pathogens

MDPI

Article

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

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; matheus.bezerra@fiocruz.br (M.F.B.); leanreis7@gmail.com (D.L.R.d.S.F.); alzira.almeida@fiocruz.br (A.M.P.d.A.)
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; tdeblock@itg.be
4 Department of Biological Sciences—Federal University of Vale do São Francisco—UNIVASF, Petrolina 56300-000, PE, Brazil; ane.snovaes@gmail.com
* Correspondence: jlpitta82@gmail.com (J.L.d.L.P.P.); antonio.rezende@fiocruz.br (A.M.R.)

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

check for updates

Citation: Pitta, J.L.d.L.P.; Bezerra, M.F.; Fernandes, D.L.R.d.S.; Block, T.d.; Novaes, A.d.S.; Almeida, A.M.P.d.; Rezende, A.M. Genomic Analysis of *Yersinia pestis* Strains from Brazil: Search for Virulence Factors and Association with Epidemiological Data. *Pathogens* **2023**, *12*, 991. https://doi.org/10.3390/pathogens12080991

Academic Editor: Ronan Jambou

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

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The plague is an infectious disease caused by *Yersinia pestis*, a nonmotile, aerobic Gram-negative bacillus that has already devastated human populations throughout its three major recorded pandemics [1]. The first pandemic, called the Plague of Justinian, devastated the civilized world between 542 and 602 AD. The second plague pandemic started in Asia and reached Europe and North Africa, enduring from the XIV until the XVI century. It claimed the lives of one third of the European population in the period from 1346 to 1353, which, nowadays, is commonly known as the Black Death [2]. Plague was introduced into Brazil during the current pandemic, in 1899, by rats and fleas present in commercial steamships in the port of Santos in the State of São Paulo [3]. Due to the transport and trade of goods, the plague spread inland through roads, railways, and river transportation, reaching, initially, countryside villages. In some regions, the plague could establish new foci by maintaining a sustainable cycle in sylvatic rodents from local ecosystems, causing many outbreaks in rural landscapes within the Caatinga biome. These outbreaks occurred across several states in the Northeast region (Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Bahia,

and Piauí) and the northeast of Minas Gerais (Vale do Jequitinhonha region), as shown in Figure 1. Moreover, outside the Caatinga, there is a smaller plague transmission area in the Atlantic Forest biome: Vale do Rio Doce in Minas Gerais and Serra dos Órgãos in the state of Rio de Janeiro [3–5]. The distribution of *Y. pestis* cases in Brazil, according to the National Reference Service in Plague reported from 1966 to 1997 (data not published), is represented in Figure 1.

Figure 1. Distribution of plague cases in Brazil confirmed by bacteriological culture and serology between 1966 and 1997. Each color represents different Brazilian biomes. Red dots on the map mark the local foci where cases were registered.

The belief that the plague has already been eradicated and that it is restricted to the distant past of humanity is a serious mistake, as it continues to be active and causes deaths throughout the world. The most recent data shows records of infection in several parts of the world, especially in Madagascar, and the African continent is the place with the highest incidence of cases. Only in Madagascar, 2323 cases were recorded between 2013 and 2018, and there are also records of 410 cases in the Democratic Republic of Congo, 36 cases in Tanzania, and 22 in Uganda [6,7]. In the same period, some cases were reported in Asia, China, and Mongolia, with five cases reported in each of these countries. In addition, in South America, 40 cases were registered in Peru and 3 in Bolivia. Even in more developed countries, with greater control and better sanitary conditions, cases have been registered, such as the 40 cases registered in the United States between 2013 and 2018 [6,7]. This large number of cases that continue to occur around the world, along with the great virulence of *Y. pestis,* are associated with a refined arsenal of gene regulation. The complex life cycle of the bacillus includes the ability to grow and multiply in diverse organisms, such as ectoparasites and mammals, and the regulation of virulence factors that are decisive to host adherence, modulation of the immune response, interruption of cellular metabolism and to obtain essential nutrients [8]. From a genomic point of view, *Y. pestis* is widely known to be highly conserved, especially when comparing the second great pandemic strains with modern strains [1]. Observing the reference strain KIM (GCF_000006645), which originated from the second pandemic, and the modern reference

strain CO92 (GCF_000009065), they both have a genome of 4.6 Mb, just over 4000 ORFs and CG content around 47%, also sharing major virulence genes and more than 95% of the genome [9,10]. The strains associated with the first pandemic differ from the modern ones, with the absence of genes such as *pheA* and *celB*. Strains from the final period of the first pandemic also have lost *mgtB* and *mgtC* genes, a fact that was associated with the end of the first pandemic [1,11]. Throughout history, the evolution of *Y. pestis* has had two notable points. First, it diverged into an ancient strain known as the Neolithic Gok2 approximately 5700 years ago, established as an independent lineage in Eurasia [12]. Secondly, during a timeframe spanning from 3300 B.P. to 5300 years ago, it diverged into different strains that acquired all the necessary genetic components for transmission through fleas. This included the integration of the *ymt locus*, which enabled the bacillus to survive within the midgut/proventriculus of fleas and inside macrophages [13–15].

2. Materials and Methods

2.1. Biological Data

The present study used 411 genomes of *Y. pestis* belonging to the National Reference Service in Plague (Serviço de Referência em Peste—SRP). The strains were isolated from rodent, flea, and human cases from epidemic and endemic periods during surveillance activities since 1966 and were maintained in plain agar stubs, in rubber-stoppered tubes, under refrigeration at 4 °C [16]. These genomes are deposited in the NCBI database under BioProject: PRJNA421720. Genomes were sequenced using the Illumina HiSeq 2000 platform (Illumina, Inc., San Diego, CA, USA) using standard protocols for the TruSeq SBS Kit v3-HS (Illumina, Code FC-401-3001) [3]. The strains belong to the *Y. pestis* cultures collection Fiocruz/CYP (http://cyp.fiocruz.br/accessed on 4 April 2019) of the Aggeu Magalhães Institute (IAM/Fiocruz-PE) [17]. The collection began as an initiative of the Brazilian government and the World Health Organization (WHO) in which a broad research program was developed for plague control in the Araripe plateau between 1966 and 1974. In 1974, the collection was transferred to the central laboratory of Garanhuns, which operated as a plague diagnosis center for the entire country of Brazil until 1982 and, finally, it was transferred to the Aggeu Magalhães Institute, receiving new strains until 1997, when the last strains were isolated. This collection is part of SpeciesLink, a distributed information system that integrates primary data from scientific collections and is registered with the World Federation Culture Collection (WFCC).

2.2. Evaluation of Genome Sequencing

The quality of the sequencing data of the 411 strains of *Y. pestis* was evaluated using FastQC 0.11.8 software [18]. FastQC generated data such as the total number of sequences, sequence sizes, percentage of GC content, and the Phred (quality measurement value) of the bases. Due to the large number of sequences and data generated by FastQC, MultiQC v1.10 [19] was used to consolidate the results that have an average 2.2 million of reads, GC content of 47%, and average depth of 42×. Sequencing data numbers are available in Supplementary Material Table S1.

