os focos estabelecidos são fruto de um único ponto de introdução da *Y. pestis* durante a terceira pandemia (Vogler *et al.*, 2019), e várias tentativas de

genotipagem dos isolados não encontraram diferenças, mesmo em hospedeiros e origens variadas (Oliveira *et al.*, 2012). A grande similaridade genética entre as cepas brasileiras torna difícil o rastreamento epidemiológico, mas algumas correlações entre grupos genéticos e suas origens temporais e geográficas já foram encontradas (Barros *et al.*, 2013, 2014; Oliveira *et al.*, 2012).

2.10 GENÔMICA COMPARATIVA E DIVERSIDADE BACTERIANA

A evolução das técnicas e ferramentas de sequenciamento nos últimos anos permitiram que sejam produzidos volumes de dados biológicos impensáveis em décadas anteriores. Com as tecnologias de Sequenciamento de Nova Geração (*Next Generation Sequencing* - NGS) passamos do sequenciamento de poucos kilobases para uma escala de sequenciamento de genomas completos, que demandam pouco tempo de sequenciamento e a custos cada vez mais acessíveis. Por outro lado, o volume de dados gerados torna a análise manual impraticável, fazendo dos métodos *in silico* (computacionais) abordagens cada vez mais relevantes.

A biologia computacional vem revolucionando as pesquisas com dados biológicos, e muitos métodos foram desenvolvidos para análise de sequências de DNA, RNA e proteínas, além de bases de dados que armazenam esses dados biológicos (Bhasin; Raghava, 2006; Crécy-Lagard; Hanson, 2013). Entre os métodos desenvolvidos, a genômica comparativa é uma ferramenta fundamental de análise onde sequências de genomas completos e conjuntos completos de genes são comparados para elucidar características e diferenças genômicas entre organismos (Crécy-Lagard; Hanson, 2013; Kawashima, 2018). Três aspectos principais devem ser considerados em uma análise de genômica comparativa: Estrutura do genoma, tipos de mudanças estruturais e distância evolucionária (Kawashima, 2018) e atualmente há várias ferramentas e métodos computacionais destinados a resolver essas questões.

O sequenciamento de genoma completo (*Whole Genome Sequencing* – WGS) de bactérias virou rotina, e a subsequente montagem *de novo* das leituras em *contigs* e/ou *scaffolds* já é algo bem estabelecido, além de ferramentas para anotação das características relevantes desses *contigs* (Seemann, 2014). Os dados de WGS são frequentemente usados para identificação de variantes, especialmente polimorfismo de nucleotídeo único (Single Nucleotide Variations – SNP) mas também podem ser associados à filogenia, sorologia ou

filogeografia, onde SNPs podem prover marcadores de mudanças evolucionárias entre genomas (Sahl *et al.*, 2016).

Outra abordagem de genômica comparativa é a análise de pangenoma, onde são analisados genes compartilhados por genomas e genes acessórios (Page *et al.*, 2015). O conceito de pangenoma parte do pressuposto que o repertório de genes de uma espécie microbiana é significativamente maior que o número de genes de uma cepa individual, sendo, portanto, um percentual significativo de cada genoma específico de uma cepa (Field; Wilson; Gast, 2006; Medini *et al.*, 2005; Tettelin *et al.*, 2008). Uma espécie bacteriana pode ser descrita em seu pangenoma, composto de genes *core* e genes acessórios, onde os genes *core* representam o conjunto de genes presentes em todas as cepas analisadas na pangenômica e os genes acessórios são os genes presentes apenas em algumas dessas cepas (Medini *et al.*, 2005; Tettelin *et al.*, 2008). O alinhamento de isolados contra um genoma de referência pode falhar ao não levar em conta sequências ausentes no genoma de referência, e a análise de pangenoma pode contornar esta limitação ao fazer a análise de montagens *de novo* sem um genoma de referência (Page *et al.*, 2015; Sherman; Salzberg, 2020). Especificamente em bactérias, onde a aquisição de genes de outros organismos é frequente, a análise de pangenômica pode prover uma melhor visão de genes conservados e genes acessórios, o que pode levar a uma melhor compreensão de processos chaves como seleção e evolução (Page *et al.*, 2015).

Uma abordagem que vem ganhando popularidade nos últimos anos é a de gene-a-gene (conhecida em inglês como *Gene-by-Gene* ou simplesmente GbG). A abordagem de GbG recai na comparação de genomas de interesse contra um esquema pré-definido, geralmente utilizando alguma técnica de alinhamento (Silva *et al.*, 2018). Esta técnica é considerada uma evolução do Multilocus Sequence Typing (MLST), que leva em consideração alguns poucos genes housekeeping na tipagem de microrganismos (Maiden *et al.*, 1998). Ao utilizar as abordagens de GbG, podemos utilizar o genoma *core* na elaboração do esquema para MLST ou ainda os genomas completos, sendo essas técnicas conhecidas como *core* genome MLST (cgMTSL) e *whole-genome* MLST (wgMLST)(Silva *et al.*, 2018; Zhou *et al.*, 2018).

Muito do conhecimento recente sobre *Y. pestis* vem de estudos utilizando genômica comparativa *in silico*. Um claro exemplo é encontrado em Keller *et al.* (2019), onde a diversidade genética de *Y. pestis* de cepas da primeira pandemia histórica foi avaliada através da análise de SNVs de genomas recuperados de ossada humanas de vítimas da peste. Em

outro estudo com cepas da China (Shi *et al.*, 2021) uma nova linhagem de *Y. pestis* foi encontrada utilizando análise de pangenômica baseado em presença e ausência de genes.

3 JUSTIFICATIVA

Considerando a epidemiologia complexa e a natureza cíclica dos surtos de peste, não se pode descartar a possibilidade do surgimento de um novo surto no Brasil. Assim, um estudo genético amplo e detalhado *Y. pestis* no território brasileiro é de extrema importância para um melhor entendimento sobre a disseminação e transmissão da bactéria na natureza e sobre a dinâmica da epidemiologia da doença.

A análise dos dados genômicos da coleção de cepas brasileiras de *Y. pestis* (Fiocruz/CYP) por meio de ferramentas de bioinformática permite avaliar a estrutura filogenética dessas cepas e comparar os dados genômicos com os dados epidemiológicos, contribuindo para o entendimento dos processos de microevolução e de adaptação às diferentes localidades e hospedeiros ao longo do tempo.

A identificação e a caracterização de subpopulações de *Y. pestis* no Brasil pode auxiliar e direcionar os esforços para identificar fatores ecológicos e ambientais associados aos potenciais riscos da doença. O estudo das cepas de *Y. pestis* isoladas de roedores, pulgas e humanos através de um gradiente local e temporal é importante para determinar quais espécies hospedeiras e de vetores estão envolvidas na manutenção enzoótica das subpopulações e delimitar seus limites geográficos, aperfeiçoando os programas de vigilância e controle da zoonose nas áreas focais.

Definir as associações entre variantes específicas e cepas observadas em um tipo de hospedeiro, em localidade específica e/ou determinado período possibilita a detecção precoce da fonte de infecção na ocorrência de um novo surto. Além disso, identificar se o isolado é endêmico e quais hospedeiros e localidades são preferenciais, direcionando melhor as estratégias de contenção.

4 PERGUNTA CONDUTORA

É possível associar variações genéticas das cepas brasileiras da coleção de *Y. pestis* da FIOCRUZ PE (Fiocruz/CYP) com seus locais, períodos e hospedeiros de isolamento através da análise de genômica comparativa *in silico*?

5 HIPÓTESE

A análise de genômica comparativa pode traçar correlações entre variações genéticas específicas das cepas brasileiras da coleção de *Y. pestis* da FIOCRUZ PE e seus respectivos dados epidemiológicos.

6 OBJETIVOS

6.1 OBJETIVO GERAL

Empregar abordagens *in silico* juntamente com dados de sequenciamento genômico para genômica comparativa de isolados brasileiros de *Y. pestis*.

6.2 OBJETIVOS ESPECÍFICOS

- a) Identificar variantes individuais em cepas de *Y. pestis*;
- b) Determinar a estrutura populacional e as relações evolutivas entre as cepas de *Y. pestis*;
- c) Pesquisar fatores de virulência e genes de resistência a antimicrobianos nos genomas de *Y. pestis*;
- d) Buscar associações entre variantes específicas e cepas que ocorrem apenas em um determinado hospedeiro, localidade e/ou período;

7 MATERIAIS E MÉTODOS

O presente estudo utilizou os dados brutos de sequenciamento genômico disponíveis para as amostras da coleção de culturas de *Y. pestis* da FIOCRUZ PE para fazer análise genômica comparativa utilizando três abordagens principais. Na primeira abordagem foi realizada uma montagem por referência das leituras de sequenciamento das cepas analisadas neste estudo contra a cepa de referência CO92 (Acesso/NCBI: GCF_000009065.1), e assim foi realizada a comparação da diversidade genética dessas cepas avaliando a presença de SNVs. Essas informações de diversidade foram utilizadas para realizar uma análise filogenética. Para a segunda abordagem, foram realizadas montagens *de novo*, seguidas de predição e anotação dos genomas, e por fim análise de pangenoma, realizando também uma análise filogenética baseada no genoma *core*, bem como no genoma acessório pela presença e ausência de genes. Na terceira abordagem, foi utilizada a técnica de GbG comparando alelos dos genes do genoma *core* através de um cgMLST, onde foram utilizados 61 genomas completos de *Y. pestis* disponíveis na base de dados do NCBI, incluindo o genoma de referência CO92, para a construção do esquema cgMLST. Os dados dessas análises foram novamente correlacionados aos dados epidemiológicos.

7.1 ISOLADOS DO ESTUDO

O Serviço de Referência Nacional em Peste (SRP) do Instituto Aggeu Magalhães (IAM) FIOCRUZ PE conta atualmente com 406 genomas sequenciados de cepas brasileiras de *Y. pestis*, pertencentes à coleção de culturas de *Y. pestis* (Fiocruz/CYP). Essas cepas foram isoladas de casos em humanos, de roedores e de pulgas em diversos focos do Nordeste brasileiro, de epidemias e de períodos endêmicos durante as ações de vigilância entre 1966 e 1997 (Rocha *et al.*, 2010). Este isolados estão mantidos em tubos de ágar simples com tampa de borracha, sob refrigeração a 4°C (Leal *et al.*, 2016). Esta coleção está inserida no SpeciesLink, sistema distribuído de informação que integra dados primários de coleções científicas, desenvolvido pelo Centro de Referência em Informação Ambiental (CRIA) e registrada na World Federation Culture Collection (WFCC). Os genomas foram sequenciados utilizando a plataforma Illumina HiSeq 2000 (Illumina, Inc., San Diego, CA) utilizando os protocolos padrão para o TruSeq SBS Kit v3-HS (Illumina, Code FC-401-3001) (Vogler *et al.*, 2019). Foi utilizada a abordagem de sequenciamento paired-end onde ambas as

extremidades de um fragmento são sequenciadas gerando dados de sequência de alta qualidade.

7.2 ANÁLISE DO SEQUENCIAMENTO DOS ISOLADOS

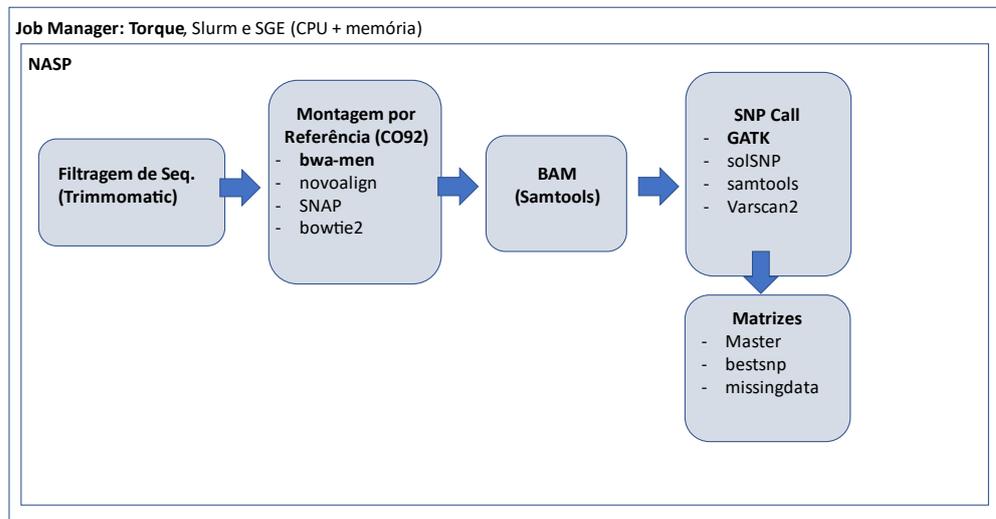
A qualidade dos dados de sequenciamentos das 406 cepas de *Y. pestis* foi avaliada utilizando o software FastQC 0.11.8 (Andrews, 2010). O FastQC gerou dados como número total de sequências, tamanhos das sequências, percentual do conteúdo GC e o Phred (valor de medição da qualidade) das bases. Dado o grande número de sequência e dados gerados no FastQC, o software MultiQC v1.10 (Ewels *et al.*, 2016) foi utilizado para consolidar os resultados onde foram observadas leituras em torno de 2.1 milhões de pb, conteúdo GC de 47% e profundidade média de 42x.

7.3 ABORDAGEM DE MONTAGEM POR REFERÊNCIA E ANÁLISE DE SNVS

7.3.1 Pipeline para mapeamento e análise de SNVs

A montagem por mapeamento das 406 sequências de *Y. pestis* contra a cepa de referência CO92 e análise de SNVs foram realizados utilizando o *pipeline* NASP (Sahl *et al.*, 2016). Este *pipeline* foi desenvolvido para fazer análise WGS, passando pela montagem/anotação de genomas e avaliação de SNVs para estudos como os de filogenia baseado em SNVs, trazendo várias opções de ferramentas computacionais que são executadas nas diversas etapas do *pipeline*. A Figura 3 mostra o esquema de funcionamento do NASP.

Figura 3 - Esquema de funcionamento do *pipeline* NASP



Fonte: O autor

Legenda: A Figura ilustra o funcionamento do NASP. O retângulo mais externo representa o gerenciador de recursos utilizado para limitar o uso de recursos computacionais. O retângulo interno representa o NASP e os vários programas utilizados. Na primeira etapa é feita a filtragem das sequências dadas de entrada no NASP. Esta etapa é seguida pela montagem por referência, onde vários programas podem ser utilizados para esta tarefa, sendo o BWA o programa escolhido neste projeto. Arquivos BAM são gerados pelo SAMTools e em seguida é feita a análise de SNVs, que possui várias opções de programas, mas neste projeto foi utilizado o GATK. Por fim são geradas as matrizes com os resultados das análises.

O primeiro passo realizado pelo NASP é a configuração das ferramentas e ambiente que serão executadas na análise. Nas configurações do pipeline foram definidas as ferramentas utilizadas para cada etapa: filtragem das sequências, mapeamento, conversão de arquivos, análise de SNVs das sequências e por fim as matrizes com os resultados. Além disso, foi configurado o ambiente de execução PBS, permitindo a gestão de recursos e filas de execução de tarefas como número específico de processadores e volume de memória utilizados por cada programa do pipeline. Todas as configurações da execução foram armazenadas em arquivo XML (eXtensible Markup Language) que é lido durante a execução do NASP.

A filtragem de sequências e remoção de adaptadores foi realizada utilizando o Trimmomatic 0.38 (Bolger; Lohse; Usadel, 2014), sendo esta ferramenta a mais utilizada para este propósito em dados de sequenciadores Illumina. Esta filtragem foi baseada na qualidade das sequências utilizando o parâmetro “-phred33” para configurar o codificador base de qualidade, e o parâmetro SLIDINGWINDOW:4:15 foi utilizado para fazer cortes de sequências com a qualidade média do quadro inferior a 15. Foi utilizado também o parâmetro AVGQUAL:30 para cortar leituras cujo a média de qualidade fosse inferior a 30 e o

parâmetro MINLEN:60 para manter apenas sequências com tamanho mínimo de 60pb. Foram utilizados ainda os parâmetros LEADING:10 e TRAILING:10 para remover respectivamente bases com baixa qualidade no começo e no fim das leituras.

O mapeamento das sequências contra o genoma de referência foi realizado utilizando o BWA 0.7.5a (Li; Durbin, 2010), através do algoritmo bwa-mem que é projetado para alinhar sequências de leitura Illumina, sendo este programa escolhido pois é muito utilizado na literatura para organismos com baixa divergência genética e por possuir otimização para execução em larga escala. Em seguida, o NASP executou o programa Samtools 0.1.19-96b5f2294a (Li *et al.*, 2009) para gerar arquivos BAM (Binary Alignment Map), que são utilizados para identificação de SNVs na etapa seguinte. Na sequência de execução do NASP, os arquivos BAM são utilizados pela função HaplotypeCaller do programa GATK 3.2-2-gec30cee (McKenna *et al.*, 2010) para fazer a análise e a chamada de variante (SNVs) das sequências. Por fim, o NASP gera matrizes com os resultados da análise de SNVs, onde pode-se destacar as matrizes chamadas “bestsnp” e “missingdata”. A matriz “bestsnp” armazena os SNVs não duplicados e que passaram por todos os filtros para todos os genomas analisados enquanto a matriz “missingdata” inclui apenas posições que são polimórficas em um subconjunto das amostras (Sahl *et al.*, 2016). Além das matrizes, esses SNVs também são armazenados em arquivos em formato fasta.

7.3.2 Análise das montagens dos genomas de *Y. pestis*

As estatísticas das montagens realizadas foram utilizadas para verificar a qualidade dos dados gerados. Para realizar esta análise foi desenvolvido um *script* em linguagem Shell Script utilizando o GenomeTools V.1.5.8 (Gremme; Steinbiss; Kurtz, 2013), extraindo as estatísticas de todos os 406 genomas montados e agregando em uma planilha.

7.3.3 Filogenia baseada em SNVs

Para a realização da análise de filogenia baseada em SNVs, os arquivos em formato fasta, gerados pelo NASP, nas etapas anteriores com base nos dados das matrizes de “bestsnp” e “missingdata” foram utilizados para estimar o modelo de substituição de nucleotídeos. Esta é uma etapa importante e que pode impactar o resultado da filogenia. Assim, o programa JModelTest 2.1.10 (Guindon; Gascuel, 2003; Posada, 2008) foi utilizado

para prever o modelo de substituição. Este software possui cinco abordagens estatísticas diferentes para prever o melhor modelo de substituição (Posada, 2008). Com o modelo de substituição definido, os arquivos em formato fasta com os dados das matrizes de “bestsnp” e “missingdata” foram utilizados novamente para gerar a árvore filogenética baseada nos SNVs. As árvores foram geradas utilizando a ferramenta PhyML 20160207 (Guindon *et al.*, 2010), que é um programa que utiliza a abordagem probabilística de máxima verossimilhança para realizar a análise filogenética. A robustez dos grupos filogenéticos foi avaliada utilizando a técnica de *bootstrap* com 1000 iterações.

7.3.4 Análise de efeitos dos SNVs

Para analisar os efeitos causados pelos SNVs identificados nos genomas analisados, foi utilizado o programa SnpEff 5.0 (Cingolani *et al.*, 2012). Este programa permite categorizar os efeitos de variantes baseado na localização genômica ou efeitos, por exemplo, como ganho ou perda de códons de iniciação ou finalização. Dado o grande volume de dados gerados, o software MultiQC v1.10 (Ewels *et al.*, 2016) foi utilizado para agregar todos os resultados do SnpEff em um único relatório, facilitando a comparação entre as amostras.

7.4 ABORDAGEM DE MONTAGEM *DE NOVO* E PANGENÔMICA

7.4.1 Montagem *de novo* e anotação dos genomas

Para cada isolado, os dados paired-end filtrados foram usados na montagem de novo dos genomas. A montagem foi realizada utilizando o script *VelvetOptimiser.pl*, desenvolvido para otimização de parâmetros para montagem *de novo* através do programa Velvet (Zerbino; Birney, 2008), sendo o mesmo configurado com função de otimização de N50 para seleção de k-mer.

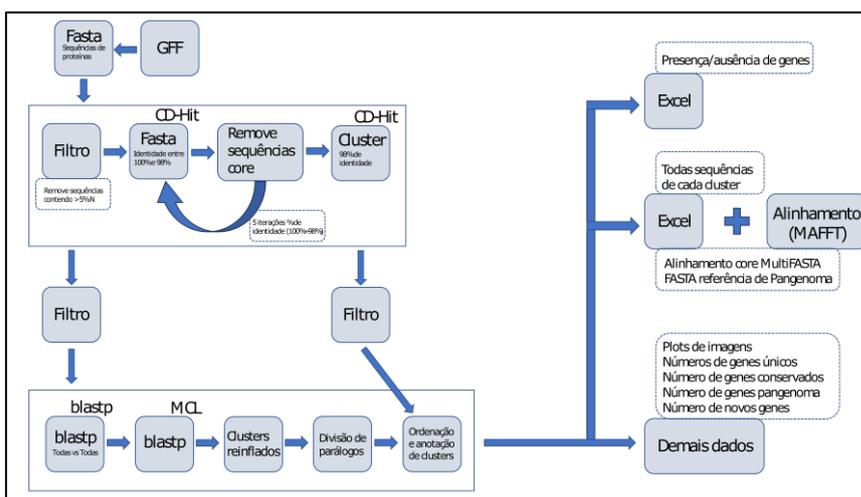
Para fazer as predições gênicas e anotações funcionais dos genomas recém montados, o *pipeline* Prokka (Seemann, 2014) foi utilizado. O Prokka foi escolhido por ser uma ferramenta específica para dados de sequenciamento de genomas de procariotos, além de ser otimizado para execução em larga escala.

7.4.2 Análise pangenômica

Os resultados das anotações realizadas pelo Prokka foram utilizados para realizar a análise de pangenômica, sendo o programa Roary versão 3.11.2 (Page *et al.*, 2015) usado para este propósito. A análise de pangenômica permitiu descrever os genes compartilhados pelos vários isolados, ou genoma *core*, e o Roary foi escolhido por possuir otimização para análise em larga escala de genomas procariotos mesmo com poucos recursos computacionais (Page *et al.*, 2015). A análise do Roary permitiu cobrir a lacuna da abordagem de mapeamento contra um genoma de referência que falha em incorporar no alinhamento as sequências ausentes no genoma de referência, permitindo assim incorporar essas sequências na análise usando apenas montagens *de novo* (Page *et al.*, 2015).

Além dos dados de genoma *core*, o Roary também gera uma matriz de presença e ausência de genes acessórios. Outro resultado também gerado automaticamente pelo Roary é uma árvore filogenética com base nos dados de presença e ausência de genes. A Figura 4 mostra o esquema de funcionamento do Roary nas várias etapas de análise executadas no estudo.

Figura 4 - Esquema de funcionamento do *pipeline* Roary



Fonte: Adaptado de Page *et al.* (2015)

Legenda: A Figura ilustra o funcionamento do Roary. Anotações em formato GFF são passadas como arquivos de entrada e transformados em formato fasta. Uma filtragem é feita para remover sequências com mais de 5% de “N’s” e o CD-HIT faz o agrupamento de proteínas com identidade superior a 98% para o genoma *core* e a clusterização dos demais grupos com identidade menor. A filtragem e clusterização também servem como otimização de recursos computacionais, diminuindo o número de grupos, que são analisados como grupos de genes ortólogos. Em seguida é feito um alinhamento e análise de ortólogos, separação de parálogos e ordenação e anotação de todos os clusters. Por fim são gerados os arquivos de resultados com planilha de presença e ausência de genes, alinhamento do genoma *core* e dados estatísticos.

Por fim, uma segunda árvore foi gerada usando o software FastTree (Price; Dehal; Arkin, 2009) com base nos dados do genoma *core*, utilizando o alinhamento gerado pelo Roary. O FastTree foi escolhido pois é otimizado para análises em larga escala, utiliza a mesma abordagem de máxima verossimilhança do PhyML e possui acurácia comparável com outros softwares incluindo o PhyML (Price; Dehal; Arkin, 2009), possibilitando a execução para o volume de dados aqui analisado.

Para gerar os resultados padrões do Roary de forma visual, foi utilizado o *script* *roary_plot.py* (www.tinyurl.com/7pnmuztk). Este *script* permite utilizar os arquivos de saída do Roary e gerar três imagens, sendo uma para a frequência de genes no genoma, um gráfico em “pizza” para mostrar a distribuição entre genes *core* e acessórios e por fim uma representação de árvore filogenética com uma matriz de presença e ausência de genes.

7.4.3 Análise estatística de presença e ausência dos genes na pangenômica

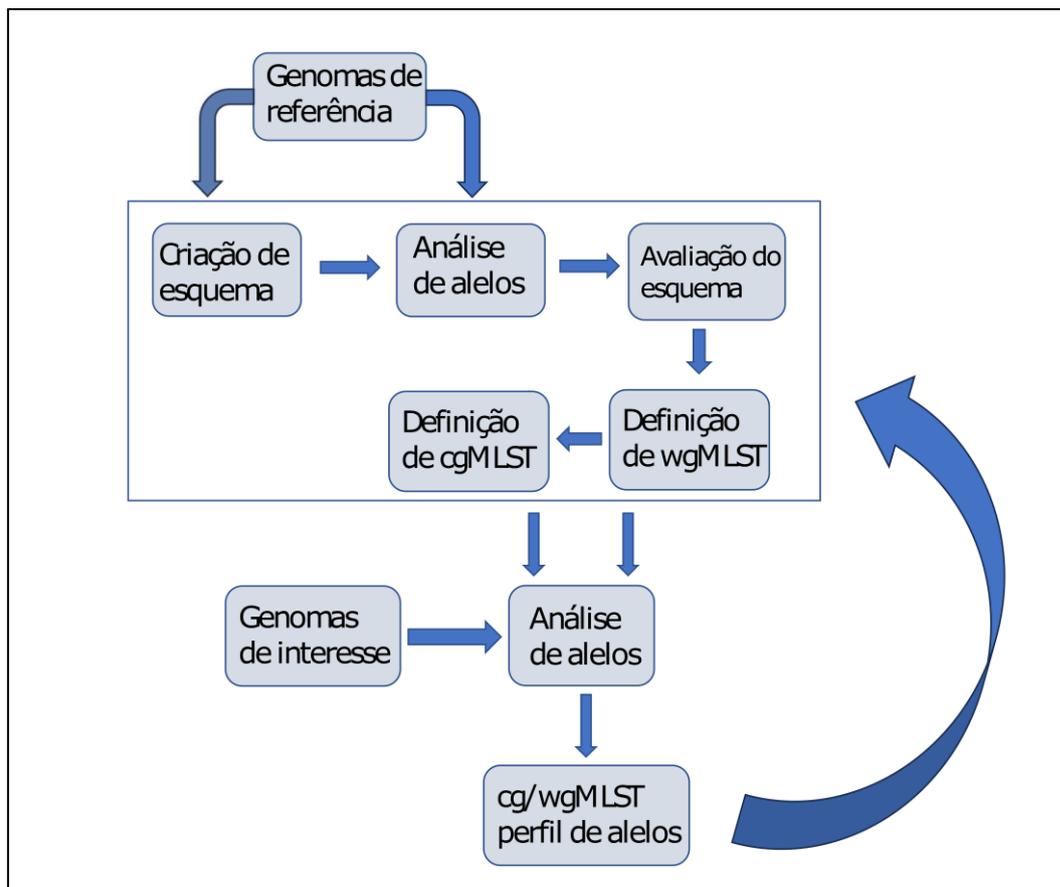
Para fazer a análise estatística dos dados de presença e ausência de genes para as cepas analisadas pelo Roary, o programa Scoary versão 1.6.16 (Brynildsrud *et al.*, 2016) foi utilizado. O Scoary foi desenvolvido para fazer associações entre os dados de genes acessórios produzidos pelo Roary às características binárias que podem ser definidas pelo usuário (Ex. Resistente ou não a um determinado antimicrobiano). O Scoary foi utilizado cruzando os dados da análise de pangenômica do Roary contra fenótipos tratados como dados binários do tipo “pertence ou não pertence” das cepas agrupadas nos clados destacados na árvore filogenética do Roary. Essa análise permite destacar genes presentes ou ausentes que são específicos desses clados, agrupados por relevância estatística. O Scoary faz essa associação entre fenótipos e os dados de pangenoma utilizando o teste exato de Fisher, que é um modelo de teste de significância estatística para análise de tabelas de contingência com distribuição hipergeométrica.

7.5 ANÁLISE GENÔMICA POR cgMLST

Para realizar a análise de gene-a-gene, a abordagem de core genome Multi Locus Sequence Typing (cgMLST) foi utilizada através da ferramenta ChewBBACA versão v.3.2.0 (Silva *et al.*, 2018). Nesta abordagem, um esquema de perfil genético baseado no genoma *core* é criado, seguido por um processo de análise de genes alelos. Primeiro, o genoma de

referência para *Y. pestis* CO92 (Acesso/NCBI: GCF_000009065.1) foi usado para treinar o modelo de predição gênica com a ferramenta Prodigal (Hyatt *et al.*, 2010). Em seguida o esquema de gene-a-gene foi estabelecido usando 60 genomas completos de *Y. pestis* pertencentes aos biovars *antiqua*, *mediaevalis*, *orientalis* e *microtus* disponíveis no banco de dados do *National Center for Biotechnology Information* (NCBI - <https://www.ncbi.nlm.nih.gov/>). Os dados referentes as 60 cepas utilizadas na montagem do esquema estão disponíveis no apêndice B. Uma nova predição gênica é realizada em todas as cepas de referência baixadas do NCBI assim como as 406 cepas do estudo nesta abordagem. Esta etapa foi seguida pela criação do *whole-genome* (wg) *Multi Locus Sequence Typing* (MLST) e do *core genome* (cg) MLST (cgMLST) usando os genomas do presente estudo (n=406) e o conjunto de dados adicional (n=60). A Figura 5 mostra o esquema da abordagem utilizada pela ferramenta Chewbbaca.

Figura 5 - Esquema de funcionamento da ferramenta Chewbbaca



Fonte: Adaptado de Silva *et al.* (2018)

O algoritmo Minimum Spanning Tree (MSTree V2) implementado no *webservice* Grapetree (Zhou *et al.*, 2018) foi usado para visualizar os grupos de loci alélicos gerados por cgMLST.

7.6 ANÁLISE FILOGENÉTICA ENTRE OS ISOLADOS DE *Y. pestis*

Para visualizar essas árvores filogenéticas geradas nas etapas anteriores, foi utilizado o *webservice* iTOL v4 (Letunic; Bork, 2019). Os arquivos em formato newick foram importados pela interface web para que fossem feitas anotações como local, década de coleta e surtos. Para fazer essas anotações, foram criados os seguintes conjuntos de dados com as anotações relativas as informações de coleta de amostras: Município, estado, hospedeiro (pulga, roedor ou humano) e década e surto. Esses conjuntos de informações foram criados visando correlacionar os dados de coletas com a topologia da árvore.

7.7 PESQUISA POR GENES DE RESISTÊNCIA A ANTIMICROBIANOS E FATORES DE VIRULÊNCIA

Para identificar genes de virulência e de resistência a antimicrobianos, foram realizadas pesquisas dos genomas montados das cepas de *Y. pestis* em bases de dados específicas para esses propósitos. Para pesquisa de dados de resistência, os genomas foram pesquisados utilizando os dados da base ResFinder (Bortolaia *et al.*, 2020). Já para a pesquisa de dados de virulência, a base do *Virulence Factors Database* (VFDB) (Chen, 2004) foi utilizada.

Devido ao volume de dados e visando automatização da análise para todas as cepas de *Y. pestis*, a ferramenta ABRicate versão V1.0.0 (<https://github.com/tseemann/abricate>) foi utilizada. O ABRicate é um programa de triagem em massa de *contigs* que foi inicialmente desenvolvido para pesquisa de resistência antimicrobiana e genes de virulência, mas que atualmente permite a pesquisa para outros tipos de dados em outras bases como plasmídeos. Através do ABRicate foi possível baixar as bases de dados descritas anteriormente e fazer um alinhamento local dos genomas com os dados dessas bases, sendo o BLASTN 2.10.1+ do pacote BLAST+ (Camacho *et al.*, 2009) utilizado pelo ABRicate para esta finalidade.

7.8 ANÁLISE DE GENES ANOTADOS COMO HIPOTÉTICOS

Considerando que alguns genes, com relevância estatística dentro dos clados foram previamente anotados como hipotéticos, uma investigação abrangente desse conjunto de genes foi realizada utilizando a ferramenta InterProScan versão V 5.56-89.0 (Jones *et al.*, 2014). Esse método emprega abordagens como o Modelo Oculto de Markov (*Hidden Markov Model* - HMM) e alinhamento local, além de utilizar modelos preditivos - conhecidos como assinaturas - de diversas bases de dados, para conduzir análises funcionais das proteínas. O InterProScan classifica as proteínas em famílias, prevê domínios e identifica elementos funcionais cruciais.

8 CONSIDERAÇÕES ÉTICAS

Os dados obtidos e utilizados ao longo do projeto são provenientes do banco de dados do Serviço de Referência Nacional em Peste (SRP). Não há nas amostras, especialmente às relacionadas aos humanos, qualquer tipo de identificação que possa ser relacionada a um paciente. Não foram utilizados humanos ou animais em nenhuma das etapas do projeto. Todos os programas computacionais usados no projeto, incluindo sistemas operacionais, banco de dados, bibliotecas, *scripts* e demais aplicativos são baseados em programas de acesso livre ou foram desenvolvidos para este trabalho.

9 RESULTADOS

A análise dos genomas montados foi realizada utilizando o programa GenomeTools 1.5.8 e os resultados foram agregados utilizando script desenvolvido para este propósito. Para se obter um panorama geral das montagens, os dados gerados pelo GenomeTools (disponíveis no Apêndice A) foram utilizados para fazer os cálculos das médias de informações básicas dos genomas, conforme pode ser visto na Tabela 1.

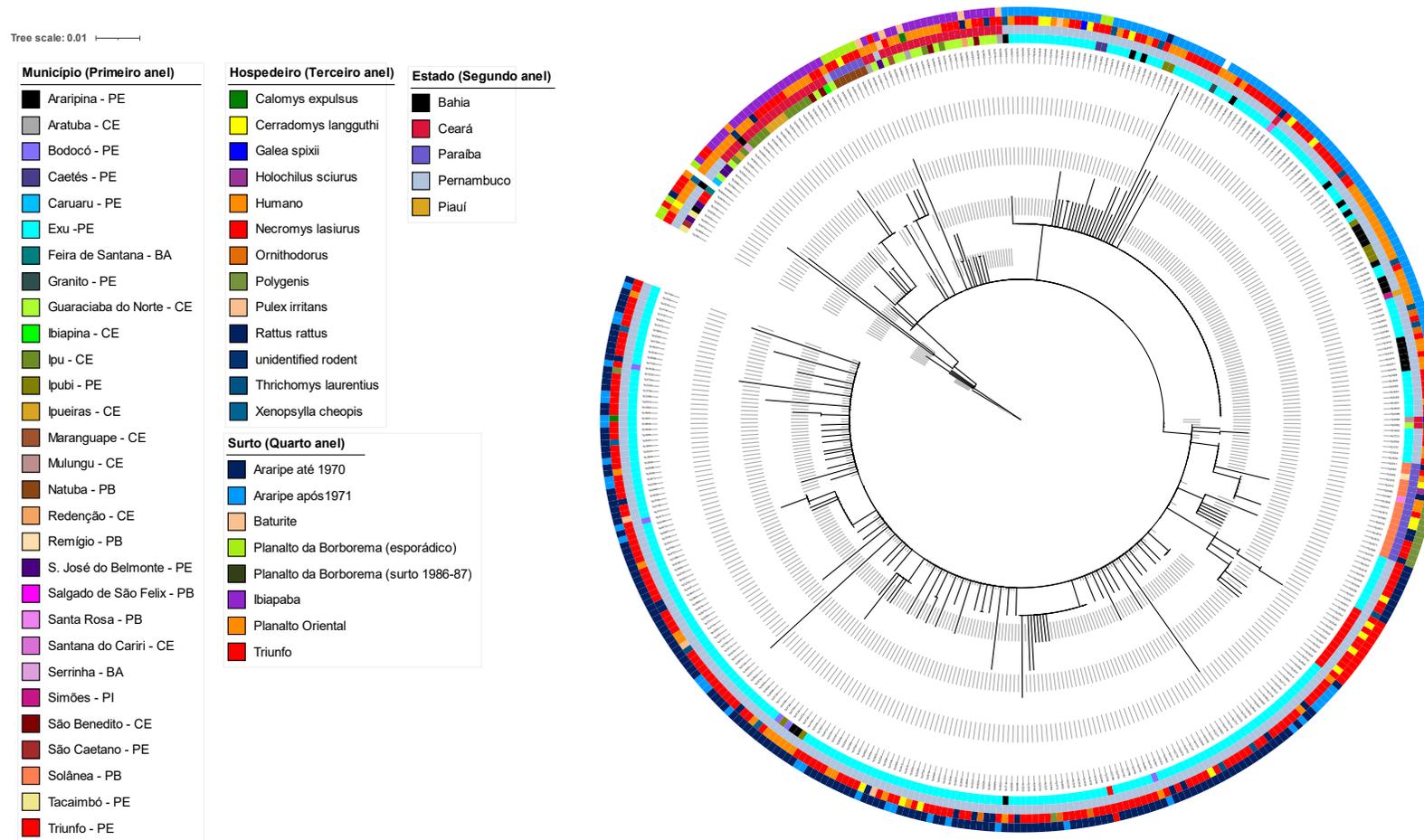
Tabela 1 - Informações gerais dos genomas

Parâmetro	Valor
Número médio de leituras	2.183.187
Tamanho médio do genoma	4.6Mb
N50 médio	367.978
Profundidade média	42x
Conteúdo GC	47%

Fonte: O autor

A partir dos dados obtidos na montagem por referência das cepas contra o genoma de referência CO92 e sua tipagem comparativa de SNVs, foram montadas árvores filogenéticas utilizando a abordagem de máxima verossimilhança e os dados de local de isolamento, hospedeiro, data de coleta e surto, porém não foi possível estabelecer padrões nessas árvores relativos aos dados epidemiológicos. A figura 6 mostra a árvore baseada em SNVs da matriz *bestsnp* enquanto a figura 7 mostra a árvore baseada na matriz *missingdata* geradas pelo *pipeline* NASP.

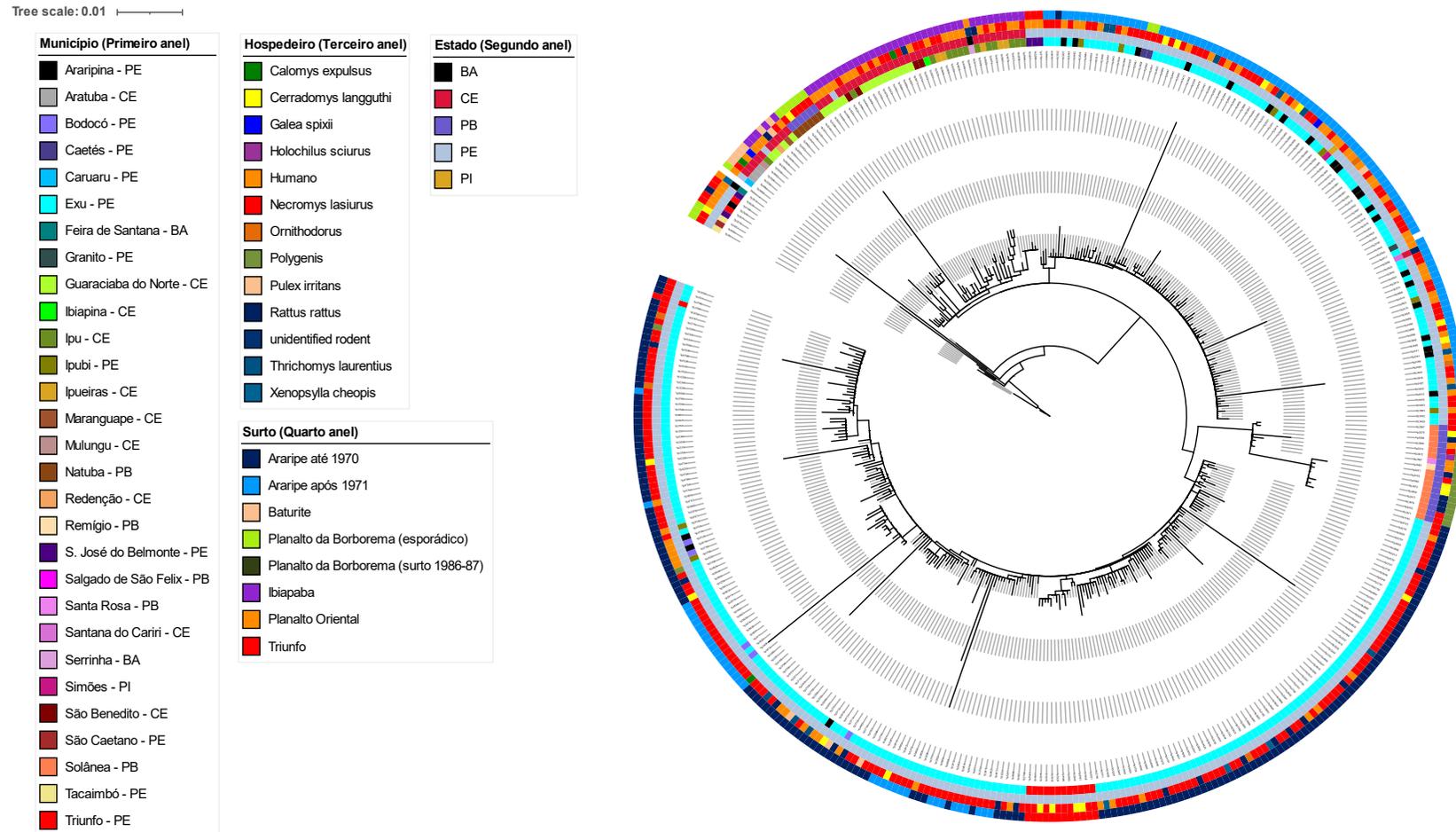
Figura 6 - Árvore baseada em SNVs da matriz de bestsnp



Fonte: O autor.

Legenda: Árvore filogenética da análise de SNVs da matriz bestsnp associada aos dados epidemiológicos. O primeiro anel, do mais interno ao mais externo, representa os municípios de coleta das cepas. Os anéis subsequentes representam os dados epidemiológicos relativos ao estado onde a cepa foi coletada, hospedeiro e o surto, sendo este último uma abstração do local e período de coleta.