2.3. Genome Assembly and Annotation

Sequencing data was filtered using Trimmomatic v. 0.38 [20] using the values of Phred 33 and the parameters LEADING:10, AVGQUAL:30, TRAILING:10, SLIDINGWINDOW:4:15, and MINLEN:60. For each isolate, the paired-end filtered data were used in de novo assembly of the genomes. The assembly was performed using the VelvetOptimiser script, developed to optimize parameters for reassembly using the Velvet program [21], which was configured with an N50 optimization function for k-mer selection. The analysis of the assembled genomes was performed using GenomeTools 1.5.8 [22] and the results were aggregated using an in-house script. The genome annotation process was performed with the Prokka pipeline [23], which includes several tools such as Prodigal [24] and Sig-

nalP [25] to annotate relevant genome features within contigs, such as coding sequences (CDS) and signal leader peptides.

After assembly, 5 strains (Yp3290, Yp3202, Yp3296, Yp3241, and Yp3959) were excluded from the study, resulting in 407 genomes for the subsequent analyses. Excluded strains had genomes sizes far below the average (4.6 Mb) expected for *Y. pestis* and the number of reads was also lower compared with other strains and could compromise the analyses. Data regarding the assembly, their respective sizes, and the number of reads, including removed strains, can be found in Supplementary Material Table S2.

2.4. Pangenome Analysis

The data from the functional annotation performed by Prokka was used as input to the pangenomics analysis using the Roary software V.3.11.2 [26]. Pangenomics analysis allowed the description of genes shared by several strains (core genome) [26]. This analysis generated a matrix of the presence and absence of accessory genes and a phylogenetic tree based on this matrix. To generate a visualization of the standard Roary results, the roary_plot.py script was used. This script uses Roary's output files to produce a pie chart showing the distribution between core and accessory genes and the phylogenetic tree with the matrix of presence and absence of genes. The data relating to the strains and their associated clades can be found in Supplementary Material Table S3.

2.5. Statistical Analysis of Presence and Absence Genes in Pangenomics

Statistical analysis of the presence and absence of genes for the 407 strains was performed with Scoary V.1.6.16 [27] using Fisher's exact test, which is a statistical significance test model for analyzing contingency tables with hypergeometric distribution. Scoary crossed the data from Roary pangenomic analysis against phenotypes treated as binary data from strains belonging to the clades highlighted in the Roary phylogenetic tree. This analysis allowed for highlighting present or absent genes that are specific to these clades, grouped by statistical relevance, in addition to bringing numbers on how many isolates a gene has within a clade, how many do not have a gene within a clade, and how many have the gene in other strains that do not belong to a clade. All of Scoary's analysis results are available in Supplementary Material Table S4.

2.6. Identification of Virulence Factors

To identify virulence genes in the assembled genomes of the 407 strains of *Y. pestis*, ABRicate V1.0.0 was used (https://github.com/tseemann/abricate accessed on 26 July 2021) with the Virulence Factor Database (VFDB) [28]. ABRicate allowed for the automatization of the analysis given the data volume. The downloaded VFDB database was used as a subject set to perform a local alignment in the predicted genome's CDS with the BLASTN 2.10.1+ algorithm [29]. Since some proteins with statistical relevance in the clades were previously annotated as hypothetical, we further investigated this group of genes using InterProScan V 5.56-89.0 [30], which employs approaches such as the Hidden Markov Model (HMM) or local alignment and utilizes predictive models, known as signatures, from multiple databases to conduct functional analysis of proteins. It classifies proteins into families, predicts domains, and identifies important functional elements.

2.7. Association of Pangenome Data and Epidemiological Data

The phylogenetic data generated during the pangenome analysis using Roary were combined with epidemiological information obtained from the surveillance sample collection. The resulting tree was then utilized as input for iTOL v4 [31], where visual and textual elements associated with the isolation location, host, and plague foci were incorporated into the tree.

2.8. Core Genome (cg) Multilocus Sequence Typing (MLST)

To perform a gene-by-gene analysis, the cgMLST method was utilized using ChewBBACA v.3.2.0 [32] which creates a genetic profiling scheme based on the core genome, followed by an allele calling process. First, the reference genome for *Y. pestis* CO92 (assembly accession GCF_000009065.1) was used to train the tool with Prodigal [24]. Next, the gene-by-gene scheme was set up using 60 complete genomes of *Y. pestis* belonging to biovars antiqua, mediaevalis, orientalis, and microtus that are available in the National Center for Biotechnology Information (NCBI) database. This step was followed by creating the whole genome (wg) Multilocus Sequence Typing (MLST) and the core genome (cg) MLST (cgMLST) using the genomes from the present study (n = 407) and the additional dataset (n = 61). Paralogs genes were excluded according to the reference genome. The minimum spanning tree algorithm (MSTree V2) implemented in Grapetree [33] was used to visualize allelic loci clusters generated by cgMLST. The complete list of the genomes used in this analysis is available in Supplementary Material Table S5.

3. Results

During the verification of the assembled genomes, it was possible to observe a homogeneous genome size and GC content in the 407 strains compatible with the known data for *Y. pestis*. The average numbers of the genomes analyzed are shown in Table 1. The data used for calculations are available in Supplementary Material Tables S1 and S2.

Table 1. General genome information.

Feature	Value
Average number of readings	2183.187
Average genome size	4.6 Mb
Average N50	367,978
Average depth	42×
GC Content	47%

Pangenome analysis resulted in a set of 5068 genes distributed among the strains. Of these genes, 3287 (64.86% of the total genome) are grouped in the core genome, while the others are distributed in the accessory genome (35.14% of the total genome). A phylogenetic tree and a matrix of the presence and absence of genes, based on this gene distribution, are shown in Figure 2. Among these clades, five main ones (C, E, F, G, and H) were found to have statistically significant genes identified through Scoary's analysis, totaling 228 genes. These clades were highlighted for further analysis.

The resulting phylogenetic tree was integrated with epidemiological data to evaluate the prevalence of gene presence or absence based on factors such as isolation collection site (municipality or state), host (flea, rodent, or human), and plague foci (representing outbreak or endemic locations and years). However, no discernible pattern could be established between the five highlighted clades and the epidemiological data. Figure 3 displays the phylogenetic tree derived from the pangenome analysis associated with the corresponding epidemiological data.

Out of the 228 statistically relevant genes found in the five main clades, 102 were initially labeled as hypothetical proteins. Following the InterProScan analysis, we successfully reannotated 49 of these genes. However, in the case of clade F, four hypothetical proteins remained unassociated with any gene annotation. Then, we analyzed all 228 genes, searching for virulence factors using the VFDB database, which resulted in 57 genes related to virulence factors. Table 2 summarizes the distribution of putative and confirmed virulence genes and nonvirulence genes in clades C, E, F, G, and H.