Figura 7 - Árvore baseada em SNVs da matriz de missingdata

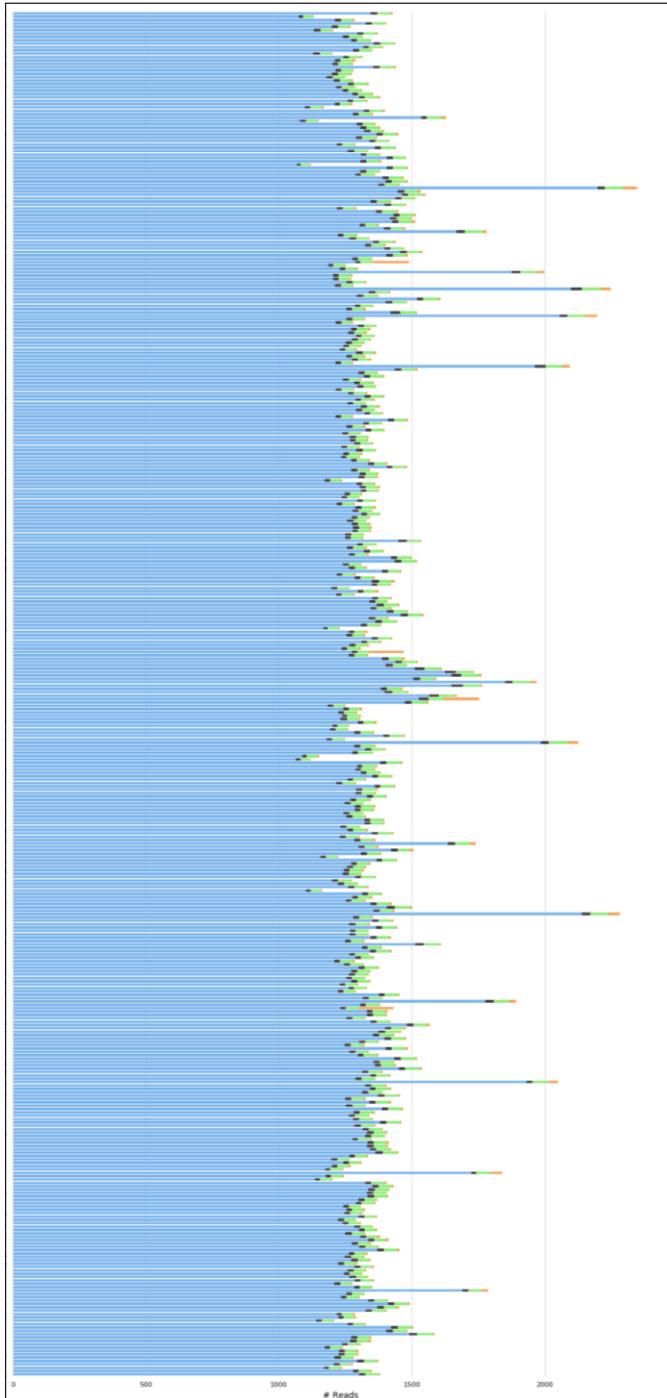


Fonte: O autor.

Legenda: Árvore filogenética da análise de SNVs da matriz missingdata associada aos dados epidemiológicos. O primeiro anel, do mais interno ao mais externo, representa os municípios de coleta das cepas. Os anéis subsequentes representam os dados epidemiológicos relativos ao estado onde a cepa foi coletada, hospedeiro e o surto, sendo este último uma abstração do local e período de coleta.

A análise de efeitos dos SNVs sobre a codificação de proteínas foi realizada através do SnpEff e foi possível observar que algumas cepas possuem um acúmulo de variações bem acima da média conforme mostrado na Figura 8.

Figura 8 - Análise de efeito de SNVs em proteínas



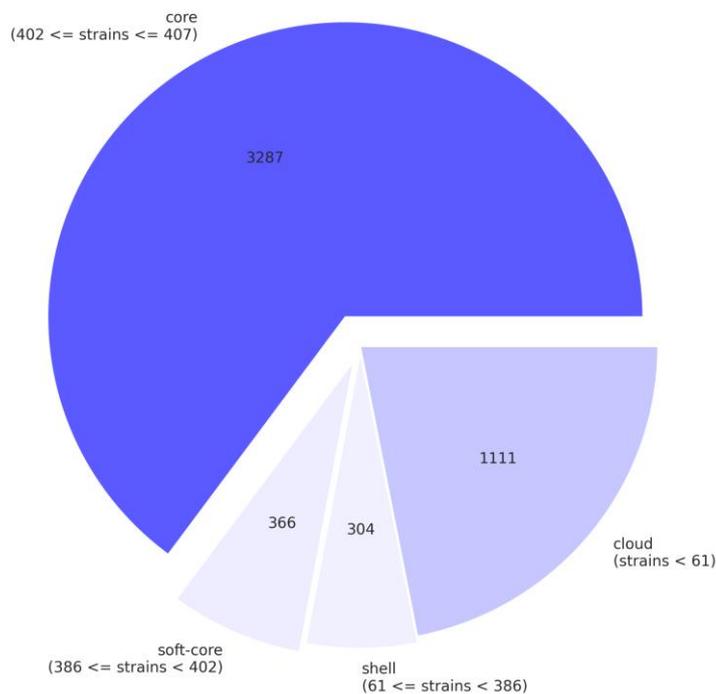
Fonte: O autor

Legenda: Análise dos efeitos dos SNVs na codificação das proteínas. Cada barra representa uma cepa do estudo. O acúmulo de SNVs é mostrado de acordo com o tamanho da barra e as regiões de codificação impactadas pelos SNVs são representados pelas cores das barras. A cor azul representa variantes sem impacto conhecido, a preta são variantes com baixo impacto, a verde com impacto moderado e a laranja com alto impacto.

Foi verificado nos resultados do SnpEff que a média de variantes processadas entre todas as amostras foi de 127 variantes em relação à cepa de referência CO92. A cepa com o maior número de variantes foi a Yp3334, com 217 variantes enquanto a cepa Yp3325 foi a que teve menos variantes, com apenas 102. Os resultados do SnpEff foram ordenados do maior número de variantes para o menor e em seguida foram cruzados com os dados epidemiológicos. Foi possível verificar que todas as cepas atribuídas aos surtos do planalto da Borborema de 1986 e o da Serra de Triunfo entre 1978 e 1979 tiveram o número de variantes acima da média.

Na análise do pangenoma através do Roary foi possível destacar a distribuição de genes entre genes *core* e genes acessórios. A Figura 9 foi gerada usando o script *roary_plot.pl* do Roary e apresenta a distribuição entre genes *core* e genes acessórios.

Figura 9 - Distribuição entre genes do núcleo e acessórios



Fonte: O autor

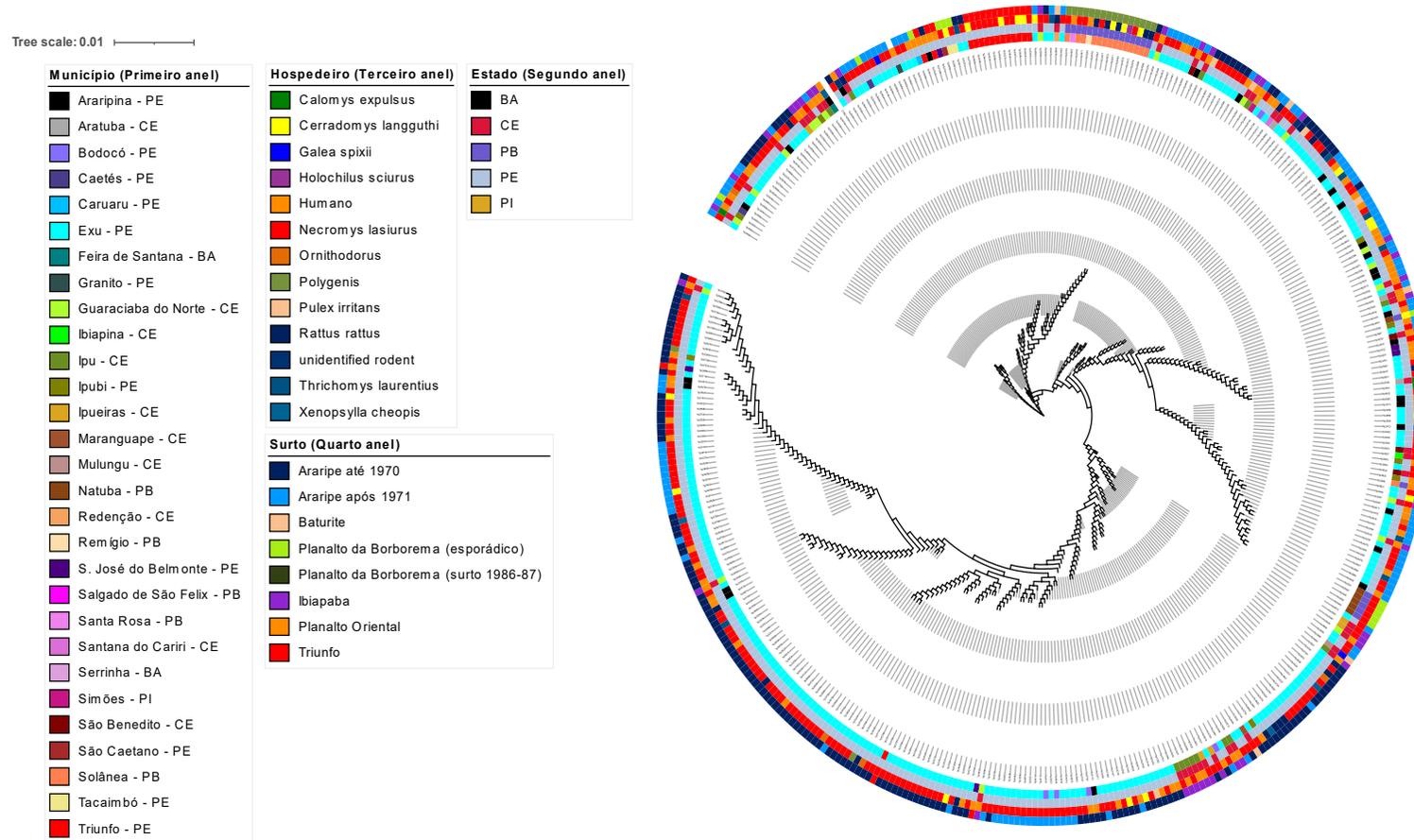
Legenda: O gráfico do tipo pizza gerado pelo *roary_plot.pl* mostra a parcela relativa ao genoma *core* destacado, além de mostrar outras porções de genes acessórios divididos pela quantidade de cepas que possuem esses números de genes acessórios.

Na Figura 9 é possível observar que a parte referente ao genoma *core* é a maior, possuindo 3.287 genes dos 5.068 genes (~65%) analisados na pangenômica.

Com o alinhamento gerado pelo Roary para o genoma *core*, uma árvore filogenética foi montada baseada no genoma *core*. Novamente foi utilizada a abordagem de máxima verossimilhança como feito nas árvores com as abordagens por SNVs e foram adicionados também os dados epidemiológicos. Não foi possível estabelecer padrão entre os dados epidemiológicos e a filogenia baseado no genoma *core* conforme é possível observar na figura 10.

Ainda sobre os dados de pangenoma, foi também montada a árvore filogenética baseada nos dados de presença e ausência de genes, conforme pode ser visto na Figura 11. Na Figura 11 é possível observar que a porção inicial da matriz é densamente azul, representando a porção de genoma *core*. À medida que se desloca na matriz representada na imagem para o lado direito, os pontos azuis ficam menos comuns, representando a porção de genes acessórios. É possível verificar na imagem que alguns clados da árvore se destacam por possuir mais genes acessórios (com mais pontos azuis localizados no fim da matriz) enquanto outros clados se destacam por grandes porções brancas, representando uma ausência de genes comum nesses clados.

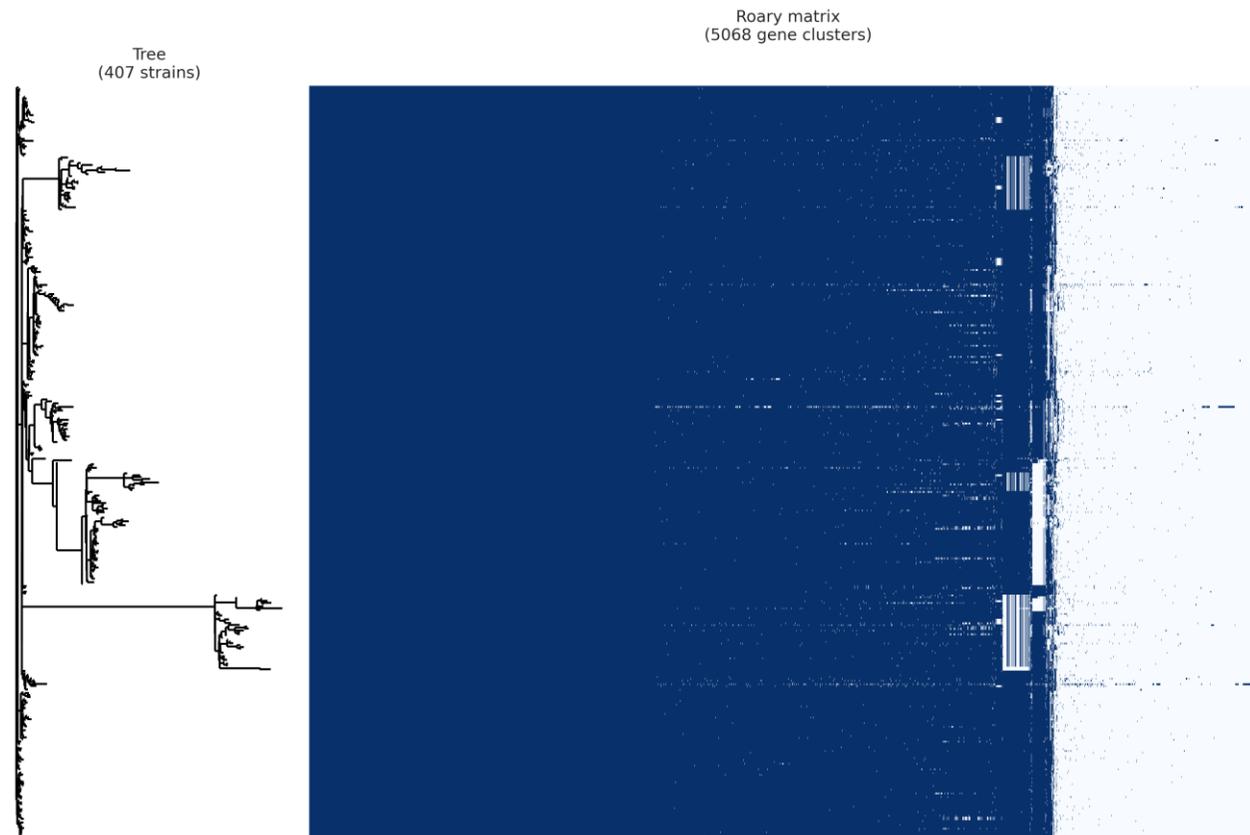
Figura 10 - Árvore baseada no genoma *core*



Fonte: O autor.

Legenda: Árvore filogenética baseada no genoma *core* associada aos dados epidemiológicos. O primeiro anel, do mais interno ao mais externo, representa os municípios de coleta das cepas. Os anéis subsequentes representam os dados epidemiológicos relativos ao estado onde a cepa foi coletada, hospedeiro e o surto, sendo este último uma abstração do local e período de coleta.

Figura 11 - Árvore demonstrando presença e ausência de genes



Fonte: O autor.

Legenda: A figura apresenta a árvore filogenética baseada na presença ou ausência de genes, que está no lado esquerdo. Em seguida é mostrada a matriz de presença e ausência de genes, onde as linhas representam as cepas e as colunas representam os genes. A presença de genes é marcada como um ponto azul, enquanto a ausência é marcada com um ponto branco.

Baseado nessa presença ou ausência de genes, alguns clados com grupos de genes destacados foram nomeados de A à I e foram analisados com o Scoary. A Figura 12 mostra estes clados destacados por cores, apresentando a mesma árvore da Figura 11 mas sem a matriz de presença e ausência de genes.

Figura 12 - Árvore filogenética com clados destacados de acordo com presença e ausência de genes



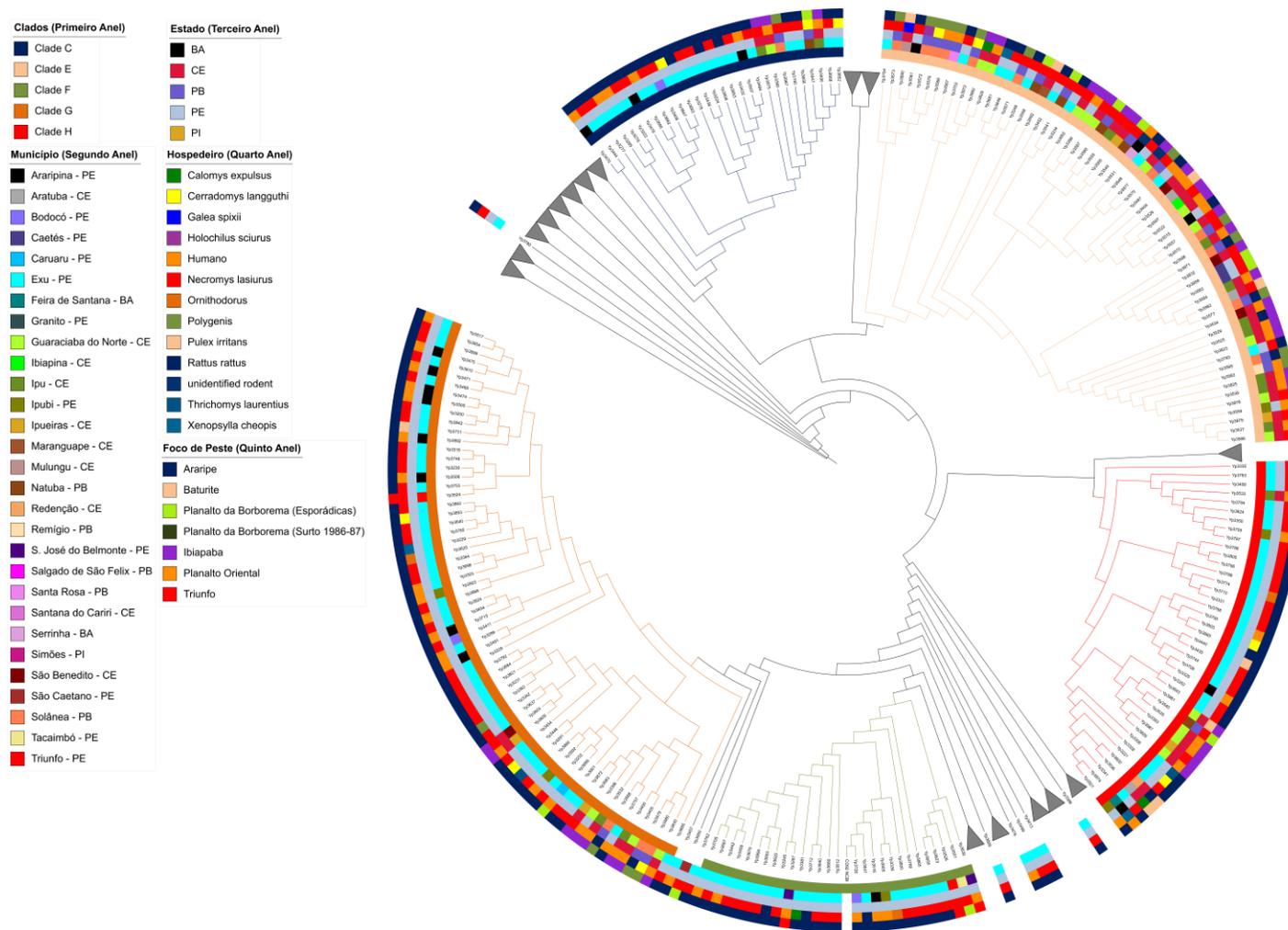
Fonte: O autor.

Legenda: A figura apresenta a árvore filogenética baseada na presença ou ausência de genes gerada pelo Roary. Os clados com grupos destacados por presença ou ausência de genes foram nomeados de cima para baixo em ordem alfabética de A a I. Cada clado destacado foi colorido visando distinguir dos demais clados.

A marcação dos clados, conforme mostrado na Figura 12, permitiu verificar cepas que estão agrupadas por terem presenças ou ausências de genes destacada na matriz de genes da Figura 11. A partir desse recorte na árvore, foi possível selecionar as cepas pertencentes a esses clados e associa-las a uma característica binárias do tipo “pertencer ou não pertencer” a um dado clado, gerando uma matriz binária que foi usada para correlacionar esses clados aos dados de presença e ausência de genes do Roary, através do Scoary.