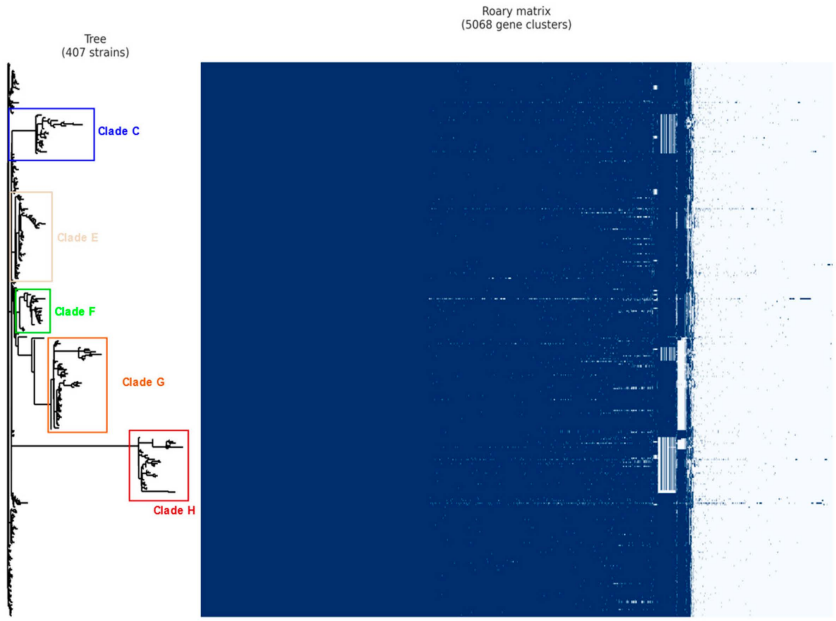

Figure 2. Phylogenetic tree based on the core genome and accessory genome from the pangenome analysis. In the matrix, blue marks indicate the presence of a specific gene, while white marks indicate its absence. The core genome, consisting of a dense portion of shared genes, is located at the beginning of the matrix, followed by the accessory genes in the final portion. The resulting clades are labeled from top to bottom as A to I and the five main clades: C, E, F, G, and H are highlighted in different colors.

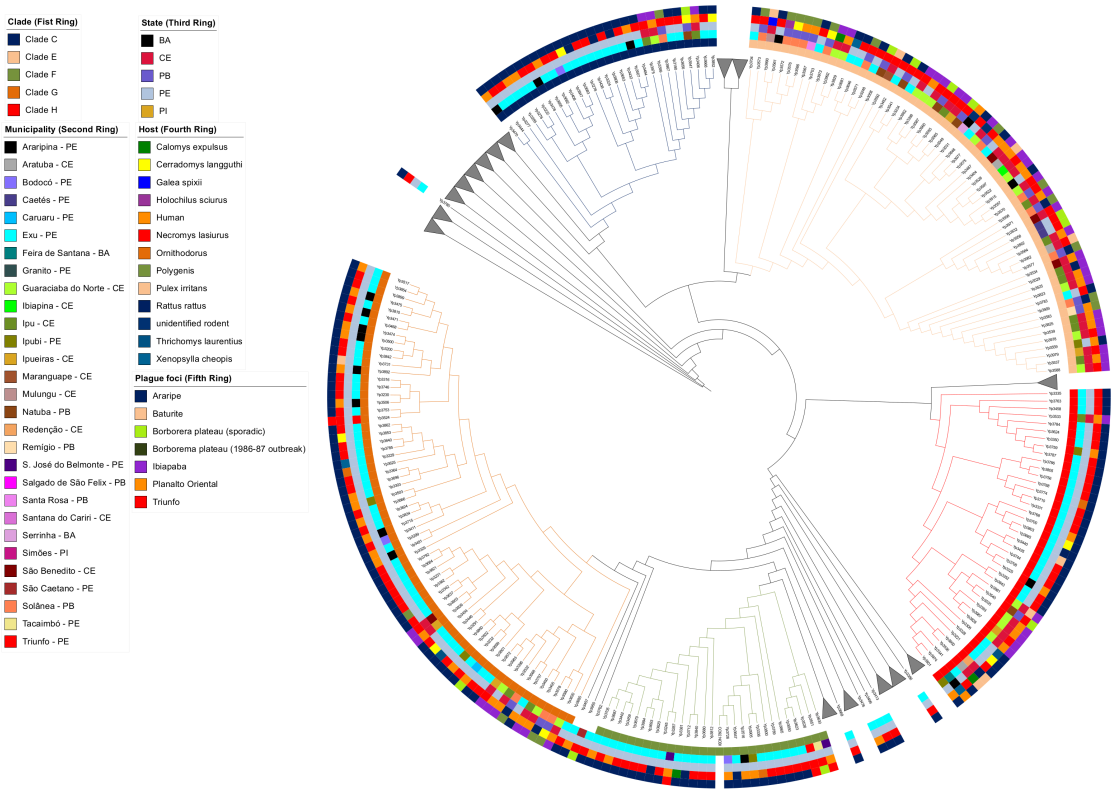

Figure 3. Phylogenetic tree from the pangenome analysis associated with the epidemiological data. The first ring, from the innermost to the outermost, represents the grouping of strains, with the five main clades distinguished by different colors. The subsequent rings display the association between clades and epidemiological data, including municipality, state, host, and plague foci.

Table 2. Distribution of virulence and non-virulence genes in clades C, E, F, G, and H.

Clade	Putative/Virulence Genes	Nonvirulence Genes
C	10	43
E	13	0
F	0	4
G	25	41
H	9	83

Figure 4 presents the virulence-related genes enriched as absent through Scoary analysis. These absent genes are grouped by clade and organized based on their corresponding *p*-values. The complete information on all genes from all clades, including those without statistical relevance or genes not associated with virulence, along with their descriptions, annotations, and corresponding statistical values, can be found in Supplementary Material Table S4.