A árvore filogenética gerada pela análise de pangenoma e baseada no genoma acessório foi integrada com dados epidemiológicos para avaliar a prevalência da presença ou ausência de genes dos clados destacados na etapa anterior com base em fatores como locais de isolamento (município ou estado), hospedeiro (pulga, roedor ou humano) e focos de peste (representando locais e anos de surto ou endemia). No entanto, nenhum padrão discernível pôde ser estabelecido entre os cinco clados destacados com genes estatisticamente relevantes e os dados epidemiológicos. A Figura 13 reproduz a árvore filogenética derivada da análise do pangenoma associada aos dados epidemiológicos

Figura 13 - Filogenia baseada na presença ou ausência de gente associada aos dados epidemiológicos



Fonte: O autor.

Legenda: Árvore filogenética da análise do pangenoma associada aos dados epidemiológicos. O primeiro anel, do mais interno ao mais externo, representa o agrupamento das cepas em cinco clados principais distinguidos por cores diferentes. Os anéis subsequentes mostram a associação entre clados e dados epidemiológicos, incluindo município, estado, hospedeiro e focos de peste.

A análise estatística do Scoary associou 228 genes com relevância estatística nos cinco clados destacados. Dos 228 genes estatisticamente relevantes encontrados nos cinco clados principais, 102 foram inicialmente anotados como proteínas hipotéticas. Após a análise do InterProScan, 49 dessas proteínas hipotéticas foram re-anotadas. No entanto, no caso do clado F, as quatro proteínas hipotéticas permaneceram sem nenhuma anotação gênica funcional. Em seguida, foram analisados todos os 228 genes em busca de fatores de virulência usando o banco de dados VFDB através da ferramenta Abricate, o que resultou em 57 genes relacionados a fatores de virulência dentro dos clados destacados. A Tabela 2 resume a distribuição dos genes de virulência putativos e confirmados e dos genes não associados à virulência nos clados C, E, F, G e H.

Tabela 2 - Genes Putativos de virulência e demais genes nos clados destacados

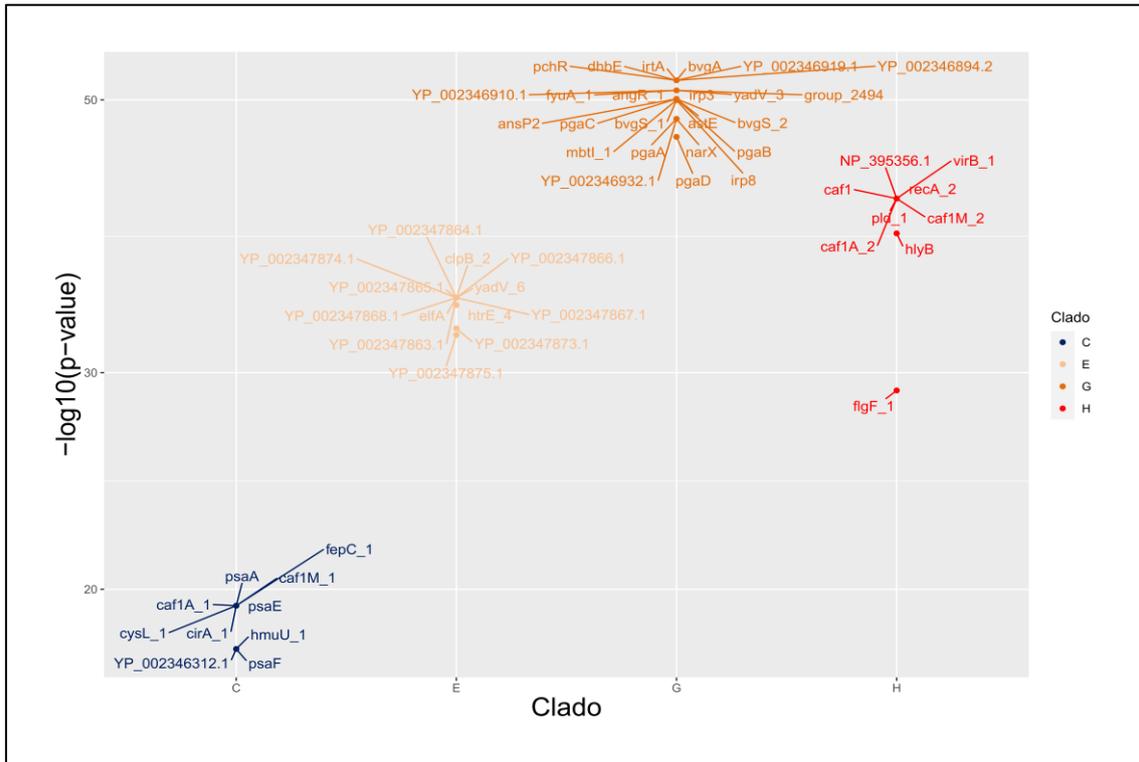
Clado	Genes Putativos de Virulência	Outros Genes
C	10	43
E	13	0
F	0	4
G	25	41
H	9	83

Fonte: O autor

Foi realizada também a pesquisa por genes de resistência a antimicrobianos. Para esta análise todos os genomas foram analisados novamente com o Abricate utilizando a base do Resfinder, porém, nenhum gene de resistência foi encontrado em nenhuma das cepas analisadas nos estudos.

A análise do Scoary gerou dados para os clados destacados associando a relevância estatística de um dado gene ao pertencimento ou não de um clado. A Figura 14 apresenta os genes relacionados à virulência enriquecidos como ausentes através da análise de Scoary. Esses genes ausentes são agrupados por clado e organizados com base em seus *p-valores* correspondentes.

Figura 14 - Fatores de virulência agrupados por clado e p -valor

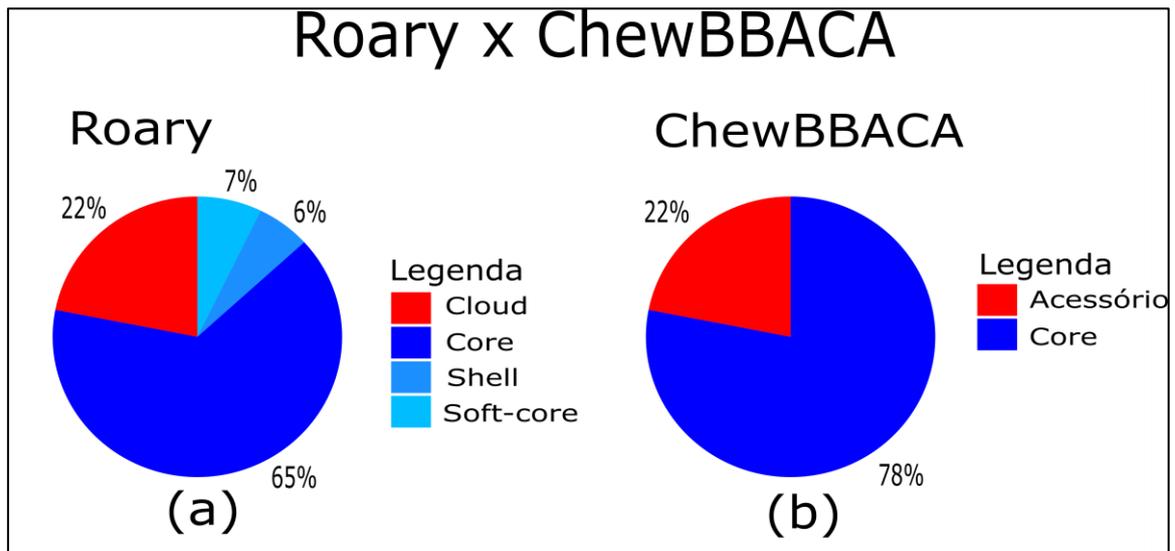


Fonte: O autor.

Legenda: A figura mostra os genes associados a fatores de virulência com relevância estatística de ausência, nos clados destacados. Os genes estão agrupados por clado destacado na análise de pangenoma e $\log_{10}(p\text{-valor})$. Cada um dos clados principais é representado por uma cor diferente.

A análise de cgMLST gerou novas anotações gênicas que resultaram em um conjunto de 4088 genes entre as cepas, dos quais 3206 estão no genoma *core* (78,42% do genoma total) e 882 genes no genoma acessório (21,58% do genoma total). A comparação da distribuição gênica entre a abordagem pangenômica usando Roary e cgMLST usando ChewBBACA é mostrada na Figura 15.

Figura 15 - Comparação da distribuição de genes entre Roary e ChewBBACA

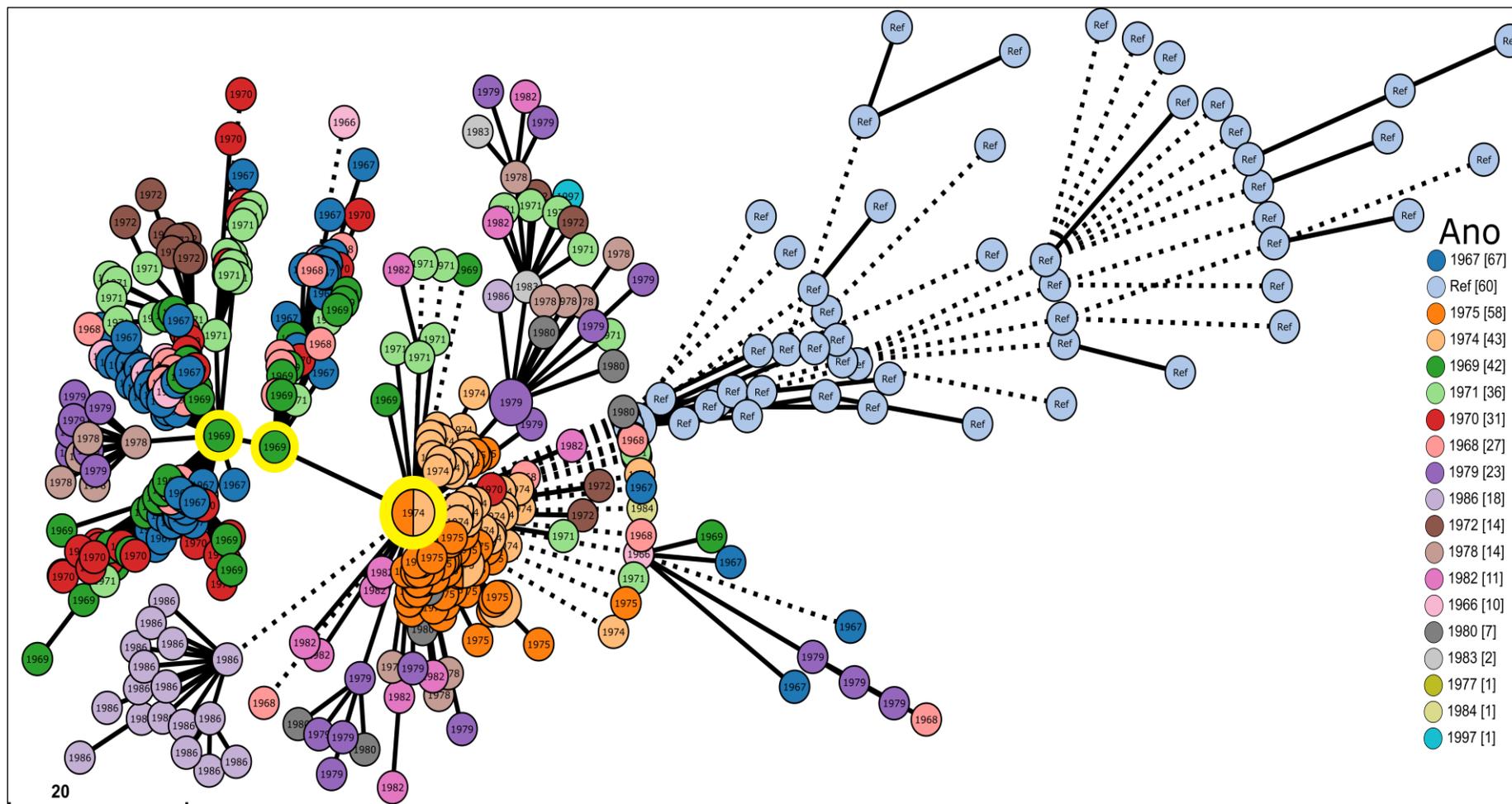


Fonte: O autor.

Legenda: A Figura mostra que a maior porção de genes está no genoma *core* com 65% (3287 genes). Os genes restantes são ainda classificados dentro do genoma acessório compreendendo 7% (386 <= cepas <= 402) no *soft-core*, 6% (61 <= cepas <= 386) no *shell* e 22% (cepas <= 61) no *cloud* (cepas <= 61). A Figura 15b mostra a maior porção de genes no genoma *core* com 78% (3206 genes).

O esquema de perfil genético gerado, seguido da análise gene-a-gene com base no genoma *core*, produziu um arquivo de estrutura de grafo usando a abordagem de *minimum spanning tree* que foi cruzada com os dados epidemiológicos na forma de metadados no *webservice* Grapetree. O grafo resultante de acordo com a data de coleta das cepas está representado na Figura 16.

Figura 16 - Agrupamentos de cepas baseados na análise gene-a-gene, por ano



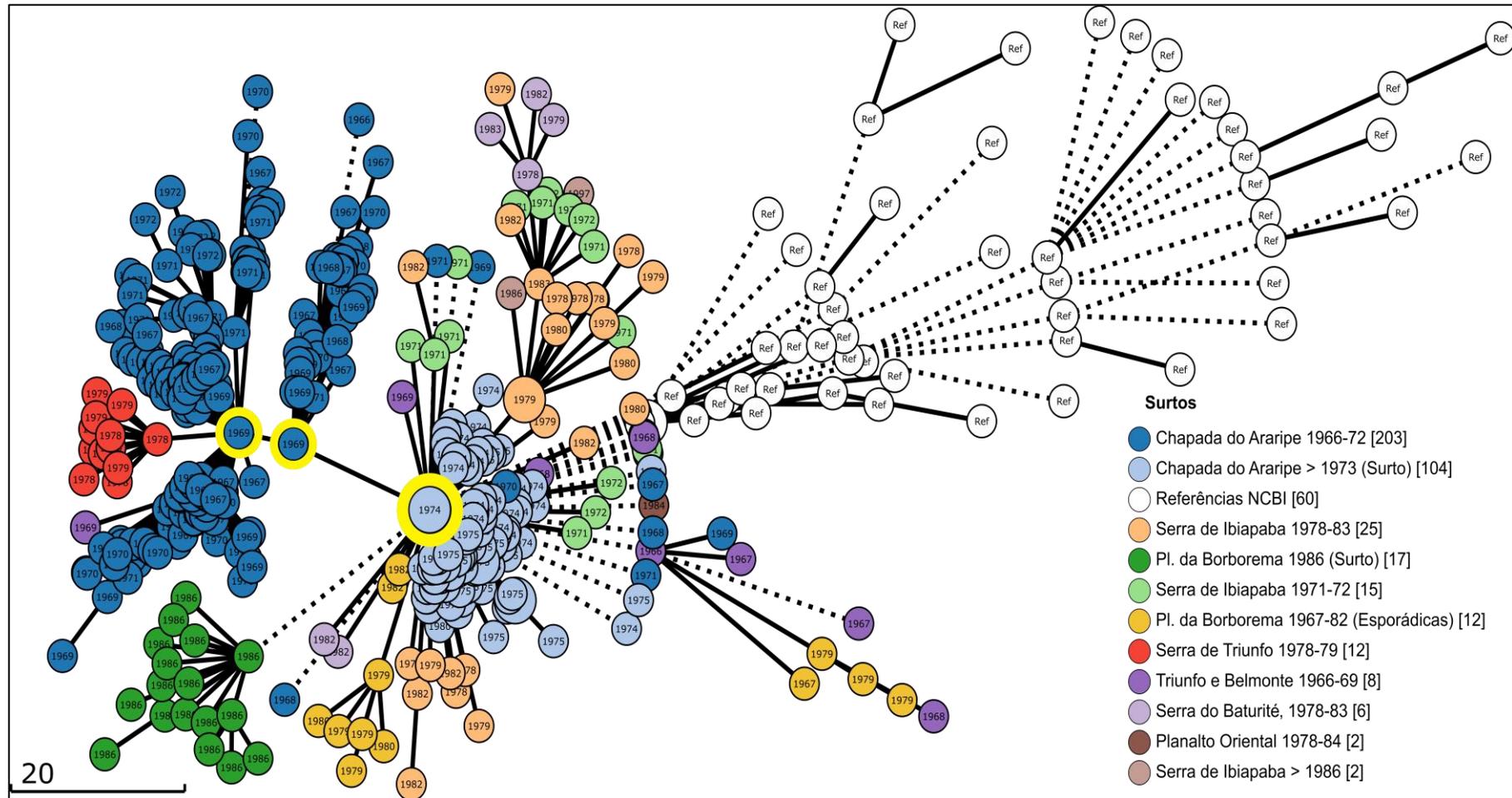
Fonte: O autor.

Legenda: Os grafos representados demonstram agrupamentos de cepas com base na análise gene-a-gene. Cada nó colorido representa o agrupamento de cepas de acordo com a similaridade alélica associada ao ano de coleta que melhor representa o agrupamento. A similaridade genética é representada pelo tamanho das arestas. Três nós principais de cepas que representam agrupamentos dos quais outros grupos de cepas geneticamente relacionadas derivam estão destacados com um círculo amarelo. O número entre colchetes ao lado do ano indica o número de cepas correspondente ao ano específico. O tamanho de cada nó é proporcional ao número de cepas agrupadas

É possível observar três nós principais dos quais derivam grupos de cepas geneticamente relacionadas (nós circulados em amarelo). Os dois primeiros, à esquerda da imagem, são caracterizados principalmente por cepas do ano de 1969. O terceiro é o maior entre os três principais nós, contendo um maior número de cepas. A maioria dessas cepas é originária dos anos de 1974 e 1975, que é representada por duas cores diferentes. Também é possível observar um pequeno grupo de nós muito característico do ano de 1986, colorido em roxo, derivado do maior nó principal. Além do ano de coleta foram realizadas análises com o grafo de cgMLST associando a outros dados epidemiológicos como local de coleta, estado de coleta e hospedeiro, porém nessas análises não foram encontrados padrões discerníveis.

Outra análise realizada para associar o grafo resultante com outros dados epidemiológicos utilizou surto de peste como parâmetro, que engloba o local de coleta e a data dos casos positivos de *Y. pestis*. A imagem resultante está representada na Figura 17. Também resultou em três nós principais como pontos de evolução (também circulados em amarelo), agora relacionados a dois surtos na Chapada do Araripe. A partir deste primeiro surto, um subgrupo de cepas parece autolimitado a um curto período entre 1966 e 1971. Outro subgrupo derivado está relacionado ao surto na Serra de Triunfo (colorido em vermelho). Destaca-se que, do maior nó principal, dois subgrupos de cepas estão ligados a surtos na Chapada da Borborema, um de 1986 está colorido em verde (e corresponde aos grupos anteriormente coloridos em roxo na figura 16) e outro colorido em amarelo com cepas coletadas entre 1979 e 1982.

Figura 17 - Agrupamentos de cepas baseados na análise gene-a-gene em relação aos surtos



Fonte: O autor.

Legenda: A figura mostra agrupamentos de cepas com base na análise gene-a-gene. Cada nó colorido representa o agrupamento de cepas de acordo com a similaridade alélica associada a surtos de *Y. pestis* no Brasil e são marcados com o ano de coleta que melhor representa o agrupamento. A similaridade genética é representada pelo tamanho das arestas. Três nós principais de cepas que representam agrupamentos dos quais outros grupos de cepas geneticamente relacionadas derivam estão destacados com um círculo amarelo. O número entre colchetes ao lado do nome do surto indica o número de cepas correspondentes ao surto específico. O tamanho de cada nó é proporcional ao número de cepas agrupadas.

10 DISCUSSÃO

Os estudos de pangenoma e gene-a-gene mostraram a maior concentração de genes no genoma *core*, o que é um achado esperado já que *Y. pestis* é conhecida por seu genoma conservado. Mesmo as cepas modernas de *Y. pestis* exibem diferenças mínimas em relação às cepas da segunda pandemia, particularmente essas diferenças se concentram nos fatores de virulência (Barbieri *et al.*, 2020). A proporção de genes no genoma *core* foi compatível entre as análises de Roary e ChewBBACA. Ao incorporar as subdivisões do Roary para o genoma *core* (*soft-core* e *shell*) ao genoma *core*, observamos que 78% dos genes pertencem ao genoma *core*. Essa proporção se alinha com as descobertas do ChewBBACA, que também produziu uma composição do genoma *core* de 78%.

O número de genes anotados *in silico* no presente estudo também é consistente com outros estudos de *Y. pestis* (Deng *et al.*, 2002), e até mesmo com um estudo que utilizou abordagens mais tradicionais em laboratório (Zhou *et al.*, 2004). Outra observação interessante na análise do pangenoma é a árvore filogenética resultante com a matriz de genes mostrada na Figura 11, onde é possível observar claramente a proporção do genoma *core* e dos genes acessórios.

Também é possível verificar que alguns clados estão bem destacados na matriz gênica por grupos de genes presentes ou ausentes em relação às outras cepas. Analisando a Figura 13, que correlaciona os dados filogenéticos com os conjuntos de dados epidemiológicos, não é possível observar padrões discerníveis entre os perfis gênicos e os locais de coleta, período, hospedeiro ou surto. O melhor exemplo disso é encontrado no clado E, onde há a maior mistura de cores representando diferentes dados epidemiológicos. O clado F é o mais próximo de ter um padrão, uma vez que uma parcela significativa das amostras está associada aos focos de peste da Chapada do Araripe, mas mesmo neste caso há cepas diferentes. Essa observação da falta de padrão entre a filogenia e os dados epidemiológicos é corroborada por outro trabalho que traçou a rota de entrada da *Y. pestis* no Brasil (Vogler *et al.*, 2019).

Os genes analisados pelo VFDB retornaram vários fatores de virulência, incluindo importantes genes bem conhecidos em *Y. pestis*, como aqueles envolvidos na formação de biofilme e aquisição de ferro (Demeure *et al.*, 2019; Hinnebusch; Erickson, 2008). Algumas das proteínas hipotéticas re-annotadas com bancos de dados do InterProScan também foram associadas a fatores de virulência, incluindo genes de aquisição de ferro e adesão celular.

Os agrupamentos gênicos retornados pelo Scoary nos clados C, E, G e H são caracterizados pela ausência de genes. O clado C mostrou a ausência de vários genes associados a importantes fatores de virulência para *Y. pestis*, tais como aquisição de ferro, formação de biofilme e adesão celular. Um notável exemplo disso é a ausência dos genes *psaA*, *psaE* e *psaF* nas 29 cepas pertencentes ao clado C. Em contraste, entre as 378 cepas de outros clados, apenas 12 cepas não possuem *psaA* e *psaE*, enquanto *psaF* está ausente em 13 dessas 378 cepas. O gene *psaA*, responsável pela produção do antígeno pH 6, desempenha um papel crucial no processo infeccioso, facilitando a adesão celular às células epiteliais. Além disso, o gene *psaA* tem a função crítica de inibir a fagocitose por macrófagos, auxiliando na evasão da resposta imune do hospedeiro pela *Y. pestis* (Bao *et al.*, 2013; Demeure *et al.*, 2019; Liu; Chen; Galván, 2006). Estudos recentes têm revelado o importante papel regulatório da *psaE* e *psaF* na transcrição do gene *psaA*. A expressão de *psaA* é dependente de condições específicas de temperatura e pH, e essa dependência é mediada por efeitos pós-transcricionais distintos sobre *psaE* e *psaF* (Quinn *et al.*, 2019).

Outro aspecto importante observado no clado C é a ausência de alguns genes associados à aquisição de ferro. A aquisição de ferro desempenha um papel importante como fator de virulência, pois *Y. pestis* necessita capturar ferro para seu crescimento, multiplicação e funções biológicas (Demeure *et al.*, 2019; Fetherston *et al.*, 2010; Hood; Skaar, 2012). Notadamente, alguns estudos têm demonstrado que a aquisição reduzida de ferro pode levar à atenuação da virulência (Fetherston *et al.*, 2010; Zauberman *et al.*, 2017). Genes ligados à aquisição de ferro em *Yersinia* como *cirA*, *fepC* e *hmuU* (Rakin *et al.*, 1994; Schubert; Fischer; Heesemann, 1999; Thompson; Jones; Perry, 1999) estão ausentes em todas as 29 cepas do clado C. Outro achado interessante entre as cepas do clado C é que o gene hipotético YP_002346312.1, foi associado ao PFAM PF03692 após análise do InterProScan e é possivelmente um sítio relacionado à aquisição de ferro ou zinco (Burnens *et al.*, 1997; Ren *et al.*, 2005). A ausência de genes relacionados à aquisição de ferro pode afetar profundamente a virulência bacteriana (Fetherston *et al.*, 2010; Zauberman *et al.*, 2017).

Uma ausência inesperada observada no clado C foram os genes *cafIA* e *cafIM* que fazem parte de um *operon* juntamente com *cafI*. O gene *cafI* é considerado específico para *Y. pestis*, sendo utilizado até mesmo para fins diagnósticos (Demeure *et al.*, 2019). Este *operon* desempenha um importante papel na virulência em *Y. pestis*, uma vez que é responsável pelo antígeno F1 que inibe a fagocitose por macrófagos (Al-Jawdah *et al.*, 2019; Du; Rosqvist; Forsberg, 2002). Mas outros trabalhos já demonstraram a ausência desses genes,

possivelmente devido à instabilidade do plasmídeo pFra (Leal-Balbino *et al.*, 2004). Outro ponto que explica essa ausência é o potencial de deleção espontânea do gene em subculturas repetidas, como demonstrado por Almeida e colaboradores (Almeida *et al.*, 1993) com o gene *irp2*. Vale ressaltar que algumas dessas cepas foram armazenadas por um longo período, datando da década de 1960 (Almeida *et al.*, 1993; Almeida *et al.*, 1994).

O gene *cysL*, um regulador transcricional lysR-like também está ausente no clado C. Este gene é descrito na literatura com um papel na formação de biofilme em *Y. pestis* e em outros organismos (Guillouard *et al.*, 2002; Vadyvaloo; Hinz, 2015). Este gene está ausente em todas as 29 cepas do clado C e em apenas 12 entre as 378 cepas de outros clados. A ausência desses genes no clado C pode afetar significativamente a virulência dessas cepas em múltiplos aspectos. A adesão celular desempenha um papel crucial em vários mecanismos de virulência, incluindo o Sistema de Secreção Tipo III (T3SS). Assim, a ausência desses genes afetando a adesão celular pode levar a uma virulência atenuada (Bohn *et al.*, 2019). A ausência de fatores associados à formação de biofilme tem um impacto significativo sobre um importante mecanismo de transmissão de bacilos pelas pulgas. Esse mecanismo depende da formação de biofilme e sua interrupção pode levar a uma diminuição nas taxas de transmissão do bacilo a partir das pulgas.

Para o clado E, os principais fatores de virulência ausentes foram relacionados à adesão celular e ao Sistema de Secreção Tipo 6 (T6SS). Um dos genes ausentes relacionados à adesão celular é o *htrE*, que é homólogo ao gene *papC* presente em *Y. pestis* (Ford *et al.*, 2010; Raina *et al.*, 1993). Outro exemplo é o *yadV*, anteriormente conhecido como *ecpD*, que é um gene homólogo ao gene *papD* presente em *Y. pestis* (Raina *et al.*, 1993). O T6SS é amplamente distribuído entre bactérias Gram-negativas e funciona como um sistema de liberação de várias proteínas patogênicas. Em *Y. pestis*, sabe-se que o T6SS possui múltiplas cópias (Alam *et al.*, 2021; Yang *et al.*, 2018). O gene *clpB*, que é primariamente reconhecido por suas funções de resposta ao estresse, mostrou-se ausente no clado E, e estudos recentes também relataram seu envolvimento como efetor dentro do T6SS (Alam *et al.*, 2021). Foi possível re-anotar todos os nove genes do clado E, inicialmente anotados como hipotéticos, através da análise do InterProScan e relacioná-los ao T6SS. Todos esses genes estão ausentes em todas as 62 cepas deste clado, exceto para o PFAM PF06812, que está presente em uma cepa. Como mencionado anteriormente em relação ao clado C, genes associados à adesão celular podem influenciar a virulência das cepas do clado E. Considerando que o T6SS é um sistema de secreção, é provável que a ausência de genes relacionados ao T6SS também afete a

virulência das cepas do clado E. No entanto, devido ao limitado conhecimento sobre esse sistema de secreção em *Y. pestis*, torna-se desafiador medir com precisão a extensão desse impacto.

O clado G exibiu o maior número de genes de virulência ausentes, particularmente genes relacionados à formação de biofilme, adesão celular e aquisição de ferro. Notavelmente, a ausência de genes da família *pga* (*pgaA*, *pgaB*, *pgaC* e *pgaD*) destacou-se nos resultados do Scoary para este clado. Esta família de genes é reconhecida por seu papel crucial na manutenção e estabilidade estrutural do biofilme em várias eubactérias (Pelludat *et al.*, 1998). Além disso, *pga* é semelhante ao gene *hms*, específico de *Y. pestis*, e conhecido pela síntese da matriz extracelular (ECM - extracellular matrix) na formação de biofilme (Hinnebusch; Erickson, 2008; Itoh *et al.*, 2008).

Outro gene de formação de biofilme ausente neste clado é o *NarX* (Mangalea; Borlee, 2022), embora seja mais conhecido por seu papel regulatório no controle da expressão da resposta ao nitrato (Cavicchioli *et al.*, 1995). Em *Yersinia*, esses genes, juntamente com outros genes da família *irp*, desempenham um papel significativo na síntese de yersiniabactin (Pelludat *et al.*, 1998). Estudos têm demonstrado que a deleção do gene *irp8* leva à atenuação da virulência da peste pneumônica (Pechous *et al.*, 2015; Pelludat *et al.*, 1998). Outro gene ausente que tem papel regulatório na yersiniabactin foi o *FyuA*. Um estudo recente demonstrou a diminuição da sobrevivência da bactéria em macrófagos devido à disfunção do gene *fyuA* (Chen *et al.*, 2022). Finalmente, a análise InterProScan atribuiu os acessos PFAMs PF01258 e PF10080 a dois genes inicialmente hipotéticos, um associado à aquisição de ferro e outro ao transporte de ferro, respectivamente. Outros genes ausentes associados à aquisição de ferro foram *irtA*, *dhbE*, *angR*, *pchR* e *mbtI* e juntas todas essas ausências podem afetar a virulência de *Y. pestis*. Vários genes relacionados à adesão celular, nomeadamente *htrE* e *yadV*, mostraram-se ausentes. Curiosamente, esses genes também estavam ausentes no clado E. Outros genes incluem os reatribuídos pelo InterProScan nos PFAM's PF05930, PF00419, e Superfamília SSF49478.

Em todas as 66 cepas do clado G, os genes relacionados à virulência discutidos estão ausentes, exceto para o gene *htrE*, que está presente em apenas uma cepa dentro deste clado. Em contraste, nos outros clados, esses genes estão ausentes em apenas 10 cepas coletivamente.

No clado H, *cafI*, *virB* e *pld* estão ausentes. A ausência de *virB* necessita de atenção dada a sua importância na composição do T3SS. O gene *pld*, também referido como toxina murina de *Yersinia* (*ymt*), é outro gene importante para *Y. pestis* e considerado uma adaptação genética que permitiu a transmissão do bacilo via pulgas, desempenhando um papel importante na sua sobrevivência no intestino médio do inseto (Chouikha; Hinnebusch, 2014; Demeure *et al.*, 2019; Hinnebusch *et al.*, 2002).

Outros genes ausentes no clado H incluem *hlyB* e *flgF*. O gene *hlyB* faz parte do Sistema de Secreção Tipo 1 (T1SS) comumente presente em cepas virulentas de Enterobacteriaceae e descrito como tendo composição mais simples quando comparado a outros sistemas de secreção (Kanonenberg; Smits; Schmitt, 2019). O gene *flgF*, reconhecido por seu papel na motilidade, também tem sido implicado na formação de biofilme em alguns estudos quando associado a outros genes comuns encontrados em *Y. pestis*, como *motA* e *cheD* (Singh *et al.*, 2018; Xu *et al.*, 2021). Finalmente, o InterProScan reanotou um gene hipotético atribuindo o PFAM PF02368, que está associado à adesão celular.

Todos esses genes estão ausentes em todas as cepas do clado H, exceto o gene *flgF*, que está presente em 6 das 41 cepas. O clado H, entretanto, carece de um gene relacionado ao T1SS que também pode comprometer a expressão de outros fatores de virulência. Em conjunto, a ausência desses genes de virulência em todos os principais clados destacados corrobora com trabalho anterior inferindo atenuação de virulência em cepas de *Y. pestis* do nordeste do Brasil (Karimi *et al.*, 1974).

A pesquisa por genes de resistência a antimicrobianos não retornou resultados para nenhum dos genomas analisados neste estudo. Este achado era esperado dado que genes de resistência a antimicrobianos não são comuns em *Y. pestis*. Um recente artigo de revisão levantou outros estudos sobre resistência a antimicrobianos, e mostrou que as cepas da América do Sul são altamente susceptíveis à antibióticos (Lei; Kumar, 2022). Apesar da ausência de genes de resistência conhecidos nas cepas do Brasil, é importante manter a pesquisa por esses genes pois alguns casos de resistência para *Y. pestis* começaram a aparecer a partir de 1987, principalmente em Madagascar na África. Três cepas foram verificadas como resistentes em Madagascar e uma na China à estreptomicina devidos a uma mutação no gene *rpsL* (Dai *et al.*, 2021). Foram reportadas também resistência por meio de transferência horizontal em Madagascar tanto a estreptomicina quanto doxiciclina (Lei; Kumar, 2022). Existe ainda um caso de cepa multirresistente a antimicrobianos também adquirida por meio de transferência horizontal do plasmídeo IP1202.

Através das análises gene-a-gene e de SNVs, dados genômicos e epidemiológicos foram associados. Ao contrário da abordagem filogenética, que considera apenas uma relação dicotômica entre uma cepa e outra, o perfil gerado no cgMLST permitiu que várias cepas fossem agrupadas no mesmo nó, possibilitando assim inferir similaridade genética com base no acúmulo de diferenças alélicas. Observando os dados epidemiológicos, a arquitetura do grafo resultante associado ao ano de coleta (Figura 16) apresenta três nós principais dos quais divergiram outros grupos de cepas geneticamente relacionados. Essa observação é consistente, uma vez que no Brasil as primeiras coletas são do ano de 1966 e vários casos surgiram até o ano de 1972, que estão representados em dois dos principais agrupamentos. Após um breve silêncio epidemiológico, vários casos surgiram nos anos de 1974 e 1975, representados pelo maior nó principal. Observou-se também que as cepas de referência, obtidas em banco de dados público e não originárias do Brasil, são geneticamente distintas da coleção brasileira de cepas. Essa observação evidencia que as cepas brasileiras passaram por uma evolução genética própria.

O grafo resultante associado aos surtos no Brasil (Figura 17) também destacou os três principais clusters, agora relacionados aos dois surtos ocorridos na Chapada do Araripe, uma grande região entre três estados do nordeste brasileiro. Dois desses nós principais pertencem ao primeiro surto (1966 a 1972) e têm dois subgrupos derivados. Um dos nós deriva para uma linhagem de cepas aparentemente autolimitadas, já que quase todas são do município de Exu e apenas três cepas são de outros municípios muito próximos entre si. Além disso, todas essas cepas são de um curto período entre 1966 e 1971. Já o segundo nó derivado apresenta todas as cepas relacionadas ao surto ocorrido no município de Triunfo, Pernambuco, entre 1978 e 1979 (grupo de nós vermelhos na figura 17). Corroborando com este achado, a análise de SNVs mostrou que todas as cepas do surto de Triunfo têm um acúmulo de variações acima da média, conforme os picos de variações mostrados na Figura 8.

A partir do segundo surto na Chapada do Araripe (iniciado em 1973) dois subgrupos foram derivados e ocorreram na região da Chapada da Borborema. Mas, apesar da Chapada da Borborema abranger 5 estados do nordeste do brasileiro, todas essas cepas são da Paraíba. O primeiro subgrupo tem cepas de 1979 a 1982 e são do município de Natuba, enquanto o outro subgrupo todas as cepas são de 1986 e estão principalmente no município de Solânea. Novamente, as cepas desse grupo foram verificadas com acúmulo de variações acima da média na análise realizada pelo SnpEff, mostrando concordância entre essas duas abordagens.

Em conjunto, esses resultados permitiram entender melhor a evolução da peste no Brasil ao longo do tempo.

Juntos, as abordagens de pangenoma, variações de SNVs e cgMLST permitiram uma análise abrangente dos genomas brasileiros de *Y. pestis* disponíveis no acervo da Fiocruz/CYP. Por meio dessa abordagem *in silico* foi possível verificar aspectos evolutivos e obter informações valiosas sobre a diversidade genética e da virulência de *Y. pestis* no Brasil. A análise de pangenoma evidenciou a ausência de vários genes de virulência associados à adesão celular, aquisição de ferro e formação de biofilme, indicando atenuação da virulência em cepas dos clados C, E, G e H. As cepas do clado G apresentaram a maior ausência de genes de virulência, incluindo genes muito característicos de *Y. pestis*, sugerindo que este foi o grupo de cepas mais afetado. O clado E foi o que mais se beneficiou com a análise aprofundada do InterProScan mostrando principalmente a ausência de genes relacionados ao T6SS, o que possivelmente indica uma atenuação da virulência nas cepas desse clado, mas destaca também a necessidade de uma análise mais aprofundada acerca do papel desse sistema de secreção na patogenicidade de *Y. pestis*. Por fim, o clado H apresentou a ausência de genes mais relacionados a plasmídeos e parece ser o grupo menos afetado pela atenuação de virulência entre os clados. Já o cgMLST realizado neste estudo forneceu informações valiosas sobre a diversidade genética, padrões evolutivos e dinâmica epidemiológica de *Y. pestis* no Brasil. Ao integrar dados genômicos e epidemiológicos, foi possível fazer correlações significativas entre os anos de isolamento e a ocorrência de surtos, lançando luz sobre a trajetória evolutiva das cepas de *Y. pestis* em circulação no país. Essas análises são corroboradas pelos resultados da análise de variações de SNV.

É preciso levar em conta nos resultados o possível impacto que o processo de cultivo e manutenção das cepas ao longo do tempo. As cepas mais antigas dessa coleção são dos anos 60 e foram mantidas em ágar simples com tampa de borracha, sob refrigeração a 4°C, sofrendo vários repiques ao longo dos anos. Todos esses fatores podem levar a deleção de genes, e podem ter influenciado as análises nas três abordagens computacionais empregadas no estudo.

Os resultados deste estudo destacam a complexidade da evolução de *Y. pestis* e sua adaptação aos ambientes locais. A identificação de grupos e linhagens distintos ao longo do tempo reflete a interação entre mutações genéticas, pressões de seleção e fatores geográficos. Outro aspecto importante é verificado na ausência de genes de resistência a antimicrobianos nas cepas brasileiras. Tais percepções não apenas aprofundam nossa compreensão da

dinâmica de transmissão do patógeno, mas também têm implicações para o planejamento dentro da área da saúde pública e estratégias de resposta.

Dessa forma, esse estudo demonstra o poder da genômica nas investigações epidemiológicas. A integração de dados genéticos com informações epidemiológicas detalhadas tem o potencial de contribuir com a compreensão da disseminação de doenças e identificar fontes de infecção com mais precisão. Conforme demonstrado, essa abordagem pode desvendar padrões ocultos, iluminar eventos de transmissão e aprimorar a capacidade de prever e controlar surtos futuros.

Em conclusão, as análises com os isolados de *Y. pestis* no Brasil forneceram um quadro abrangente da paisagem genética do patógeno e sua relação dinâmica com eventos epidemiológicos, contribuindo para uma compreensão mais ampla da evolução de *Y. pestis*. O estudo ainda lança luz sobre fatores de virulência pouco estudados em *Y. pestis* além de enfatizar a importância dos dados genômicos no avanço do conhecimento da dinâmica de doenças infecciosas. À medida que as análises genômicas continuam a evoluir, mais estudos podem ser elaborados sobre essas bases, levando a estratégias mais eficazes para prevenir e gerenciar surtos de *Y. Pestis*, visto que a utilização de informações genômicas para investigações epidemiológicas e rastreamento de surtos tem mostrado um potencial promissor.

11 CONCLUSÃO

- a) De acordo com o perfil de virulência das cepas de *Y. pestis*, foi possível agrupá-las em clados filogenéticos distintos;
- b) As cepas brasileiras de *Y. pestis* tiveram seu perfil de virulência atenuado ao longo do tempo e ainda não apresentam perfil de resistência a antimicrobianos;
- c) Foi possível associar dados epidemiológicos e genômicos de modo a observar perfis evolutivos distintos ao longo do tempo em surtos específicos da doença;
- d) Foi possível agrupar cepas isoladas de surtos no município de Triunfo e na Chapada da Borborema onde foi observado perfis clonais semelhantes entre as cepas isoladas nestas localidades, em um mesmo período, com evoluções distintas para essas localidades. Entretanto, não foi possível associar cepas que ocorrem apenas em um determinado hospedeiro.
- e) Foi possível observar um acúmulo de variantes acima da média nos grupos de cepas dos surtos no município de Triunfo e na Chapada da Borborema, corroborando com os agrupamentos encontrados nos perfis clonais.