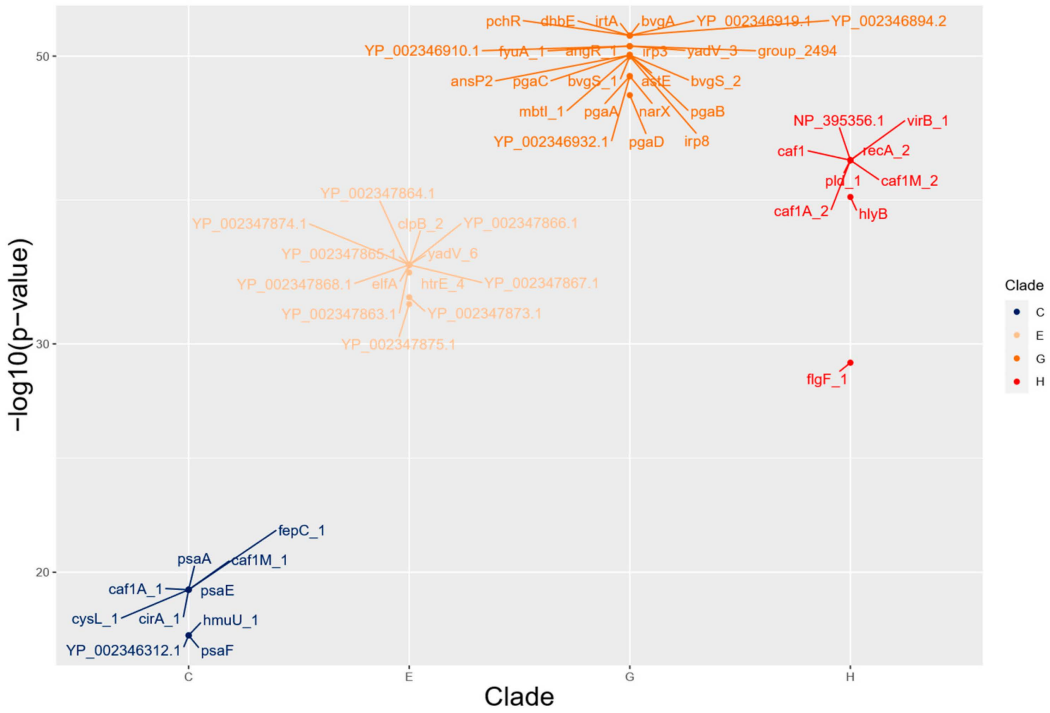

Figure 4. Absent virulence-related genes grouped by clade and log10(*p*-value). Each of the main clades is represented by a different color.

The cgMLST analysis generated new gene annotations that resulted in a set of 4088 genes among the strains, of which 3206 are in the core genome (78.42% of the total genome) and 882 genes in the accessory genome (21.58% of the total genome). The comparison of the gene distribution between the pangenomic approach using Roary and cgMLST using ChewBBACA is shown in Figure 5.

The genetic-profiling scheme generated, followed by a gene-by-gene analysis based on the core genome, produced a graph structure file using the minimum spanning tree approach which was crossed with the epidemiological data in the form of metadata in the Grapetree webservice. The resulting graph, according to the strain collection date, is represented in Figure 6. It is possible to observe three main clusters from which several evolutions depart. The first two, on the left of the image, are mainly characterized by strains from the year 1969. The third is the largest among the three main clusters, containing a greater number of strains. Most of these strains originate from the years 1974 and 1975, and

are represented by two different colors. It is also possible to observe a small cluster, very characteristic of the year 1986, colored in purple derived from the main largest cluster.

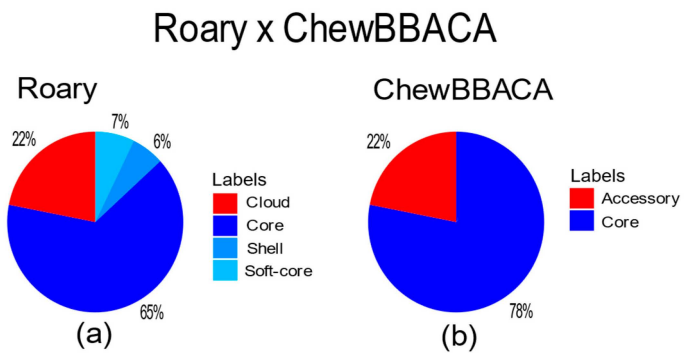

Figure 5. Distribution of genes after pangenome and cgMLST analysis. (**a**) shows that the largest portion of genes is in the core genome with 65% (3287 genes). The remaining genes are further classified within the accessory genome comprising 7% (386 $\leq$ strains $\leq$ 402) into soft core, 6% (61 $\leq$ strains $\leq$ 386) into shell, and 22% into cloud (strains $\leq$ 61). (**b**) shows the largest portion of genes in the core genome with 78% (3206 genes).

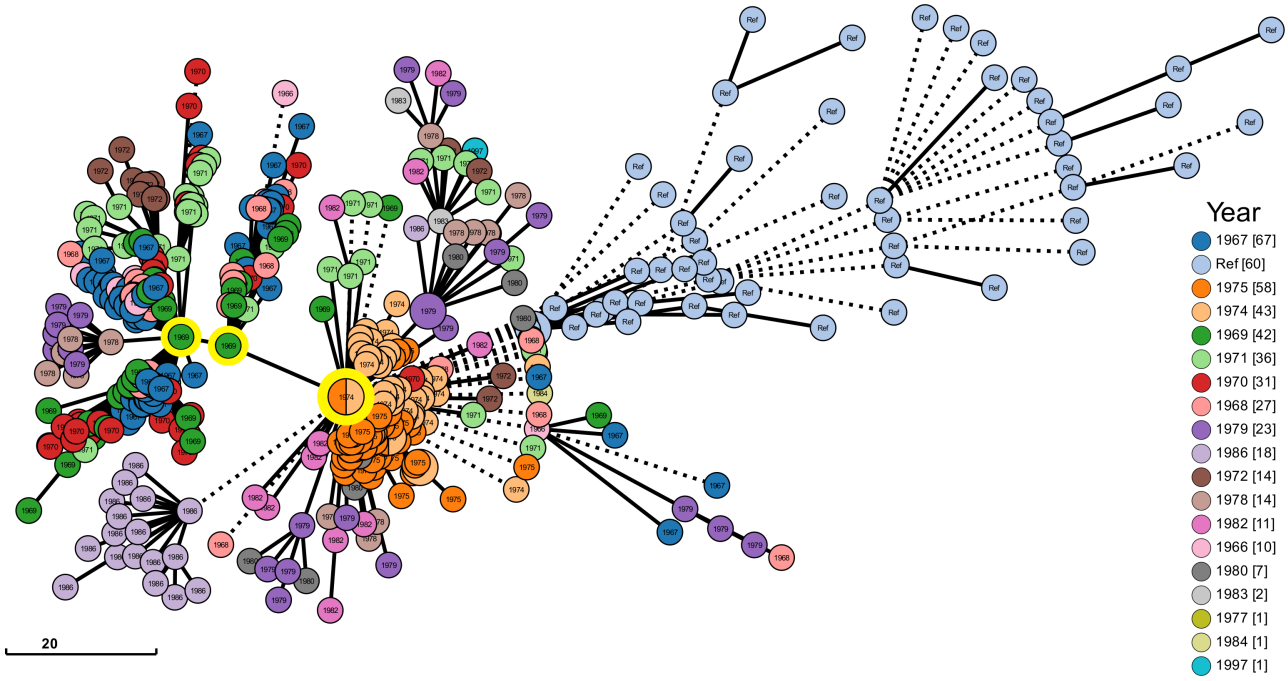

Figure 6. Graph showing clusters of strains based on gene-by-gene analysis. Each colored node represents a cluster of strains grouped according to allele similarity associated with the year of collection that best represents the cluster. Genetic similarity is represented by edge sizes. Three main clusters of strains representing evolutions departing are highlighted with a yellow circle. The number in brackets next to the year indicates the number of strains corresponding to the specific year. The size of each node is proportional to the number of strains clustered.