REFERÊNCIAS

- ALAM, A. *et al.* The Role of ClpB in Bacterial Stress Responses and Virulence. **Frontiers in Molecular Biosciences**, v. 8, 2021.
- AL-JAWDAH, A. D. *et al.* Induction of the immunoprotective coat of *Yersinia pestis* at body temperature is mediated by the Caf1R transcription factor. **BMC microbiology**, v. 19, n. 1, p. 68, 29, 2019.
- ALMEIDA, A. M. P. DE *et al.* **PESTE: Serviço de Referência**, 2002. Disponível em: https://www.fiocruz.br/biosseguranca/Bis/manuais/epidemiologia/manual_pestes.pdf, 2002.
- ALMEIDA, A. M. P. DE *et al.* Chromosomal *irp2* gene in *Yersinia*: distribution, expression, deletion and impact on virulence. **Microbial Pathogenesis**, v. 14, n. 1, p. 9–21, 1993.
- ALMEIDA, A. M. P. DE *et al.* Survey of the *irp2* gene among *Yersinia pestis* strains isolated during several plague outbreaks in northeast Brazil. **Memórias do Instituto Oswaldo Cruz**, v. 89, n. 1, p. 87–92, 1994.
- ALMEIDA, A. M. P. DE *et al.* Does the Plague Still Threaten Us? **Revista da Sociedade Brasileira de Medicina Tropical**, v. 53, 2020.
- ANDREWS, S. **FastQC: A Quality Control Tool for High Throughput Sequence Data**, 2010. Disponível em: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- ANISIMOV, A. P.; AMOAKO, K. K. Treatment of plague: promising alternatives to antibiotics. **Journal of Medical Microbiology**, v. 55, n. 11, p. 1461–1475, 2006.
- ANISIMOV, A. P.; LINDER, L. E.; PIER, G. B. Intraspecific Diversity of *Yersinia pestis*. **Clinical Microbiology Reviews**, v. 17, n. 2, p. 434–464, 2006.
- AYYADURAI, S. *et al.* Long-term persistence of virulent *Yersinia pestis* in soil. **Microbiology**, v. 154, n. 9, p. 2865–2871, 2008.
- BACOT, A. W.; MARTIN, C. J. Observations on the mechanism of the transmission of plague by fleas. **Journal of Hygiene**, v. 13, p. 423–439, 1914.
- BAO, R. *et al.* Structural basis for the specific recognition of dual receptors by the homopolymeric pH 6 antigen (Psa) fimbriae of *Yersinia pestis*. **Proceedings of the National Academy of Sciences**, v. 13, n. 3, p. 1065–1070, 2013.
- BARBIERI, R. *et al.* *Yersinia pestis*: the Natural History of Plague. **Clinical Microbiology Reviews**, v. 34, n. 1, p. 1–44, 2020.
- BARROS, M. P. S. *et al.* Subtyping Brazilian *Yersinia pestis* strains by pulsed-field gel electrophoresis. **Genetics and Molecular Research**, v. 12, n. 2, p. 1294–1302, 2013.
- BARROS, M. P. S. *et al.* Dynamics of CRISPR Loci in Microevolutionary Process of *Yersinia pestis* Strains. **PLoS ONE**, v. 9, n. 9, p. e108353, 2014.

BERTHERAT, E. *et al.* Plague Reappearance in Algeria after 50 Years, 2003. **Emerging Infectious Diseases**, v. 13, n. 10, p. 1459–1462, 2007.

_____. Plague around the world in 2019. **Weekly Epidemiological Record = Relevé épidémiologique hebdomadaire**, v. 94, n. 25, p. 289–292, 2019.

BHASIN, M.; RAGHAVA, G. P. S. Computational Methods in Genome Research. **Applied Mycology and Biotechnology**. v. 6, p. 179–207, 2006.

BITAM, I. *et al.* New Rural Focus of Plague, Algeria. **Emerging Infectious Diseases**, v. 16, n. 10, p. 1639–1640, 2010.

BOHN, E. *et al.* Bacterial adhesion and host cell factors leading to effector protein injection by type III secretion system. **International Journal of Medical Microbiology**, v. 309, n. 5, p. 344–350, 2019.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014.

BOLOTIN, E.; HERSHBERG, R. Gene Loss Dominates As a Source of Genetic Variation within Clonal Pathogenic Bacterial Species. **Genome Biology and Evolution**, v. 7, n. 8, p. 2173–2187, 2015.

BONACORSI, S. P. *et al.* Assessment of a fluoroquinolone, three beta-lactams, two aminoglycosides, and a cycline in treatment of murine *Yersinia pestis* infection. **Antimicrobial Agents and Chemotherapy**, v. 38, n. 3, p. 481–486, 1994.

BORTOLAIA, V. *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. **Journal of Antimicrobial Chemotherapy**, v. 75, n. 12, p. 3491–3500, 2020.

BOS, K. I. *et al.* Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. **eLife**, v. 5, p. 1–11, 2016.

BRAMANTI, B. *et al.* Plague: A Disease Which Changed the Path of Human Civilization. **Advances in experimental medicine and biology**, v. 918, p. 1–26, 2016.

BRYNILDSRUD, O. *et al.* Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. **Genome Biology**, v. 17, n. 1, p. 238, 2016.

BURNENS, A. P. *et al.* The flagellin N-methylase gene *fliB* and an adjacent serovar-specific IS200 element in *Salmonella typhimurium*. **Microbiology**, v. 143, p. 1539–1547, 1997.

BUTLER, T. Plague Gives Surprises in the First Decade of the 21st Century in the United States and Worldwide. **The American Journal of Tropical Medicine and Hygiene**, v. 89, n. 4, p. 788–793, 2013.

BUTLER, T. Plague history: Yersin's discovery of the causative bacterium in 1894 enabled, in the subsequent century, scientific progress in understanding the disease and the development of treatments and vaccines. **Clinical Microbiology and Infection**, v. 20, n. 3, p. 202–209, 2014.

- CABANEL, N. *et al.* Plague Outbreak in Libya, 2009, Unrelated to Plague in Algeria. **Emerging Infectious Diseases**, v. 19, n. 2, p. 230–236, 2013.
- CAMACHO, C. *et al.* BLAST plus: architecture and applications. **BMC Bioinformatics**, v. 10, n. 421, p. 1, 2009.
- CAVICCHIOLI, R. *et al.* The NarX and NarQ sensor-transmitter proteins of *Escherichia coli* each require two conserved histidines for nitrate-dependent signal transduction to NarL. **Journal of Bacteriology**, v. 177, n. 9, p. 2416–2424, 1995.
- CENTERS FOR DISEASE CONTROL AND PREVENTION. Fatal laboratory-acquired infection with an attenuated *Yersinia pestis* Strain - Chicago, Illinois, 2009. **MMWR. Morbidity and mortality weekly report**, v. 60, n. 7, p. 201–5, 2011.
- CHANTEAU, S. *et al.* Current epidemiology of human plague in Madagascar. **Microbes and Infection**, v. 2, n. 1, p. 25–31, 2000.
- CHEN, L. *et al.* VFDB: a reference database for bacterial virulence factors. **Nucleic Acids Research**, v. 33, p. D325–D328, 2004.
- CHEN, Y. *et al.* Attenuation of *Yersinia pestis* *fyuA* Mutants Caused by Iron Uptake Inhibition and Decreased Survivability in Macrophages. **Frontiers in cellular and infection microbiology**, v. 12, p. 874773, 2022.
- CHOUIKHA, I.; HINNEBUSCH, B. J. Silencing urease: a key evolutionary step that facilitated the adaptation of *Yersinia pestis* to the flea-borne transmission route. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 52, p. 18709–14, 2014.
- CINGOLANI, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. **Fly**, v. 6, n. 2, p. 80–92, 2012.
- CRÉCY-LAGARD, V. DE; HANSON, A. Comparative Genomics. *Em: Brenner's Encyclopedia of Genetics*. v. 2, p. 102–105, 2013.
- DAI, R. *et al.* A novel mechanism of streptomycin resistance in *Yersinia pestis*: Mutation in the *rpsL* gene. **PLoS neglected tropical diseases**, v. 15, n. 4, p. e0009324, 2021.
- DAMGAARD, P. DE B. *et al.* 137 ancient human genomes from across the Eurasian steppes. **Nature**, v. 557, n. 7705, p. 369–374, 2018.
- DEMEURE, C. E. *et al.* *Yersinia pestis* and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. **Genes & Immunity**, v. 20, n. 5, p. 357–370, 2019.
- DENG, W. *et al.* Genome sequence of *Yersinia pestis* KIM. **Journal of bacteriology**, v. 184, n. 16, p. 4601–11, 2002.

DEVIGNAT, R. Variétés de l'espèce *Pasteurella pestis*; nouvelle hypothèse. **Bull World Health Organization**, v.4, n. 2, p. 247-63,1951.

DITCHBURN, J. L.; HODGKINS, R. *Yersinia pestis*, a problem of the past and a re-emerging threat. **Biosafety and Health**, v. 1, n. 2, p. 65–70, 2019.

DRANCOURT, M. *et al.* Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. **Proceedings of the National Academy of Sciences**, v. 95, n. 21, p. 12637–12640, 1998.

DU, Y.; ROSQVIST, R.; FORSBERG, A. Role of fraction 1 antigen of *Yersinia pestis* in inhibition of phagocytosis. **Infection and immunity**, v. 70, n. 3, p. 1453–60, 2002.

EISEN, R. J. *et al.* Early-phase transmission of *Yersinia pestis* by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics. **Proceedings of the National Academy of Sciences**, v. 103, n. 42, p. 15380–15385, 2006.

EWELS, P. *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. **Bioinformatics**, v. 32, n. 19, p. 3047–3048, 2016.

FERNANDES, D. L. R. DA S. *et al.* Rodent hosts and flea vectors in Brazilian plague foci: a review. **Integrative Zoology**, v.6, n.6, p.810-819, 2021.

FERNANDES, D. L. R. DA S. *et al.* Spatiotemporal analysis of bubonic plague in Pernambuco, northeast of Brazil: Case study in the municipality of Exu. **PLOS ONE**, v. 16, n. 4, p. e0249464, 2021.

FETHERSTON, JACQUELINE D *et al.* The Yersiniabactin Transport System Is Critical for the Pathogenesis of Bubonic and Pneumonic Plague. **Infection and Immunity**, v. 78, n. 5, p. 2045–2052, 2010.

FIELD, D.; WILSON, G.; GAST, C. VAN DER. How do we compare hundreds of bacterial genomes? **Current Opinion in Microbiology**, v. 9, n. 5, p. 499–504, 2006.

FORD, B. *et al.* Structural homology between the C-terminal domain of the PapC usher and its plug. **Journal of bacteriology**, v. 192, n. 7, p. 1824–31, 2010.

GAGE, K. L.; KOSOY, M. Y. NATURAL HISTORY OF PLAGUE: Perspectives from More than a Century of Research. **Annual Review of Entomology**, v. 50, n. 1, p. 505–528, 2005.

GALIMAND, M. *et al.* Multidrug Resistance in *Yersinia pestis* Mediated by a Transferable Plasmid. **New England Journal of Medicine**, v. 337, n. 10, p. 677–681, 1997.

GALIMAND, M.; CARNIEL, E.; COURVALIN, P. Resistance of *Yersinia pestis* to Antimicrobial Agents. **ANTIMICROBIAL AGENTS AND CHEMOTHERAPY**, v. 50, n. 10, p. 3233–3236, 2006.

GILES, J.; PETERSON, A. T.; ALMEIDA, A. Ecology and Geography of Plague Transmission Areas in Northeastern Brazil. **PLoS Neglected Tropical Diseases**, v. 5, n. 1, p. e925, 2011.

- GREEN, M. H. *et al.* *Yersinia pestis* and the three plague pandemics. **The Lancet Infectious Diseases**, v. 14, n. 10, p. 918, 2014.
- GREMME, G.; STEINBISS, S.; KURTZ, S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. **IEEE/ACM transactions on computational biology and bioinformatics**, v. 10, n. 3, p. 645–56, 2013.
- GUERN, A. S. LE *et al.* *Yersinia artesianiana* sp. nov., *Yersinia proxima* sp. nov., *Yersinia alsatica* sp. nov., *Yersinia vastinensis* sp. nov., *Yersinia thracica* sp. nov. and *Yersinia occitanica* sp. nov., isolated from humans and animals. **International Journal of Systematic and Evolutionary Microbiology**, v. 70, n. 10, p. 5363–5372, 2020.
- GUILLOUARD, I. *et al.* Identification of *Bacillus subtilis* CysL, a regulator of the *cysJI* operon, which encodes sulfite reductase. **Journal of bacteriology**, v. 184, n. 17, p. 4681–9, 2002.
- GUINDON, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. **Systematic Biology**, v. 59, n. 3, p. 307–321, 2010.
- GUINDON, S.; GASCUEL, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. **Systematic Biology**, v. 52, n. 5, p. 696–704, 2003.
- GUIYOULE, A. Transferable Plasmid-Mediated Resistance to Streptomycin in Clinical Isolate of *Yersinia pestis*. **Emerging Infectious Diseases**, v. 7, n. 1, p. 43–48, 2001.
- HIGGINS, J. A. *et al.* 5' Nuclease PCR Assay To Detect *Yersinia pestis*. **Journal of Clinical Microbiology**, v. 36, n. 8, p. 2284–2288, 1998.
- HINNEBUSCH, B JOSEPH *et al.* Role of *Yersinia murine* toxin in survival of *Yersinia pestis* in the midgut of the flea vector. **Science**, v. 296, n. 5568, p. 733–5, 2002.
- HINNEBUSCH, B. JOSEPH *et al.* High-frequency conjugative transfer of antibiotic resistance genes to *Yersinia pestis* in the flea midgut. **Molecular Microbiology**, v. 46, n. 2, p. 349–354, 2002.
- HINNEBUSCH, B. J.; ERICKSON, D. L. *Yersinia pestis* biofilm in the flea vector and its role in the transmission of plague. **Current topics in microbiology and immunology**, v. 322, p. 229–48, 2008.
- HINNEBUSCH, B. J.; JARRETT, C. O.; BLAND, D. M. “Fleaing” the Plague: Adaptations of *Yersinia pestis* to Its Insect Vector That Lead to Transmission. **Annual Review of Microbiology**, v. 71, n. 1, p. 215–232, 2017.
- HINNEBUSCH, B. J.; PERRY, R. D.; SCHWAN, T. G. Role of the *Yersinia pestis* Hemin Storage (*hms*) Locus in the Transmission of Plague by Fleas. **Science**, v. 273, n. 5273, p. 367–370, 1996.
- HINNEBUSCH, J.; SCHWAN, T. G. New method for plague surveillance using polymerase chain reaction to detect *Yersinia pestis* in fleas. **Journal of Clinical Microbiology**, v. 31, n. 6, p. 1511–1514, 1993.

HOOD, M. I.; SKAAR, E. P. Nutritional immunity: transition metals at the pathogen–host interface. **Nature Reviews Microbiology**, v. 10, n. 8, p. 525–537, 2012.

HURST, M. R. H. *et al.* Temperature-Dependent *Galleria mellonella* Mortality as a Result of *Yersinia entomophaga* Infection. **Applied and Environmental Microbiology**, v. 81, n. 18, p. 6404–6414, 2015.

HYATT, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC Bioinformatics**, v. 11, n. 1, p. 119, 2010.

ITOH, Y. *et al.* Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. **Journal of bacteriology**, v. 190, n. 10, p. 3670–80, 2008.

JONES, P. *et al.* InterProScan 5: genome-scale protein function classification. **Bioinformatics**, v. 30, n. 9, p. 1236–1240, 2014.

KANONENBERG, K.; SMITS, S. H. J.; SCHMITT, L. Functional Reconstitution of HlyB, a Type I Secretion ABC Transporter, in Saposin-A Nanoparticles. **Scientific reports**, v. 9, n. 1, p. 8436, 2019.

KARIMI, Y. *et al.* Particularités des souches de *Yersinia pestis* isolées dans le nord-est du Brésil. **Annales de microbiologie**, v. 125A, n. 2, p. 243–6, 1974.

KAWASHIMA, T. Comparative and evolutionary genomics. **Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics**, v. 1–3, p. 257–267, 2018.

KELLER, M. *et al.* Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). **Proceedings of the National Academy of Sciences**, v. 116, n. 25, p. 12363–12372, 2019.

KUBERSKI, T.; ROBINSON, L.; SCHURGIN, A. A Case of Plague Successfully Treated with Ciprofloxacin and Sympathetic Blockade for Treatment of Gangrene. **Clinical Infectious Diseases**, v. 36, n. 4, p. 521–523, 2003.

LEAL, N. C. *et al.* Viability of *Yersinia pestis* subcultures in agar stabs. **Letters in Applied Microbiology**, v. 62, n. 1, p. 91–95, 2016.

LEAL, N. C.; ALMEIDA, A. M. Diagnosis of plague and identification of virulence markers in *Yersinia pestis* by multiplex-PCR. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 41, n. 6, p. 339–42, 1999.

LEAL-BALBINO, T. *et al.* Differences in the stability of the plasmids of *Yersinia pestis* cultures in vitro: impact on virulence. **Memórias do Instituto Oswaldo Cruz**, v. 99, n. 7, p. 727–732, 2004.

LEI, C.; KUMAR, S. *Yersinia pestis* antibiotic resistance: a systematic review. **Osong Public Health and Research Perspectives**, v. 13, n. 1, p. 24–36, 2022.

- LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. **Nucleic Acids Research**, v. 47, n. 47, p. W256–W259, 2019.
- LI, H. *et al.* The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.
- LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589–595, 2010.
- LIU, F.; CHEN, H.; GALVÁN, E. M.; *et al.* Effects of Psa and F1 on the Adhesive and Invasive Interactions of *Yersinia pestis* with Human Respiratory Tract Epithelial Cells. **Infection and Immunity**, v. 74, n. 10, p. 5636–5644, 2006.
- LOÑEZ, C. *et al.* Detection of *Yersinia pestis* in Sputum by Real-Time PCR. **Journal of Clinical Microbiology**, v. 41, n. 10, p. 4873–4875, 2003.
- MAIDEN, M. C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 6, p. 3140–5, 1998.
- MAHMOUDI, A. *et al.* Plague reservoir species throughout the world. **Integrative Zoology**, v. 16, n. 6, p. 820–833, 2021.
- MANGALEA, M. R.; BORLEE, B. R. The NarX-NarL two-component system regulates biofilm formation, natural product biosynthesis, and host-associated survival in *Burkholderia pseudomallei*. **Scientific Reports**, v. 12, n. 203, 2022.
- MCKENNA, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, v. 20, n. 9, p. 1297–1303, 2010.
- MEDINI, D. *et al.* The microbial pan-genome. **Current Opinion in Genetics & Development**, v. 15, n. 6, p. 589–594, 2005.
- MEYER, K. F. *et al.* EFFECTIVE TREATMENT OF PLAGUE. **Annals of the New York Academy of Sciences**, v. 55, n. 6, p. 1228–1274, 1952.
- MONTMINY, S. W. *et al.* Virulence factors of *Yersinia pestis* are overcome by a strong lipopolysaccharide response. **Nature Immunology**, v. 7, n. 10, p. 1066–1073, 2006.
- NAMOUCI, A. *et al.* Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. **Proceedings of the National Academy of Sciences**, v. 115, n. 50, p. E11790–E11797, 2018.
- NAUTH, T. *et al.* Visualization of translocons in *Yersinia* type III protein secretion machines during host cell infection. **PLOS Pathogens**, v. 14, n. 12, p. e1007527, 2018.
- NGUYEN, S. V. *et al.* *Yersinia hibernica* sp. nov., isolated from pig-production environments. **International Journal of Systematic and Evolutionary Microbiology**, v. 69, n. 7, p. 2023–2027, 2019.

OLIVEIRA, M. B. M. *et al.* Genetic diversity of yersinia pestis in Brazil. **Genetics and Molecular Research**, v. 11, n. 3, p. 3414–3424, 2012.

PAGE, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. **Bioinformatics**, v. 31, n. 22, p. 3691–3693, 2015.

PECHOUS, R. D. *et al.* In vivo transcriptional profiling of yersinia pestis reveals a novel bacterial mediator of pulmonary inflammation. **mBio**, v. 6, n. 1, p. e02302-14, 2015.

PELLUDAT, C. *et al.* The Yersiniabactin Biosynthetic Gene Cluster of Yersinia enterocolitica: Organization and Siderophore-Dependent Regulation. **Journal of Bacteriology**, v.180, n.3, p.538-46, 1998.

PERRY, R. D.; FETHERSTON, J. D. Yersinia pestis - Etiologic agent of plague. **Clinical Microbiology Reviews**, v.10, n. 1, p. 35-66, 1997.

POSADA, D. jModelTest: Phylogenetic Model Averaging. **Molecular Biology and Evolution**, v. 25, n. 7, p. 1253–1256, 2008.

PRICE, M. N.; DEHAL, P. S.; ARKIN, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. **Molecular Biology and Evolution**, v. 26, n. 7, p. 1641–1650, 2009.

QUINN, J. D. *et al.* Temperature Control of psaA Expression by PsaE and PsaF in Yersinia pestis. **Journal of bacteriology**, v. 201, n. 16, p. e00217-19, 2019.

RADNEDGE, L. *et al.* Genome plasticity in Yersinia pestis. **Microbiology**, v. 148, n. 6, p. 1687–1698, 2002.

RAHALISON, L. *et al.* Diagnosis of Bubonic Plague by PCR in Madagascar under Field Conditions. **Journal of clinical microbiology**, v. 38, n. 1, p. 260–263, 2000.

RAINA, S. *et al.* Identification and Transcriptional Analysis of the Escherichia coli htrE Operon Which Is Homologous to pap and Related Pilin Operons. **Journal of bacteriology**, v.175, n. 16, p. 5009-21, 1993.

RAKIN, A. *et al.* The pesticin receptor of Yersinia enterocolitica: a novel virulence factor with dual function. **Molecular microbiology**, v. 13, n. 2, p. 253–63, 1994.

RAOULT, D. *et al.* Plague: History and contemporary analysis. **Journal of Infection**, v. 66, n. 1, p. 18–26, 2013.

RASCOVAN, N. *et al.* Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. **Cell**, v. 176, n. 1–2, p. 295- 305.e10, 2019.

RASMUSSEN, S. *et al.* Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. **Cell**, v. 163, n. 3, p. 571–582, 2015.