An additional analysis was performed to associate the resulting graph structure with other epidemiological data. This analysis utilized the outbreak parameter, which combines the collection site and the date of positive cases of *Y. pestis*. The resulting image is represented in Figure 7. It also resulted in three main clusters with evolution points, now related to two outbreaks at Chapada do Araripe. From this first outbreak, one subcluster of strains appears self-limited to a short period of time between 1966 and 1971. But, another

derived subcluster was related to an outbreak in Serra de Triunfo (colored in red). It is noteworthy that from the largest main cluster, two subclusters of strains are linked to outbreaks in Chapada da Borborema, one from 1986 colored in green (corresponding to the clusters previously colored in purple in Figure 6), and another colored in yellow (from 1979 to 1982).

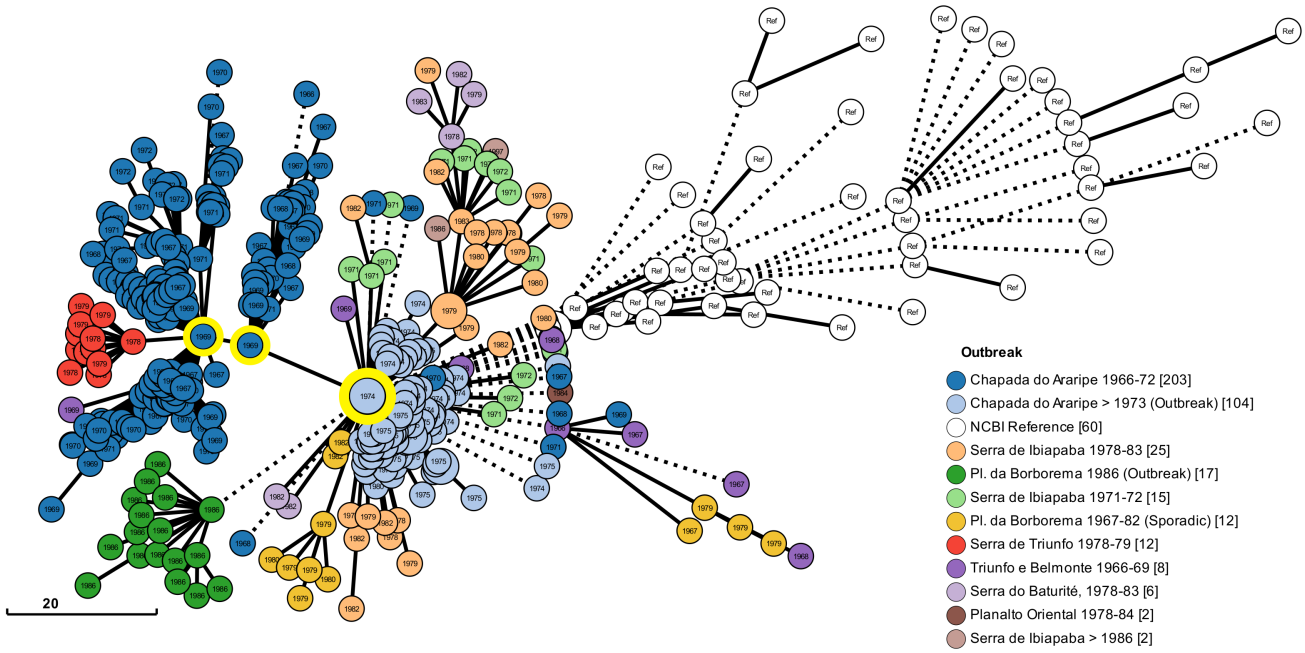

Figure 7. Graph showing clusters of strains based on gene-by-gene analysis. Each colored node represents a cluster of strains grouped according to allele similarity associated with *Y. pestis* outbreaks in Brazil and is labeled with the year of collection that best represents the cluster. Genetic similarity is represented by edges sizes. Three main clusters of strains representing evolutions departing are highlighted with a yellow circle. The number in brackets next to the outbreak name indicates the number of strains corresponding to the specific outbreak. The size of each node is proportional to the number of strains clustered.

4. Discussion

The pangenome and gene-by-gene studies showed the largest concentration of genes in the core genome, which is an expected finding since *Y. pestis* is known for its conserved genome. Even the modern strains of *Y. pestis* exhibit minimal differences from the strains of the second pandemic, particularly in terms of virulence factors [1]. The proportion of genes in the core genome was compatible between the Roary and ChewBBACA analyses. The number of genes annotated in silico in the present study is also consistent with another *Y. pestis* study [9], and even with a study that used more traditional approaches in the laboratory [34]. Another interesting finding in relation to the pangenome phylogenetic tree is the clear distinction of certain clades in the gene matrix based on groups of genes that are either present or absent in comparison to other strains.

Regarding the epidemiological features of the strains, we observed no discernible patterns between the genomic profile and the collection location, period, or host. While clade E showed a high diversity of epidemiological features, clade F showed a clearer pattern with a significant portion of the samples associated with plague foci of Araripe. The genes analyzed using the VFDB returned several virulence factors, including important well-known genes in *Y. pestis*, such as those involved in biofilm formation, iron acquisition, and cell adhesion genes [35–37]. The gene clusters returned by Scoary in clades C, E, G, and H are characterized by the absence of virulence genes; however, it is necessary to consider the collection period of these strains (some from the 1960s), the prolonged

laboratory handling and storage process in agar stabs, in rubber-stoppered tubes, and under refrigeration at 4 °C [16].

One notable observation was the absence of genes from the *pga* family (*pgaA*, *pgaB*, *pgaC*, and *pgaD*), specifically within clade G. This gene family is renowned for its crucial role in maintaining and providing structural stability to the biofilm in various eubacteria [38] and is orthologous to the *hms* family, specific to *Y. pestis*, known for the synthesis of the extracellular matrix (ECM) in biofilm formation [36,38]. Another striking absence observed in the data from clade G was the absence of the *irp3* and *irp8* genes. This is noteworthy because genes belonging to the *irp* family play a crucial role in virulence, particularly in the synthesis of Yersiniabactin [39,40]. A study has demonstrated that the deletion of *irp8* results in the reduced virulence of the pneumonic plague [41] but the absence of these genes may be attributed to repeated subculturing, as previously demonstrated with *irp2* [42].

An unexpected absence was observed in the caf operon genes (*caf1*, *caf1A,* and *caf1M*). The *caf1* gene is considered specific for *Y. pestis*, being used even for diagnostic purposes [35], and is absent in clade H while the genes *caf1M* and *caf1A* are absent in clades H and C. This operon is important for *Y. pestis* virulence since it is responsible for the F1 antigen that inhibits the phagocytosis by macrophages [43] but other work has already demonstrated the absence of these genes, possibly due to the instability of the pFra plasmid [44]. Furthermore, it has been demonstrated that the activation of *caf1* occurs at 37 °C [43] and the gene's function is dependent on the host and its genetic markers [45]. Another gene that is considered host dependent and, similar to *caf1*, is absent in clade C, is *psaA* [45,46]. This gene is responsible for the expression of the pH6 antigen, which plays a crucial role in the virulence of the plague by inhibiting phagocytosis by macrophages and mediating the binding of the bacillus to epithelial cells [35,46]. In addition to the *psaA*, other genes from the *psa* family (*psaE* and *psaF*) are also absent in clade C. These genes play an important regulatory role in *psaA*, mediated through distinct post-transcriptional effects under the optimal conditions of temperature and pH for *psaA* expression [47]. The absence of *caf1* and *psaA* in the strains of clade C is significant because the full pathogenicity of *Y. pestis* requires these two genes [45].