- REN, C.-P. *et al.* The Flag-2 locus, an ancestral gene cluster, is potentially associated with a novel flagellar system from *Escherichia coli*. **Journal of bacteriology**, v. 187, n. 4, p. 1430–40, 2005.
- ROCHA, A. *et al.* Primeiro workshop interno dos serviços de referência do Centro de Pesquisas Aggeu Magalhães da Fundação Oswaldo Cruz. **Revista De Patologia Tropical**, v. 38, n. 4, p. 299–309, 2010.
- RUSSELL, P. *et al.* Doxycycline or ciprofloxacin prophylaxis and therapy against experimental *Yersinia pestis* infection in mice. **Journal of Antimicrobial Chemotherapy**, v. 37, n. 4, p. 769–774, 1996.
- SAHL, J. W. *et al.* NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. **Microbial Genomics**, v. 2, n. 8, p. e000074, 2016.
- SAVIN, C. *et al.* The *Yersinia pseudotuberculosis* complex: Characterization and delineation of a new species, *Yersinia wautersii*. **International Journal of Medical Microbiology**, v. 304, n. 3–4, p. 452–463, 2014.
- _____. Genus-wide *Yersinia* core-genome multilocus sequence typing for species identification and strain characterization. **Microbial Genomics**, v. 5, n. 10, p. e000301, 2019.
- SCHUBERT, S.; FISCHER, D.; HEESEMANN, J. Ferric enterochelin transport in *Yersinia enterocolitica*: molecular and evolutionary aspects. **Journal of bacteriology**, v. 181, n. 20, p. 6387–95, 1999.
- SCHULTE, R.; CORNELIS, G. R. Mutational analysis of the *Yersinia enterocolitica* virC operon: characterization of yscE, F, G, I, J, K required for Yop secretion and yscH encoding YopR. **Molecular Microbiology**, v.18, n. 2, p.343-55, 1995.
- SEEMANN, T. Prokka: rapid prokaryotic genome annotation. **Bioinformatics**, v. 30, n. 14, p. 2068–2069, 2014.
- SHERMAN, R. M.; SALZBERG, S. L. Pan-genomics in the human genome era. **Nature Reviews Genetics**, v. 21, n. 4, p. 243–254, 2020.
- SHI, L. *et al.* New Genotype of *Yersinia pestis* Found in Live Rodents in Yunnan Province, China. **Frontiers in Microbiology**, v. 12, p. 628335, 2021.
- SILVA, M. *et al.* chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. **Microbial Genomics**, v. 4, n. 3, p. e000166, 2018.
- SINGH, A. *et al.* Anti-biofilm and anti-virulence potential of 3,7-dimethyloct-6-enal derived from *Citrus hystrix* against bacterial blight of rice caused by *Xanthomonas oryzae* pv. *oryzae*. **Microbial pathogenesis**, v. 115, p. 264–271, 2018.
- SONG, Y. Complete Genome Sequence of *Yersinia pestis* Strain 91001, an Isolate Avirulent to Humans. **DNA Research**, v. 11, n. 3, p. 179–197, 2004.

SPYROU, M. A. *et al.* Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. **Nature Communications**, v. 9, n. 1, p. 2234, 2018.

_____. Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. **Nature Communications**, v. 10, n. 1, 2019.

SPYROU, MARIA A *et al.* Ancient pathogen genomics as an emerging tool for infectious disease research. **Nature Reviews Genetics**, v. 20, n. 6, p. 323–340, 2019.

SUSAT, J. *et al.* *Yersinia pestis* strains from Latvia show depletion of the *pla* virulence gene at the end of the second plague pandemic. **Scientific Reports**, v. 10, n. 1, p. 1–10, 2020.

TAVARES, C. *et al.* Plague in Brazil: From Now and Then. **Advances in Experimental Medicine and Biology**. v. 954, p. 69–77, 2012.

_____. Peste, uma zoonose esquecida. **Journal of Health & Biological Sciences**, v. 8, n. 1, p. 1, 2020.

TETTELIN, H. *et al.* Comparative genomics: the bacterial pan-genome. **Current Opinion in Microbiology**, v. 11, n. 5, p. 472–477, 2008.

THOMPSON, J. M.; JONES, H. A.; PERRY, R. D. Molecular characterization of the hemin uptake locus (*hmu*) from *Yersinia pestis* and analysis of *hmu* mutants for hemin and hemoprotein utilization. **Infection and immunity**, v. 67, n. 8, p. 3879–92, 1999.

THOMSON, J. J.; PLECHA, S. C.; KRUKONIS, E. S. Ail provides multiple mechanisms of serum resistance to *Yersinia pestis*. **Molecular Microbiology**, v. 111, n. 1, p. 82–95, 2019.

TRAN, T.-N.-N. *et al.* High Throughput, Multiplexed Pathogen Detection Authenticates Plague Waves in Medieval Venice, Italy. **PLoS ONE**, v. 6, n. 3, p. e16735, 2011.

TROSKY, J. E.; LIVERMAN, A. D. B.; ORTH, K. *Yersinia* outer proteins: Yops. **Cellular Microbiology**, v. 10, n. 3, p. 557–565, 2008.

TSANG, T. M.; FELEK, S.; KRUKONIS, E. S. Ail binding to fibronectin facilitates *Yersinia pestis* binding to host cells and Yop delivery. **Infection and immunity**, v. 78, n. 8, p. 3358–68, 2010.

VADYVALOO, V.; HINZ, A. K. A LysR-Type Transcriptional Regulator, RovM, Senses Nutritional Cues Suggesting that It Is Involved in Metabolic Adaptation of *Yersinia pestis* to the Flea Gut. **PLoS ONE**, v. 10, n. 9, p. e0137508, 2015.

VALLÈS, X. *et al.* Human plague: An old scourge that needs new answers. **PLoS Neglected Tropical Diseases**, v. 14, n. 8, p. e0008251, 2020.

VALLES, X. *et al.* PLOS NEGLECTED TROPICAL DISEASES Human plague: An old scourge that needs new answers. **PLoS Neglected Tropical Diseases**, v. 14, n. 18, p. e0008251, 2020.

VOGLER, A. J. *et al.* A single introduction of *Yersinia pestis* to Brazil during the 3rd plague pandemic. **PLOS ONE**, v. 14, n. 1, p. e0209478, 2019.

VOGLER, A. J.; KEIM, P; WAGNER, D. M. A review of methods for subtyping *Yersinia pestis*: From phenotypes to whole genome sequencing. **Infection, Genetics and Evolution**, v. 37, p. 21-36, 2016.

WORLD HEALTH ORGANIZATION. **Plague – Madagascar**. Disponível em: <https://www.who.int/emergencies/disease-outbreak-news/item/plague---madagascar>, 2018.

XU, Z. *et al.* Impact of *pmrA* on *Cronobacter sakazakii* planktonic and biofilm cells: A comprehensive transcriptomic study. **Food microbiology**, v. 98, p. 103785, 2021.

YANG, R. *et al.* *Yersinia pestis* and Plague: Some Knowns and Unknowns. **Zoonoses**, v. 3, n.1, 2023.

YANG, X. *et al.* Type VI Secretion Systems Present New Insights on Pathogenic *Yersinia*. **Frontiers in cellular and infection microbiology**, v. 8, p. 260, 2018.

YU, H. *et al.* Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia. **Cell**, v. 181, n. 6, p. 1232- 1245, 2020.

YUAN, Y. *et al.* The characteristics and genome analysis of the novel *Y. pestis* phage JC221. **Virus Research**, v. 283, p. 197982, 2020.

ZAUBERMAN, A. *et al.* Host Iron Nutritional Immunity Induced by a Live *Yersinia pestis* Vaccine Strain Is Associated with Immediate Protection against Plague. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. 277, 2017.

ZERBINO, D. R.; BIRNEY, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, v. 18, n. 5, p. 821–829, 2008.

ZHOU, D. *et al.* DNA Microarray Analysis of Genome Dynamics in *Yersinia pestis*: Insights into Bacterial Genome Microevolution and Niche Adaptation. **Journal of Bacteriology**, v. 186, n. 15, p. 5138–5146, 2004.

ZHOU, Z. *et al.* GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. **Genome Research**, v. 28, n. 9, p. 1395–1404, 2018.

ZIMBLER, D. L. *et al.* Inactivation of Peroxiredoxin 6 by the Pla Protease of *Yersinia pestis*. **Infection and Immunity**, v. 84, n. 1, p. 365–374, 2016.

APÊNDICE A - DADOS DO SEQUENCIAMENTO E MONTAGEM DAS CEPAS DE *Y.pestis* DA FIOCRUZ/CYP

Os dados resultantes dos sequenciamentos realizados e utilizados para mensurar a qualidade das montagens na tabela 1 desde apêndice.

Tabela 1 - Dados do sequenciamento e montagem das cepas de *Y.pestis* da FIOCRUZ/CYP

CEPA	READS	CONTIGS	GENOMA (PB)	N50	CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3620	3.191.594	78	4758590	2454675	Yp3978	963.907	149	4942187	4017970
Yp3800	2.459.324	93	4771249	4655928	Yp3668	3.012.075	155	6407278	3242189
Yp3341	1.335.008	110	4586472	4512600	Yp3726	1.466.374	162	4874917	3251655
Yp3406	1.943.095	110	4871894	2691298	Yp3853	1.491.574	162	5029749	4170700
Yp3588	1.769.638	113	4838674	4674038	Yp3459	2.552.638	163	5158223	1604120
Yp3501	2.253.985	116	4933204	1521370	Yp3647	1.138.323	174	4969425	3314093
Yp3950	1.480.927	120	4670904	1243015	Yp3865	1.947.303	174	5179643	4352122
Yp3229	1.941.372	121	5004701	1898267	Yp3583	1.163.141	189	5034425	3245377
Yp3789	3.644.285	121	4900826	2469988	Yp3559	1.515.334	190	4770708	3756745
Yp3862	1.152.620	121	4621914	2370784	Yp3679	1.581.346	200	5003666	827416
Yp3979	1.402.846	122	4818594	2545820	Yp3792	2.524.070	202	4605136	803494
Yp3245	845.582	124	4674850	3929414	Yp3974	2.347.057	211	4691965	2403760
Yp3516	2.048.378	124	5125422	4769504	Yp3825	910.996	214	5037003	2006034
Yp3336	1.504.724	126	4902145	4715551	Yp3840	3.288.447	228	5108562	4020656
Yp3693	2.675.537	126	4796976	3186572	Yp3537	3.025.702	233	4745856	2247330
Yp3287	2.985.028	129	4727872	2823254	Yp3994	983.487	235	5015972	3877210
Yp3785	1.516.191	129	4657246	4430143	Yp3705	782.485	237	4851605	4186688
Yp3600	1.714.540	135	4714867	1042222	Yp3640	3.324.977	255	4738138	2378262
Yp3852	1.236.673	144	4862862	4487567	Yp3680	2.491.473	260	4840136	796316
Yp3442	2.111.913	147	4817683	4516885	Yp3905	691.615	261	4764795	858784
Yp3381	2.091.191	149	5593711	4694460	Yp3547	1.232.017	264	4628945	984924
Yp3539	1.869.159	149	4775305	1125678	Yp3569	1.006.839	265	4576532	1042216
Yp3783	1.330.782	316	4716035	2947747	Yp3536	2.485.593	267	4544209	1125370
Yp3454	2.129.676	396	4441392	26722	Yp3712	1.603.683	269	4736102	2891279
Yp3797	2.916.716	405	4615800	26076	Yp3823	1.801.906	270	4823622	743968

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3532	2.608.034	407	4432388	26949
Yp3471	4.136.430	408	4416294	27271
Yp3446	3.818.361	416	4400665	26953
Yp3651	3.391.389	416	4627717	26965
Yp3821	3.338.908	418	4475490	25605
Yp3452	3.481.674	424	4604089	26722
Yp3625	3.465.653	426	4527343	26048
Yp3803	2.709.489	428	4533794	27254
Yp3200	1.642.355	430	4452788	24690
Yp3858	1.924.268	430	4611194	26056
Yp3444	4.336.057	431	4578220	27250
Yp3483	3.524.426	434	4622838	26726
Yp3540	2.942.556	435	4533320	26076
Yp3692	3.382.689	435	4626198	25947
Yp3684	3.300.298	438	4384582	26048
Yp3571	1.599.009	439	4565420	25927
Yp3952	2.372.284	439	4602027	26730
Yp3636	2.516.017	440	4629007	26036
Yp3303	1.999.276	441	4405946	25434
Yp3744	2.252.867	441	4534395	26953
Yp3805	3.003.997	442	4546220	26080
Yp3650	3.571.192	444	4629137	26969
Yp3529	2.335.201	445	4540859	26730
Yp3977	6.183.595	445	4622052	27291
Yp3842	2.341.778	446	4526735	25968
Yp3849	2.549.260	446	4628384	26068
Yp3967	2.667.373	446	4551116	25974

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3405	1.734.112	283	4687642	385747
Yp3812	1.072.535	297	4709712	2777517
Yp3687	1.119.895	330	4959722	1482136
Yp3661	3.553.995	447	4634960	27279
Yp3492	2.915.467	448	4624619	26028
Yp3646	2.514.490	448	4620379	27486
Yp3794	3.340.730	448	4631762	27250
Yp3462	5.148.379	449	4616806	26044
Yp3622	3.576.171	450	4646296	27283
Yp3669	7.681.195	450	4653153	27270
Yp3855	1.254.307	450	4627831	25801
Yp3328	1.970.538	451	4522523	26417
Yp3414	16.186.018	452	4646533	26762
Yp3490	3.492.542	452	4629687	26076
Yp3575	4.203.303	452	4574117	26746
Yp3807	3.229.839	452	4633297	26076
Yp3685	4.683.637	453	4546803	26739
Yp3841	3.281.335	453	4637353	26969
Yp3844	3.486.115	454	4638480	26048
Yp3551	3.962.759	455	4632190	27182
Yp3723	1.982.869	455	4625776	25847
Yp3756	2.323.645	455	4543283	24881
Yp3816	2.882.158	456	4626088	26730
Yp3700	3.662.991	457	4539854	26742
Yp3526	3.116.419	458	4628635	26048
Yp3653	2.887.022	458	4568002	26044
Yp3525	2.382.800	459	4477281	25939

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3480	5.488.043	447	4636950	26080
Yp3505	2.498.202	447	4595365	26722
Yp3475	3.497.048	462	4530892	26953
Yp3479	4.877.574	462	4627445	25943
Yp3735	1.961.082	462	4631915	26727
Yp3346	2.492.206	463	4622451	26726
Yp3637	2.946.463	463	4456109	24671
Yp3703	3.420.678	463	4622917	27258
Yp3655	2.815.910	464	4627798	26072
Yp3708	1.660.583	464	4531707	24636
Yp3793	2.924.108	464	4636061	26044
Yp3528	3.150.745	465	4613082	26027
Yp3715	1.430.062	465	4524933	25578
Yp3325	1.836.785	466	4534522	27478
Yp3447	2.356.935	466	4640633	26063
Yp3603	1.307.407	466	4431111	25422
Yp3623	3.584.834	466	4631473	26743
Yp3506	1.393.982	467	4519104	25426
Yp3768	2.422.510	467	4534317	26036
Yp3476	2.832.254	468	4645289	25605
Yp3587	2.796.522	468	4610844	26738
Yp3901	2.783.545	468	4637730	25605
Yp3398	2.745.744	469	4607353	25604
Yp3458	3.786.428	469	4442920	26028
Yp3511	2.399.824	469	4632248	26953
Yp3533	3.117.822	469	4359639	26031
Yp3630	4.267.029	469	4644396	26973

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3837	3.062.920	459	4636417	26965
Yp3298	1.545.019	460	4629371	26937
Yp3400	1.910.390	460	4515004	26726
Yp3457	4.611.370	460	4587284	25454
Yp3531	3.365.336	460	4616999	26965
Yp3795	1.931.298	460	4624245	25186
Yp3535	3.880.523	461	4542044	26973
Yp3656	3.888.514	470	4555569	26751
Yp3557	2.824.041	471	4614384	26726
Yp3766	1.926.606	471	4639173	25438
Yp3809	3.232.206	471	4605891	26044
Yp3819	3.401.855	471	4597090	26076
Yp3838	2.574.141	471	4637082	25829
Yp3432	3.020.191	472	4570589	26044
Yp3517	2.738.512	472	4540326	25605
Yp3308	1.774.686	473	4529900	26028
Yp3544	3.359.503	473	4638250	26048
Yp3686	2.193.700	473	4592129	26036
Yp3704	2.251.241	473	4554770	26040
Yp3761	2.333.558	473	4634609	26951
Yp3329	2.275.236	474	4543852	26953
Yp3413	2.928.968	474	4644051	25821
Yp3619	2.907.324	474	4630833	25381
Yp3765	2.313.658	474	4625590	25821
Yp3470	3.871.802	475	4578132	26718
Yp3519	3.143.757	475	4629904	25817
Yp3678	2.199.524	475	4626289	26036

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3649	3.160.151	469	4637919	26044
Yp3652	2.555.643	469	4564455	25438
Yp3802	2.557.539	469	4637871	26044
Yp3473	3.094.103	470	4633258	26036
Yp3487	2.911.470	470	4612563	26064
Yp3465	4.607.624	477	4641564	26059
Yp3701	2.587.663	477	4626093	26953
Yp3752	1.812.339	478	4631209	26028
Yp3981	1.197.754	478	4521656	24737
Yp3352	2.042.968	479	4528575	26028
Yp3455	2.757.092	479	4631718	25821
Yp3570	2.446.973	479	4568349	26727
Yp3867	2.454.860	479	4632319	27242
Yp3949	1.629.328	479	4630398	25994
Yp3682	1.090.539	480	4556036	24487
Yp3813	3.113.481	480	4617213	25454
Yp3201	2.121.715	481	4559606	24330
Yp3322	2.316.231	481	4635826	25817
Yp3372	2.362.429	481	4642464	25442
Yp3387	1.719.037	481	4580019	24485
Yp3808	2.628.933	481	4580279	25001
Yp3811	1.981.119	481	4600602	26727
Yp3848	2.713.134	481	4620076	26044
Yp3204	1.869.961	482	4628797	25186
Yp3484	3.702.421	482	4565132	24583
Yp3499	2.537.692	482	4646907	26072
Yp3555	1.748.492	482	4530964	26064

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3788	2.284.741	475	4537611	26056
Yp3815	2.401.556	475	4630950	25589
Yp3340	1.794.871	476	4580598	24643
Yp3489	3.153.309	476	4629754	25813
Yp3507	2.379.157	476	4563573	26032
Yp3524	2.709.371	476	4527310	26926
Yp3648	2.154.926	476	4606155	26730
Yp3753	1.864.148	476	4531973	25442
Yp3817	2.942.735	476	4629343	25454
Yp3438	2.945.749	477	4536409	25947
Yp3851	1.917.203	484	4633106	27242
Yp3342	1.559.415	485	4417063	25430
Yp3364	1.863.941	485	4521176	24981
Yp3696	1.978.243	485	4527966	25914
Yp3746	1.696.698	485	4529495	25186
Yp3810	1.969.375	485	4534405	25935
Yp3822	3.675.510	485	4649376	26044
Yp3274	1.896.621	486	4624609	23890
Yp3448	2.498.469	486	4634567	25597
Yp3234	1.893.278	487	4608608	24493
Yp3439	3.144.851	487	4634192	26060
Yp3463	3.244.826	487	4625235	25935
Yp3472	2.363.060	487	4632940	26032
Yp3498	1.752.546	487	4631571	25438
Yp3521	2.354.213	487	4626926	25939
Yp3565	1.829.489	487	4607374	26726
Yp3763	1.428.705	487	4468307	25581

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3639	2.450.757	482	4623947	25943
Yp3786	2.597.009	482	4548287	25198
Yp3192	1.927.009	483	4618350	25178
Yp3248	2.231.940	483	4637795	25424
Yp3474	4.098.956	483	4531707	26028
Yp3478	3.090.035	483	4577165	25947
Yp3522	2.643.569	483	4611173	25937
Yp3624	2.186.536	484	4454861	25939
Yp3845	2.452.511	484	4634148	26730
Yp3567	1.465.417	491	4612792	24698
Yp3829	1.105.295	491	4596164	25198
Yp3231	1.361.336	492	4427773	24568
Yp3332	1.970.047	492	4626299	25750
Yp3370	2.533.065	492	4641774	25803
Yp3508	1.765.480	492	4585527	25418
Yp3527	3.028.574	492	4635164	24501
Yp3847	1.607.143	492	4560588	24703
Yp3947	2.149.416	492	4633545	25179
Yp3309	1.938.056	493	4626818	26064
Yp3334	2.312.137	493	4643428	25817
Yp3396	1.750.578	493	4504067	25592
Yp3485	2.549.918	493	4621573	25585
Yp3504	2.338.974	493	4615653	24640
Yp3659	3.493.430	493	4636331	25846
Yp3338	1.956.572	494	4631808	25813
Yp3866	1.452.448	494	4523186	25368
Yp3899	1.957.507	494	4532221	25589

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3860	1.414.575	487	4490874	23789
Yp3316	1.861.484	488	4502844	26024
Yp3592	1.691.667	488	4542105	25182
Yp3731	1.372.489	488	4517814	24647
Yp3787	911.832	488	4292710	22724
Yp3854	2.272.730	488	4539299	25943
Yp3856	1.785.422	488	4637008	25592
Yp3518	2.228.302	489	4615514	25581
Yp3747	1.848.125	489	4625813	25941
Yp3297	1.993.621	490	4630507	26024
Yp3801	1.024.002	490	4526413	25418
Yp3299	1.786.684	491	4540317	25585
Yp3435	2.522.850	491	4471770	25931
Yp3491	2.303.168	497	4532200	25421
Yp3496	2.385.624	497	4627940	25581
Yp3660	3.356.488	497	4653858	25462
Yp3850	2.242.797	497	4632911	25935
Yp3500	2.003.974	498	4528685	25364
Yp3707	1.601.650	498	4576773	24489
Yp3774	1.381.143	498	4500422	26024
Yp3186	1.869.571	499	4631353	26027
Yp3358	1.864.740	499	4621561	24981
Yp3503	2.042.169	499	4525907	25438
Yp3512	2.215.729	499	4624570	25801
Yp3520	2.391.889	499	4635541	25194
Yp3580	1.038.033	499	4464004	23257
Yp3453	2.610.061	500	4595940	25364

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3330	2.062.220	495	4619831	25178
Yp3349	1.618.443	495	4634270	24326
Yp3706	1.325.328	495	4625728	25418
Yp3709	2.111.599	495	4644436	26726
Yp3404	2.323.548	496	4615447	25817
Yp3466	3.561.984	496	4630463	26028
Yp3665	3.414.116	496	4520772	25198
Yp3327	2.049.089	497	4628910	25927
Yp3337	1.855.504	497	4624632	25702
Yp3399	2.312.964	497	4556935	26027
Yp3386	2.241.627	505	4644411	25442
Yp3412	2.233.298	505	4634143	25817
Yp3509	2.437.739	505	4627348	23805
Yp3482	2.326.190	506	4632308	24804
Yp3839	2.314.179	506	4615413	25186
Yp3683	1.753.733	507	4567810	26024
Yp3725	1.117.876	507	4622318	24263
Yp3931	1.519.206	507	4628297	25182
Yp3224	1.821.397	508	4527196	24485
Yp3331	1.660.538	508	4508539	25182
Yp3481	2.367.767	508	4640344	25585
Yp3278	2.325.912	510	4563411	24568
Yp3335	1.555.390	510	4513773	24647
Yp3742	1.632.715	510	4643149	25186
Yp3757	1.540.631	510	4508591	25244
Yp3356	1.791.921	511	4597261	24978
Yp3534	2.280.839	511	4618239	25821

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3464	1.940.701	500	4631758	26060
Yp3722	1.873.507	500	4627738	25430
Yp3191	2.317.646	501	4625422	24624
Yp3382	1.702.556	501	4638121	24349
Yp3431	3.546.052	501	4629751	26718
Yp3597	1.536.558	501	4611490	25182
Yp3306	2.298.169	502	4622266	26024
Yp3440	2.564.056	502	4491386	25589
Yp3515	2.469.507	502	4612627	25813
Yp3558	2.519.905	502	4628418	25935
Yp3843	1.592.821	502	4528581	25182
Yp3542	2.775.944	503	4600717	25438
Yp3292	2.427.477	504	4638587	25588
Yp3411	2.100.461	504	4534435	24986
Yp3892	1.771.426	504	4532703	24585
Yp3362	1.916.317	516	4470201	24651
Yp3543	1.696.700	516	4630839	24489
Yp3497	2.153.808	517	4602269	24637
Yp3247	2.106.321	518	4642191	22235
Yp3272	1.666.822	518	4580303	23170
Yp3541	2.453.082	518	4568292	25430
Yp3277	2.232.597	519	4577804	24485
Yp3291	2.060.781	519	4538075	25598
Yp3951	1.414.534	519	4628724	25430
Yp3393	1.854.918	523	4515481	25182
Yp3461	4.003.976	523	4643057	24493
Yp3502	1.255.153	523	4536053	24306

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3621	1.428.005	511	4630704	24632
Yp3780	1.141.169	511	4625249	24643
Yp3495	1.650.709	512	4640641	25585
Yp3834	1.637.882	512	4532454	24982
Yp3408	1.727.792	513	4567953	24978
Yp3545	1.564.021	513	4633438	25438
Yp3568	1.360.233	513	4499113	23193
Yp3824	1.916.091	513	4527317	25929
Yp3833	1.284.606	513	4583878	25430
Yp3638	1.355.103	514	4632830	25799
Yp3755	1.515.865	514	4626137	25174
Yp3754	1.394.889	515	4627611	24562
Yp3185	2.185.023	516	4640366	24489
Yp3681	3.192.468	537	4624635	23805
Yp3820	1.198.972	537	4624219	24477
Yp3578	998.887	538	4460402	24290
Yp3222	1.754.628	539	4559921	22482
Yp3285	935.089	540	4570568	24142
Yp3510	1.270.222	541	4627146	25426
Yp3631	2.257.899	541	4617829	23181
Yp3711	1.113.623	541	4629222	23797
Yp3724	1.432.457	541	4624953	25580
Yp3828	903.131	542	4502476	24470
Yp3572	1.571.889	545	4613953	24376
Yp3784	1.062.464	545	4428474	24560
Yp3562	1.144.009	546	4600775	25162
Yp3971	1.158.837	546	4609249	22740

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3548	2.477.514	523	4634823	24576
Yp3799	883.523	523	4609564	24461
Yp3759	1.343.735	525	4452679	23209
Yp3872	1.132.940	525	4501057	24310
Yp3213	2.137.168	526	4629354	23801
Yp3710	2.146.704	526	4455089	20934
Yp3832	1.171.629	526	4603819	24560
Yp3494	1.443.085	527	4618217	24565
Yp3972	1.101.556	530	4558736	24635
Yp3560	1.171.581	531	4600875	25591
Yp3743	1.465.733	531	4637425	24643
Yp3230	1.527.618	532	4524998	24639
Yp3806	1.074.681	532	4618410	24631
Yp3576	1.206.779	533	4607994	24372
Yp3573	1.176.494	535	4614353	22864
Yp3598	1.251.156	535	4610657	24314
Yp3826	883.898	535	4408810	24538
Yp3350	1.674.728	536	4409040	24490
Yp3975	919.413	562	4551056	24552
Yp3737	1.142.115	563	4625545	22486
Yp3468	981.516	566	4519206	22547
Yp3197	803.811	568	4612281	21561
Yp3992	934.763	570	4604681	23058
Yp3925	964.201	576	4615756	25550
Yp3695	859.375	577	4554467	23050
Yp3561	1.119.810	583	4619178	22548
Yp3698	1.131.617	583	4629201	24134

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3549	1.064.611	548	4593295	25414
Yp3584	975.687	548	4599061	24548
Yp3830	955.444	549	4482059	24465
Yp3983	803.655	549	4491648	22971
Yp3935	962.140	550	4625741	24290
Yp3208	1.557.593	551	4638043	21973
Yp3460	1.031.833	552	4626887	25348
Yp3749	1.105.254	553	4619745	24552
Yp3990	1.312.443	553	4614840	24556
Yp3924	1.126.945	554	4623290	22736
Yp3699	812.256	555	4517418	22720
Yp3713	1.022.390	555	4622098	22175
Yp3233	1.522.142	559	4629361	21827
Yp3690	1.116.481	561	4633019	24556
Yp3791	1.049.864	556	4616408	23177
Yp3690	1.116.481	561	4633019	24556
Yp3958	1.231.822	562	4607281	24265

CEPA	READS	CONTIGS	GENOMA (PB)	N50
Yp3790	753.641	596	4611577	20809
Yp3982	741.416	601	4567720	22034
Yp3884	716.996	609	4614124	21553
Yp3762	3.520.768	635	4663968	24342
Yp3987	677.660	642	4501792	19465
Yp3566	732.748	643	4607708	19437
Yp3658	3.079.014	663	4635042	19185
Yp3228	655.848	698	4581074	15602
Yp3190	550.423	703	4526757	17529
Yp3232	561.250	721	4495419	16978
Yp3227	592.283	733	4628217	15914
Yp3193	407.908	807	4591631	16615
Yp3276	373.622	808	4528925	14319
Yp3221	323.814	881	4486235	16274
Yp3577	512.926	1009	4558331	9985
Yp3286	397.307	1440	4597187	6480

Fonte: O autor

APÊNDICE B - CEPAS DE BASE DE DADOS PÚBLICA UTILIZADAS NO ESQUEMA DO cgMLST

NOME DA CEPA	ACESSO/NCBI	BIOPROJECT	BIOSAMPLE
			(continua)
KIM10+	GCF_000006645.1	PRJNA288	SAMN02604093
91001	GCF_000007885.1	PRJNA10638	SAMN02602970
CO92	GCF_000009065.1	PRJNA34	SAMEA1705942
NEPAL516	GCF_000013805.1	PRJNA16646	SAMN02598370
ANTIQUA	GCF_000013825.1	PRJNA16645	SAMN02598369
PESTOIDES F	GCF_000016445.1	PRJNA16700	SAMN02598373
ANGOLA	GCF_000018805.1	PRJNA16067	SAMN02604042
D106004	GCF_000022805.1	PRJNA36507	SAMN02603871
Z176003	GCF_000022845.1	PRJNA36547	SAMN02603873
HARBIN 35	GCF_000186725.1	PRJNA30505	SAMN02603222
A1122	GCF_000222975.1	PRJNA67155	SAMN02603531
EV NIEG	GCF_000590535.2	PRJNA234351	SAMN02598688
PBM19	GCF_000834235.1	PRJNA240104	SAMN03074081
HARBIN35	GCF_000834275.1	PRJNA240100	SAMN02769655
PESTOIDES F	GCF_000834315.1	PRJNA239345	SAMN03121016
SHASTA	GCF_000834335.1	PRJNA239341	SAMN03120838
EL DORADO	GCF_000834495.1	PRJNA240085	SAMN03145064
A1122	GCF_000834755.1	PRJNA240102	SAMN03074123
DODSON	GCF_000834775.1	PRJNA240106	SAMN03161124
ANTIQUA	GCF_000834825.1	PRJNA240103	SAMN03174669
ANGOLA	GCF_000834845.1	PRJNA240101	SAMN03177051
NICHOLISK 41	GCF_000834885.1	PRJNA240108	SAMN03216612
PESTOIDES B	GCF_000834925.1	PRJNA240105	SAMN03218157
PESTOIDES G	GCF_000834985.1	PRJNA240082	SAMN03202985
1412	GCF_001188695.1	PRJNA208250	SAMN03782999
1522	GCF_001188715.1	PRJNA208252	SAMN03783057
8787	GCF_001188755.1	PRJNA208258	SAMN03943569
3770	GCF_001188775.1	PRJNA208257	SAMN03943654
3067	GCF_001188795.1	PRJNA208254	SAMN03943655
2944	GCF_001188815.1	PRJNA208249	SAMN03941711
1413	GCF_001188935.1	PRJNA208251	SAMN03783058
CO92	GCF_001293415.1	PRJNA235224	SAMN03160774
CADMAN	GCF_001693595.1	PRJNA326509	SAMN05282282
FDAARGOS_603	GCF_003798205.1	PRJNA231221	SAMN10228583
FDAARGOS_601	GCF_003798225.1	PRJNA231221	SAMN10228581
FDAARGOS_602	GCF_003798345.1	PRJNA231221	SAMN10228582
SCPM-O-B-6899 (231)	GCF_009295925.1	PRJNA269675	SAMN12991205
SCPM-O-DNA-18 (I-3113)	GCF_009295945.1	PRJNA269675	SAMN12991206
SCPM-O-B-5935 (I-1996)	GCF_009295965.1	PRJNA269675	SAMN12991208
SCPM-O-B-6530	GCF_009295985.1	PRJNA269675	SAMN07176224

NOME DA CEPA	(Continuação)		
	ACESSO/NCBI	BIOPROJECT	BIOSAMPLE
SCPM-O-B-6291 (C-25)	GCF_009296005.1	PRJNA269675	SAMN12991209
SCPM-O-B-5942 (I-2638)	GCF_009363195.1	PRJNA269675	SAMN12991207
14D	GCF_015159615.2	PRJNA669044	SAMN16431992
S19960127	GCF_015190655.1	PRJNA577233	SAMN13021983
C-781	GCF_015336085.1	PRJNA673664	SAMN16619712
M2029	GCF_015336265.1	PRJNA673664	SAMN16619718
I-1252	GCF_015336465.1	PRJNA673664	SAMN16619721
M2086	GCF_015336695.1	PRJNA673664	SAMN16619723
M-1974	GCF_015336865.1	PRJNA673664	SAMN16619719
C-792	GCF_015337085.2	PRJNA673664	SAMN16619714
C-783	GCF_015337285.1	PRJNA673664	SAMN16619713
KM 567	GCF_015337445.1	PRJNA673664	SAMN16619715
M-1482	GCF_015337645.1	PRJNA673664	SAMN16619716
M-1770	GCF_015337825.2	PRJNA673664	SAMN16619717
M2085	GCF_015338045.2	PRJNA673664	SAMN16619720
C-830	GCF_015338205.1	PRJNA673664	SAMN16619722
R	GCF_024498375.1	PRJNA766482	SAMN21848823
94	GCF_024498395.1	PRJNA766481	SAMN21848822
20	GCF_024498415.1	PRJNA766484	SAMN21848824
EV76-CN	GCF_024758685.1	PRJNA830668	SAMN27736294

Fonte: O autor

APÊNDICE C – ARTIGOS CIENTÍFICOS PUBLICADOS

1 - Artigo científico publicado relacionado a tese, disponível em:

<https://www.mdpi.com/2076-0817/12/8/991>

<https://doi.org/10.3390/pathogens12080991>



Open Access Article

Genomic Analysis of *Yersinia pestis* Strains from Brazil: Search for Virulence Factors and Association with Epidemiological Data

by João Luiz de Lemos Padilha Pitta ^{1,2,*} Matheus Filgueira Bezerra ¹ ,
 Diego Leandro Reis da Silva Fernandes ¹ , Tessa de Block ³ , Ane de Souza Novaes ⁴ ,
 Alzira Maria Paiva de Almeida ¹ and Antonio Mauro Rezende ^{1,2,*}

¹ Microbiology Department of Aggeu Magalhães Institute—FIOCRUZ PE, Recife 50740-465, PE, Brazil

² Bioinformatics Platform of Aggeu Magalhães Institute—FIOCRUZ PE, Recife 50740-465, PE, Brazil

³ Department of Clinical Sciences—Institute of Tropical Medicine, 2000 Antwerp, Belgium

⁴ Department of Biological Sciences—Federal University of Vale do São Francisco—UNIVASF, Petrolina 56300-000, PE, Brazil

* Authors to whom correspondence should be addressed.

Pathogens **2023**, *12*(8), 991; <https://doi.org/10.3390/pathogens12080991>

Received: 14 June 2023 / Revised: 13 July 2023 / Accepted: 22 July 2023 / Published: 28 July 2023

(This article belongs to the Special Issue **An Ancient Terror, *Yersinia pestis*: What's New?**)

Abstract

Yersinia pestis, the etiological agent of the plague, is considered a genetically homogeneous species. Brazil is currently in a period of epidemiological silence but plague antibodies are still detected in sentinel animals, suggesting disease activity in the sylvatic cycle. The present study deployed an in silico approach to analyze virulence factors among 407 Brazilian genomes of *Y. pestis* belonging to the Fiocruz Collection (1966–1997). The pangenome analysis associated several known virulence factors of *Y. pestis* in clades according to the presence or absence of genes. Four main strain clades (C, E, G, and H) exhibited the absence of various virulence genes. Notably, clade G displayed the highest number of absent genes, while clade E showed a significant absence of genes related to the T6SS secretion system and clade H predominantly demonstrated the absence of plasmid-related genes. These results suggest attenuation of virulence in these strains over time. The cgMLST analysis associated genomic and epidemiological data highlighting evolutionary patterns related to the isolation years and outbreaks of *Y. pestis* in Brazil. Thus, the results contribute to the understanding of the genetic diversity and virulence within *Y. pestis* and the potential for utilizing genomic data in epidemiological investigations.

Keywords: *Yersinia pestis*; plague; virulence factor; comparative genomics

2 – Outras publicações científicas

2.1 REVISTA PEERJ, 2021



In silico predictions of protein interactions between Zika virus and human host

João Luiz de Lemos Padilha Pitta^{1,*}, Crhislane Rafele dos Santos Vasconcelos^{2,*}, Gabriel da Luz Wallau³, Túlio de Lima Campos⁴ and Antonio Mauro Rezende¹

¹ Microbiology Department, Aggeu Magalhães Institute-FIOCRUZ/PE, Recife, PE, Brasil

² Genetics Department, Federal University of Pernambuco, Recife, PE, Brasil

³ Entomology Department, Aggeu Magalhães Institute-FIOCRUZ/PE, Recife, PE, Brasil

⁴ Bioinformatics Platform, Aggeu Magalhães Institute-FIOCRUZ/PE, Recife, PE, Brasil

* These authors contributed equally to this work.

ABSTRACT

Background: The ZIKA virus (ZIKV) belongs to the *Flaviviridae* family, was first isolated in the 1940s, and remained underreported until its global threat in 2016, where drastic consequences were reported as Guillan-Barre syndrome and microcephaly in newborns. Understanding molecular interactions of ZIKV proteins during the host infection is important to develop treatments and prophylactic measures; however, large-scale experimental approaches normally used to detect protein-protein interaction (PPI) are onerous and labor-intensive. On the other hand, computational methods may overcome these challenges and guide traditional approaches on one or few protein molecules. The prediction of PPIs can be used to study host-parasite interactions at the protein level and reveal key pathways that allow viral infection.

Results: Applying Random Forest and Support Vector Machine (SVM) algorithms, we performed predictions of PPI between two ZIKV strains and human proteomes. The consensus number of predictions of both algorithms was 17,223 pairs of proteins. Functional enrichment analyses were executed with the predicted networks to access the biological meanings of the protein interactions. Some pathways related to viral infection and neurological development were found for both ZIKV strains in the enrichment analysis, but the JAK-STAT pathway was observed only for strain PE243 when compared with the FSS13025 strain.

Conclusions: The consensus network of PPI predictions made by Random Forest and SVM algorithms allowed an enrichment analysis that corroborates many aspects of ZIKV infection. The enrichment results are mainly related to viral infection, neuronal development, and immune response, and presented differences among the two compared ZIKV strains. Strain PE243 presented more predicted interactions between proteins from the JAK-STAT signaling pathway, which could lead to a more inflammatory immune response when compared with the FSS13025 strain. These results show that the methodology employed in this study can potentially reveal new interactions between the ZIKV and human cells.

Subjects Bioinformatics, Microbiology, Molecular Biology, Virology

Keywords Zika virus, Protein-protein interactions, Machine learning, Metabolic pathway

Submitted 25 November 2020

Accepted 23 June 2021

Published 24 August 2021

Corresponding authors

João Luiz de Lemos Padilha Pitta,

jlpitta82@gmail.com

Antonio Mauro Rezende,

antonio.rezende@fiocruz.br

Academic editor

Brett Pickett

Additional Information and
Declarations can be found on
page 17

DOI 10.7717/peerj.11770

© Copyright

2021 Pitta et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Epitranscriptome machinery in Trypanosomatids: New players on the table?

Suellen Rodrigues Maran¹ | João Luiz de Lemos Padilha Pitta² | Chrisllane Rafael dos Santos Vasconcelos² | Suzanne M. McDermott³ | Antonio Mauro Rezende² | Nilmar Silvio Moretti³ 

¹Laboratory of Molecular Biology of Pathogens, Department of Microbiology, Immunology and Parasitology, Federal University of Sao Paulo, São Paulo, Brazil

²Department of Microbiology, Aggeu Magalhães Institute/FIOCRUZ, Recife, Brazil

³Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, USA

Correspondence

Antonio Mauro Rezende, Department of Microbiology, Aggeu Magalhães Institute/FIOCRUZ, Recife, PE, Brazil.
Email: antonio.rezende@cpqam.fiocruz.br

Nilmar Silvio Moretti, Department of Microbiology, Immunology and Parasitology, Federal University of Sao Paulo, São Paulo, SP, Brazil.
Email: nilmar.moretti@unifesp.br

Funding information

Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2018/09948-0 and 2019/13765-1; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 424729/2018-0; Fundação Oswaldo Cruz, Grant/Award Number: 400739/2019-4

Abstract

Trypanosoma and *Leishmania* parasites cause devastating tropical diseases resulting in serious global health consequences. These organisms have complex life cycles with mammalian hosts and insect vectors. The parasites must, therefore, survive in different environments, demanding rapid physiological and metabolic changes. These responses depend upon regulation of gene expression, which primarily occurs post-transcriptionally. Altering the composition or conformation of RNA through nucleotide modifications is one posttranscriptional mechanism of regulating RNA fate and function, and modifications including N6-methyladenosine (m6A), N1-methyladenosine (m1A), N5-methylcytidine (m5C), N4-acetylcytidine (ac4C), and pseudouridine (Ψ), dynamically regulate RNA stability and translation in diverse organisms. Little is known about RNA modifications and their machinery in Trypanosomatids, but we hypothesize that they regulate parasite gene expression and are vital for survival. Here, we identified Trypanosomatid homologs for writers of m1A, m5C, ac4C, and Ψ and analyze their evolutionary relationships. We systematically review the evidence for their functions and assess their potential use as therapeutic targets. This work provides new insights into the roles of these proteins in Trypanosomatid parasite biology and treatment of the diseases they cause and illustrates that Trypanosomatids provide an excellent model system to study RNA modifications, their molecular, cellular, and biological consequences, and their regulation and interplay.

KEYWORDS

ac4C, m1A, m5C, pseudouridine, RNA modification, Trypanosomatids