The *pld* gene, also referred to as *ymt* (Yersinia murine toxin), and highlighted as absent in clade H, is another important gene for *Y. pestis* and is considered a recent genetic adaptation that allowed transmission of the bacillus via fleas, playing an important role in their midgut survival [14,15,35]. Another important aspect observed in the results was the absence of genes associated with iron acquisition and iron transport in clades G and C. The acquisition of iron plays an important role as a virulence factor because *Y. pestis* needs to capture iron for its growth, multiplication, and biological functions [35,48,49]. Sebbane et al. (2006) [50] demonstrated the importance of iron-transport genes over iron-acquisition genes, possibly due to redundant systems in *Y. pestis*; in addition, some studies have demonstrated that reduced iron acquisition can lead to virulence attenuation [49,51].

After reannotation using InterProScan, all nine genes from clade E, which were initially annotated as hypothetical, were found to be related to the T6SS (Type VI Secretion System). The T6SS is widely distributed among Gram-negative bacteria, and functions as a delivery system for various pathogenic proteins, and *Y. pestis* is known to have multiple copies [52]. Due to limited knowledge about this secretion system in *Y. pestis*, it becomes challenging to precisely measure the extent of this impact on the virulence. Together, the absence of these virulence genes in all the main clades highlighted corroborates with previous works inferring virulence attenuation in *Y. pestis* strains from northeastern Brazil [53].

Through the gene-by-gene analysis performed by cgMLST genomic and epidemiological data were associated. Unlike the phylogenetic approach, which considers only a dichotomic relationship between one strain and another, the profile generated in the cgMLST allowed several strains to be grouped in the same node and it was also possible to infer genetic similarity based on the accumulation of allele differences. Looking at the epidemiological data, the architecture of the resulting graph associated with the year of collection (Figure 6) has three main clusters from which many evolutions departed. This is

consistent given that, in Brazil, the first collections are from the year 1966 and several cases appeared until the year 1972, which are represented in two of the main clusters. After a brief epidemiological silence, several cases occurred in the years 1974 and 1975, represented by the largest main cluster. It was also observed that the reference strains, obtained from a public database and not originating from Brazil, are genetically distinct from the Brazilian collection of strains. This observation highlights that the Brazilian strains have undergone their own genetic evolution.

The resulting graph associated with outbreaks in Brazil (Figure 7) also highlighted the three main clusters, now related to the two outbreaks that occurred in the Chapada do Araripe, a large region between three northeast states of Brazil. Two of these main clusters belong to the first outbreak (1966 to 1972) and have two subclusters deriving. One shows a lineage of strains that are apparently self-limiting since almost all of them are from the municipality of Exu and only three strains are from other municipalities very close to each other. In addition, all these strains are from a short period of time (1966 and 1971). But the second derived subcluster has all the strains related to the outbreak that occurred in the municipality of Triunfo, state of Pernambuco, between 1978 and 1979 (red clusters in Figure 7). From the second outbreak in Chapada do Araripe (that began in 1973) two subclusters were derived and occurred in the Chapada da Borborema region. But, despite Chapada da Borborema encompassing five states in Northeast Brazil, all these strains are from the state of Paraíba. The first subcluster has strains from 1979 and 1982 and is from the municipality of Natuba, while the other subcluster has all strains from 1986 and is mainly in the municipality of Solânea. Together, these results enabled further understanding of the evolution of the plague in Brazil over time.

5. Conclusions

The pangenome and cgMLST together allowed a comprehensive analysis of Brazilian *Y. pestis* genomes available at the Fiocruz/CYP collection. The absence of several virulence genes associated with cell adhesion, iron acquisition, and biofilm formation indicates attenuation of virulence in strains from clades C, E, G, and H. The strains of clade G showed the most absence of virulence genes, including very characteristic genes of *Y. pestis* suggesting that this was the most affected group of strains. Clade E greatly benefited from the more indepth analysis of InterProScan showing the absence of genes related to T6SS, which possibly indicates an attenuation of virulence in the strains from this clade but highlights the need to further study the role of this secretion system in *Y. pestis*. Finally, clade H presented the absence of genes mostly related to plasmids and appears to be the group less affected by virulence attenuation among the clades. Through this in silico approach, it was possible to integrate genomic and epidemiological data, shedding light on evolutionary aspects and providing valuable insights into the genetic diversity and virulence of *Y. pestis* in Brazil. Moreover, the utilization of genomic information for epidemiological investigations and outbreak tracing has shown promising potential.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pathogens12080991/s1.

Author Contributions: Conceptualization, A.M.R. and A.M.P.d.A.; methodology, J.L.d.L.P.P., T.d.B., A.d.S.N., D.L.R.d.S.F. and A.M.R.; software, J.L.d.L.P.P., T.d.B., A.d.S.N., D.L.R.d.S.F. and A.M.R.; validation, J.L.d.L.P.P., M.F.B., A.M.R. and A.M.P.d.A.; formal analysis, J.L.d.L.P.P., M.F.B., A.M.R. and A.M.P.d.A.; investigation, J.L.d.L.P.P., M.F.B., A.M.R. and A.M.P.d.A.; resources, A.M.R. and A.M.P.d.A.; data curation, J.L.d.L.P.P., M.F.B., A.M.R. and A.M.P.d.A.; writing—original draft preparation, J.L.d.L.P.P.; writing—review and editing, J.L.d.L.P.P., M.F.B., A.M.R. and A.M.P.d.A.; visualization, J.L.d.L.P.P., M.F.B. and A.M.R.; supervision, A.M.R. and A.M.P.d.A.; project administration, A.M.R. and A.M.P.d.A.; funding acquisition, A.M.R. and A.M.P.d.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article and Supplementary Materials.

Acknowledgments: The authors thank Aggeu Magalhães Institute/FIOCRUZ-PE for the use of its facilities through its Bioinformatics research group and the Reference Service in Plague for all the support. Thanks to The Pathogen and Microbiome Institute—Northern Arizona University for supporting the sequencing of the strains. We also thank the Brazilian funding agencie Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) for their support providing the computational facilities.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Barbieri, R.; Signoli, M.; Chevé, D.; Costedoat, C.; Tzortzis, S.; Aboudharam, G.; Raoult, D.; Drancourt, M. Yersinia Pestis: The Natural History of Plague. *Clin. Microbiol. Rev.* **2020**, *34*, e00044-19. [CrossRef] [PubMed]
2. Bramanti, B.; Stenseth, N.C.; Walløe, L.; Lei, X. Plague: A Disease Which Changed the Path of Human Civilization. In *Advances in Experimental Medicine and Biology*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 918, pp. 1–26. [CrossRef]
3. Vogler, A.J.; Sahl, J.W.; Leal, N.C.; Sobreira, M.; Williamson, C.H.D.; Bollig, M.C.; Birdsell, D.N.; Rivera, A.; Thompson, B.; Nottingham, R.; et al. A Single Introduction of Yersinia Pestis to Brazil during the 3rd Plague Pandemic. *PLoS ONE* **2019**, *14*, e0209478. [CrossRef] [PubMed]
4. da Silva Fernandes, D.L.R.; de Souza Gomes, E.C.; Bezerra, M.F.; de Paula Souza E Guimarães, R.J.; de Almeida, A.M.P. Spatiotemporal Analysis of Bubonic Plague in Pernambuco, Northeast of Brazil: Case Study in the Municipality of Exu. *PLoS ONE* **2021**, *16*, e0249464. [CrossRef]
5. Giles, J.; Peterson, A.T.; Almeida, A. Ecology and Geography of Plague Transmission Areas in Northeastern Brazil. *PLoS Negl. Trop. Dis.* **2011**, *5*, e925. [CrossRef] [PubMed]
6. Vallès, X.; Stenseth, N.C.; Demeure, C.; Horby, P.; Mead, P.S.; Cabanillas, O.; Ratsitorahina, M.; Rajerison, M.; Andrianaivoarimanana, V.; Ramasindrazana, B.; et al. Human Plague: An Old Scourge That Needs New Answers. *PLoS Negl. Trop. Dis.* **2020**, *14*, e0008251. [CrossRef]
7. Bertherat, E. Plague around the World in 2019. *Wkly. Epidemiol. Rec. = Relevé Épidémiologique Hebd.* **2019**, *94*, 289–292.
8. Perry, R.D.; Fetherston, J.D. Yersinia Pestis—Etiologic Agent of Plague. *Clin. Microbiol. Rev.* **1997**, *10*, 35–66. [CrossRef]
9. Deng, W.; Burland, V.; Plunkett, G.; Boutin, A.; Mayhew, G.F.; Liss, P.; Perna, N.T.; Rose, D.J.; Mau, B.; Zhou, S.; et al. Genome Sequence of Yersinia Pestis KIM. *J. Bacteriol.* **2002**, *184*, 4601–4611. [CrossRef]
10. Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; Sebaihia, M.; James, K.D.; Churcher, C.; Mungall, K.L.; et al. Genome Sequence of Yersinia Pestis, the Causative Agent of Plague. *Nature* **2001**, *413*, 523–527. [CrossRef]
11. Keller, M.; Spyrou, M.A.; Scheib, C.L.; Neumann, G.U.; Kröpelin, A.; Haas-Gebhard, B.; Päffgen, B.; Haberstroh, J.; Ribera i Lacomba, A.; Raynaud, C.; et al. Ancient Yersinia Pestis Genomes from across Western Europe Reveal Early Diversification during the First Pandemic (541–750). *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 12363–12372. [CrossRef]
12. Rascovan, N.; Sjögren, K.-G.; Kristiansen, K.; Nielsen, R.; Willerslev, E.; Desnues, C.; Rasmussen, S. Emergence and Spread of Basal Lineages of Yersinia Pestis during the Neolithic Decline. *Cell* **2019**, *176*, 295–305.e10. [CrossRef]
13. Andrades Valtueña, A.; Neumann, G.U.; Spyrou, M.A.; Musralina, L.; Aron, F.; Beisenov, A.; Belinskiy, A.B.; Bos, K.I.; Buzhilova, A.; Conrad, M.; et al. Stone Age Yersinia Pestis Genomes Shed Light on the Early Evolution, Diversity, and Ecology of Plague. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116722119. [CrossRef]
14. Bland, D.M.; Miarinjara, A.; Bosio, C.F.; Calarco, J.; Hinnebusch, B.J. Acquisition of Yersinia Murine Toxin Enabled Yersinia Pestis to Expand the Range of Mammalian Hosts That Sustain Flea-Borne Plague. *PLoS Pathog.* **2021**, *17*, e1009995. [CrossRef]
15. Hinnebusch, B.J.; Rudolph, A.E.; Cherepanov, P.; Dixon, J.E.; Schwan, T.G.; Forsberg, Å. Role of Yersinia Murine Toxin in Survival of Yersinia Pestis in the Midgut of the Flea vector. *Science* **2002**, *296*, 733–735. [CrossRef]
16. Leal, N.C.; Sobreira, M.; Araújo, A.F.Q.; Magalhães, J.L.O.; Vogler, A.J.; Bollig, M.C.; Nottingham, R.; Keim, P.; Wagner, D.M.; Almeida, A.M.P. Viability of Yersinia Pestis Subcultures in Agar Stabs. *Lett. Appl. Microbiol.* **2016**, *62*, 91–95. [CrossRef]
17. Rocha, A.; Barbosa, C.S.; Brandão Filho, S.P.; Oliveira, C.M.F.d.; Almeida, A.M.P.d.; Gomes, Y.d.M. Primeiro Workshop Interno Dos Serviços de Referência Do Centro de Pesquisas Aggeu Magalhães Da Fundação Oswaldo Cruz. *Rev. Patol. Trop./J. Trop. Pathol.* **2010**, *38*, 299–309. [CrossRef]
18. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 21 April 2020).
19. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef]
20. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

21. Zerbino, D.R.; Birney, E. Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs. *Genome Res.* **2008**, *18*, 821–829. [CrossRef]
22. Gremme, G.; Steinbiss, S.; Kurtz, S. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2013**, *10*, 645–656. [CrossRef]
23. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]
24. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef] [PubMed]
25. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat. Methods* **2011**, *8*, 785–786. [CrossRef]
26. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef] [PubMed]
27. Brynildsrud, O.; Bohlin, J.; Scheffer, L.; Eldholm, V. Rapid Scoring of Genes in Microbial Pan-Genome-Wide Association Studies with Scoary. *Genome Biol.* **2016**, *17*, 238. [CrossRef]
28. Chen, L. VFDB: A Reference Database for Bacterial Virulence Factors. *Nucleic Acids Res.* **2004**, *33*, D325–D328. [CrossRef]
29. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST plus: Architecture and Applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]
30. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef]
31. Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef]
32. Silva, M.; Machado, M.P.; Silva, D.N.; Rossi, M.; Moran-Gilad, J.; Santos, S.; Ramirez, M.; Carriço, J.A. ChewBBACA: A Complete Suite for Gene-by-Gene Schema Creation and Strain Identification. *Microb. Genom.* **2018**, *4*, e000166. [CrossRef]
33. Zhou, Z.; Alikhan, N.-F.; Sergeant, M.J.; Luhmann, N.; Vaz, C.; Francisco, A.P.; Carriço, J.A.; Achtman, M. GrapeTree: Visualization of Core Genomic Relationships among 100,000 Bacterial Pathogens. *Genome Res.* **2018**, *28*, 1395–1404. [CrossRef]
34. Zhou, D.; Han, Y.; Song, Y.; Tong, Z.; Wang, J.; Guo, Z.; Pei, D.; Pang, X.; Zhai, J.; Li, M.; et al. DNA Microarray Analysis of Genome Dynamics in Yersinia Pestis: Insights into Bacterial Genome Microevolution and Niche Adaptation. *J. Bacteriol.* **2004**, *186*, 5138–5146. [CrossRef]
35. Demeure, C.E.; Dussurget, O.; Mas Fiol, G.; Le Guern, A.-S.; Savin, C.; Pizarro-Cerdá, J. Yersinia Pestis and Plague: An Updated View on Evolution, Virulence Determinants, Immune Subversion, Vaccination, and Diagnostics. *Genes Immun.* **2019**, *20*, 357–370. [CrossRef]
36. Hinnebusch, B.J.; Erickson, D.L. Yersinia Pestis Biofilm in the Flea Vector and Its Role in the Transmission of Plague. *Curr. Top. Microbiol. Immunol.* **2008**, *322*, 229–248. [CrossRef]
37. Tsang, T.M.; Felek, S.; Krukonis, E.S. Ail Binding to Fibronectin Facilitates Yersinia Pestis Binding to Host Cells and Yop Delivery. *Infect. Immun.* **2010**, *78*, 3358–3368. [CrossRef]
38. Itoh, Y.; Rice, J.D.; Goller, C.; Pannuri, A.; Taylor, J.; Meisner, J.; Beveridge, T.J.; Preston, J.F.; Romeo, T. Roles of PgaABCD Genes in Synthesis, Modification, and Export of the Escherichia Coli Biofilm Adhesin Poly-Beta-1,6-N-Acetyl-D-Glucosamine. *J. Bacteriol.* **2008**, *190*, 3670–3680. [CrossRef]
39. Pelludat, C.; Rakin, A.; Jacobi, C.A.; Schubert, S.; Heesemann, J. The Yersiniabactin Biosynthetic Gene Cluster of Yersinia Enterocolitica: Organization and Siderophore-Dependent Regulation. *J. Bacteriol.* **1998**, *180*, 538–546. [CrossRef]
40. de Almeida, A.M.P.; Guiyoule, A.; Guilvout, I.; Iteman, I.; Baranton, G.; Carniel, E. Chromosomal Irp2 Gene in Yersinia: Distribution, Expression, Deletion and Impact on Virulence. *Microb. Pathog.* **1993**, *14*, 9–21. [CrossRef]
41. Pechous, R.D.; Broberg, C.A.; Stasulli, N.M.; Miller, V.L.; Goldman, W.E. In Vivo Transcriptional Profiling of Yersinia Pestis Reveals a Novel Bacterial Mediator of Pulmonary Inflammation. *mBio* **2015**, *6*, e02302-14. [CrossRef]
42. Almeida, A.M.P.d.; Guiyoule, A.; Leal, N.C.; Carniel, E. Survey of the Irp2 Gene among Yersinia Pestis Strains Isolated during Several Plague Outbreaks in Northeast Brazil. *Mem. Inst. Oswaldo Cruz* **1994**, *89*, 87–92. [CrossRef]
43. Du, Y.; Rosqvist, R.; Forsberg, A. Role of Fraction 1 Antigen of Yersinia Pestis in Inhibition of Phagocytosis. *Infect. Immun.* **2002**, *70*, 1453–1460. [CrossRef] [PubMed]
44. Leal-Balbino, T.; Leal, N.; Lopes, C.; Almeida, A. de Differences in the Stability of the Plasmids of Yersinia Pestis Cultures in Vitro: Impact on Virulence. *Mem. Inst. Oswaldo Cruz* **2004**, *99*, 727–732. [CrossRef]
45. Weening, E.H.; Cathelyn, J.S.; Kaufman, G.; Lawrenz, M.B.; Price, P.; Goldman, W.E.; Miller, V.L. The Dependence of the Yersinia Pestis Capsule on Pathogenesis Is Influenced by the Mouse Background. *Infect. Immun.* **2011**, *79*, 644–652. [CrossRef] [PubMed]
46. Anisimov, A.P.; Bakhteeva, I.V.; Panfertsev, E.A.; Svetoch, T.E.; Kravchenko, T.B.; Platonov, M.E.; Titareva, G.M.; Kombarova, T.I.; Ivanov, S.A.; Rakin, A.V.; et al. The Subcutaneous Inoculation of PH 6 Antigen Mutants of Yersinia Pestis Does Not Affect Virulence and Immune Response in Mice. *J. Med. Microbiol.* **2009**, *58*, 26–36. [CrossRef] [PubMed]
47. Quinn, J.D.; Weening, E.H.; Miner, T.A.; Miller, V.L. Temperature Control of PsaA Expression by PsaE and PsaF in Yersinia Pestis. *J. Bacteriol.* **2019**, *201*, e00217-19. [CrossRef] [PubMed]
48. Hood, M.I.; Skaar, E.P. Nutritional Immunity: Transition Metals at the Pathogen–Host Interface. *Nat. Rev. Microbiol.* **2012**, *10*, 525–537. [CrossRef]

49. Fetherston, J.D.; Kirillina, O.; Bobrov, A.G.; Paulley, J.T.; Perry, R.D. The Yersiniabactin Transport System Is Critical for the Pathogenesis of Bubonic and Pneumonic Plague. *Infect. Immun.* **2010**, *78*, 2045–2052. [CrossRef]
50. Sebbane, F.; Lemaître, N.; Sturdevant, D.E.; Rebeil, R.; Virtaneva, K.; Porcella, S.F.; Hinnebusch, B.J. Adaptive Response of Yersinia Pestis to Extracellular Effectors of Innate Immunity during Bubonic Plague. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11766–11771. [CrossRef]
51. Zauberman, A.; Vagima, Y.; Tidhar, A.; Aftalion, M.; Gur, D.; Rotem, S.; Chitlaru, T.; Levy, Y.; Mamroud, E. Host Iron Nutritional Immunity Induced by a Live Yersinia Pestis Vaccine Strain Is Associated with Immediate Protection against Plague. *Front. Cell Infect. Microbiol.* **2017**, *7*, 277. [CrossRef]
52. Yang, X.; Pan, J.; Wang, Y.; Shen, X. Type VI Secretion Systems Present New Insights on Pathogenic Yersinia. *Front. Cell Infect. Microbiol.* **2018**, *8*, 260. [CrossRef]
53. Karimi, Y.; de Almeida, C.R.; de Almeida, A.M.; Keyvanfar, A.; Bourdin, M. Particularités Des Souches de Yersinia Pestis Isolées Dans Le Nord-Est Du Brésil. *Ann. Microbiol.* **1974**, *125A*, 243–246.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.